Training Strategies for Improving Listeners’ Comprehension of Foreign-Accented Speech

Holly Krech Thomas
University of Colorado Boulder

Follow this and additional works at: https://scholar.colorado.edu/print_theses

Recommended Citation
https://scholar.colorado.edu/print_theses/69
Training Strategies for Improving Listeners’ Comprehension of Foreign-Accented Speech

by

Holly Krech Thomas
B.A., University of Colorado at Denver, 1996
M.A., University of Colorado at Boulder, 1998

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
Department of Linguistics
2004
This thesis entitled:
Training Strategies for Improving Listeners’ Comprehension
of Foreign-Accented Speech
written by Holly Krech Thomas
has been approved for the
Department of Linguistics

Lise Menn (co-chair)

Alice F. Healy (co-chair)

May 28, 2004

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above-mentioned discipline.

HRC protocol #0203.09
ABSTRACT

Holly Krech Thomas (Ph.D., Linguistics and Cognitive Science)

Training Strategies for Improving Listeners’ Comprehension of Foreign-Accented Speech

Directed by Professor Lise Menn, Department of Linguistics, and Professor Alice F. Healy, Department of Psychology, University of Colorado at Boulder

This dissertation examines training methodology for improving native English speakers’ comprehension of foreign-accented speech. The training tasks, imitating and paraphrasing accented speech, were developed out of theoretical considerations about bottom-up and top-down processes in speech perception. Paraphrasing speech encourages a focus on the meaning and leads to practice of top-down processing, whereas imitating speech promotes attention to the accented pronunciation and allows practice of bottom-up processing. Additionally, the tasks support active, implicit learning. Results show an improvement in the comprehension of accented speech after about an hour of training. The imitation task primarily improves understanding of relatively short, decontextualized utterances with few semantic and syntactic cues, which suggests that attention to accented pronunciation helps with perception of discrete segments of speech. In contrast, the paraphrase task tends to improve perception of longer speech samples, implying that attention to meaning may be more beneficial for comprehension of conversational speech. There is evidence that training is durable and transferable, but the exact nature of its durability and transferability needs to be explored in future experiments.
DEDICATION

To Ben, my editor and chef,

and

to my mother, who taught me

that I can do all things through Christ who strengthens me.
ACKNOWLEDGEMENTS

I appreciate all the insightful comments and suggestions the members of my dissertation committee gave me about this work, especially my co-chairs, Lise Menn and Alice F. Healy. I am also grateful to family and friends who helped me pilot test the materials and score the data. This research was partially supported by student research grants from the Institute of Cognitive Science at the University of Colorado, Boulder.
# Table of Contents

**Chapter 1: Introduction**

- Variability in the Speech Signal ................................................................. 4
- Speech Perception: Nativist, Invariant Feature Approaches .................. 11
- Speech Perception: Direct Realist Approach ............................................. 18
- Speech Perception instead of Segment Perception .............................. 26
- Theoretical Bases for Training Conditions: Speech Perception Research..... 31
- Theoretical Bases for Training Conditions: Other Research .................. 34
- The Training Tasks ..................................................................................... 39
- Format of Training Experiments ............................................................... 42

**Chapter 2: Speech Sample Collection and Accent Rating** ...................... 46

- Interviews ........................................................................................................ 47
  - Method ........................................................................................................... 47
  - Participants ................................................................................................. 47
  - Procedure ...................................................................................................... 48
- Accent Rating ................................................................................................ 49
  - Method .......................................................................................................... 49
  - Participants ................................................................................................. 49
  - Procedure ...................................................................................................... 50
  - Results .......................................................................................................... 51
  - Discussion .................................................................................................... 55

**Chapter 3: Experiment 1—Focus of Attention** ...................................... 58
LIST OF TABLES

Table 1. Language Background Data for Interviewees (all ages given in years)......48
Table 2. Accent Rating: Balanced Latin Square for the Order of Native Language Groups.................................................................50
Table 3. Accent Rating: The Two Presentation Orders for the Speech Samples......51
Table 4. Mean Ratings for Aspects of Non-Native Pronunciation, by Native Language Group.................................................................52
Table 5. Intraclass Correlations of Raters for Pronunciation Aspects..............53
Table 6. Adjusted p Values for Contrasts Between Mean Ratings of Language Groups..................................................................................54
Table 7. Segmental Aspects of the Accents of the Native Mandarin, Russian, Thai, and Turkish Speakers in These Speech Samples.........................56
Table 8. Experiment 1: Mean Test Scores by Condition.................................65
Table 9. Experiment 2: Training Samples Information....................................73
Table 10. Experiment 2: Raw Pretest Score Means.......................................80
Table 11. Experiment 3: Characteristics of the Speech Samples Used for Comprehension Measures.................................................................92
Table 12. Experiment 3: Raw Pretest Score Means.......................................96
Table 13. Experiment 4: Native English Mean Test Scores for Each Version of the Test.................................................................111
Table 14. Recall Test Means for Experiment 4 (Native) and Experiment 3 Pretest (Accented Pretest) for Each Test Version.................................117
LIST OF FIGURES

Figure 1. Experiment 2: Z-score means for written recall and multiple-choice comprehension measures on the pre- and posttests across conditions.........................................................82

Figure 2. Experiment 2: Z-score means for the transcription measure on the pre- and posttests across conditions.................................................................83

Figure 3. Experiment 2: Z-score means for the accent discrimination measure on the pre- and posttests across conditions.........................................................84

Figure 4. Experiment 3: Z-score means for the multiple-choice comprehension measure on the pre-, post-, and retention tests across conditions, with scores on the Turkish-accented tests separate from scores on the Thai-accented tests.................................................................97

Figure 5. Experiment 3: Z-score means for the recall comprehension measure on the pre-, post-, and retention tests across conditions, with scores on the Turkish-accented tests separate from scores on the Thai-accented tests............98

Figure 6. Experiment 3: Z-score means for the transcription section on the pre-, post-, and retention tests across conditions, with scores on the Turkish-accented tests separate from scores on the Thai-accented tests...............99

Figure 7. Experiment 3: Z-score means for the accent discrimination section on the pre-, post-, and retention tests across conditions, with scores on the Turkish-accented tests separate from scores on the Thai-accented tests.....100
Figure 8. Experiments 2 and 4 comparison: Multiple-choice scores for test versions A and B, according to the group of the participants (native, preA, preB).

Figure 9. Experiments 2 and 4 comparison: Recall scores for test versions A and B, according to the group of the participants (native, preA, preB).

Figure 10. Multiple-choice test scores for Experiment 4 (native), Experiment 3 pretest (accented pre), and Experiment 3 retention test (accented retention), according to the accent type of the original test speaker.

Figure 11. Transcription scores for all test versions according to the group of the participants in Experiment 4 (native), or in Experiment 3 for the pretest (accented pre) and the retention test (accented retention).
CHAPTER 1: 
INTRODUCTION

Foreign-accented speech can cause problems not only for international scholars and businesspeople, but also for the native English speakers with whom they interact. One international journalist recalls, “I once worked with a Chinese fellow in England who when things went wrong would mutter darkly, ‘Bruddy haiko!’ which I took to be some ancient Cantonese invective; it was not until many months later that I realized he was just saying, ‘Bloody hell’” (Bryson, 1990, p. 84). Students at American universities tend to be less amused and more frustrated by their interactions with non-native English speakers who are their instructors and teaching assistants; they avoid classes taught by non-native speakers, and when confronted with foreign-looking or sounding instructors, students tend to rate the instructors’ teaching ability harshly (Rubin, 1992; Rubin & Smith, 1990).

One approach to resolving, or at least mitigating, the communication difficulty between native and non-native speakers is to help the non-native speakers reduce or eliminate their accent. Pronunciation programs and classes are available for this purpose, but their effectiveness is not guaranteed (Derwing, Munro, & Weibe, 1998; Johnson, 2000). A number of factors influence the degree of non-native speakers’ accents, including the age at which they acquired their second language (Bongaerts, van Summeren, Planken, & Schils, 1997; Flege, Yeni-Komshian, & Liu, 1999; Munro, Derwing, & Flege, 1999), the amount of continued use of the first language (Flege, Frieda, & Nozawa, 1997; Piske, 2001), and the pronunciation proficiency—or
lack of foreign accent—in the first language (Yeni-Komshian, Flege, & Liu, 2000). In other words, the younger people are when they learn their second language, the more likely they will be to speak that language with little or no foreign accent. Similarly, if people do not use their first language regularly, they will be more likely to speak the second language without an accent. However, bilingual speakers who speak their first language fluently with no trace of a foreign accent from the interference of their second language will probably have a noticeable accent in their second language: It is unusual to speak two or more languages with native-like pronunciation, as there is usually a dominant language whose pronunciation patterns influence the pronunciation patterns in the other languages the speaker uses. Although speakers have some control over how much they continue to use their first language and could minimize its influence over their pronunciation of the second language by choosing not to use the first language on a regular basis, speakers cannot turn back the clock and begin learning the second language at an earlier age. Other qualities that affect proficiency in a second language, such as language aptitude and willingness to take risks (Harley & Hart, 1997; Lightbown & Spada, 1999), also remain largely beyond the speakers’ control. Therefore, even though helping non-native speakers reduce their accent may be an important part of fostering better communication between native and non-native speakers, eliminating an accent entirely seems difficult, if not impossible, for many speakers.

Because speech communication entails both production and perception of speech, another way to tackle the breakdown in communication that can occur as a result of a foreign accent is to deal with the perception end of the process, or the
native listener. How listeners perceive speech, extracting meaning from the continuously changing input of the acoustic signal, remains a much-debated matter, in large part because of the variability in how phonemes and words are realized at any given point by any given speaker. Listeners are generally unaware of this variability, and somehow they compensate for it with little effort as they listen to and understand speech. However, foreign-accented speech presents challenges for speech perception that unaccented speech does not. Listeners are aware of the shifts in pronunciation that accents cause, and they often have to expend conscious effort to compensate for these shifts. The increased processing load created by this challenge can lead to negative affective responses from the listeners, who then become less willing to put forth the effort necessary to foster effective communication. The listeners’ attitude toward the speaker has received a good deal of attention, and some work on designing a program to help native English listeners become better at understanding accented speech has incorporated cultural awareness as part of the training (Derwing, Rossiter, & Munro, 2002). Nevertheless, a key piece of a training program for native listeners must be exposure to accented speech and practice comprehending it, for listeners who have developed strategies to deal with the shifts in speech patterns found in foreign-accented speech can understand non-native speakers more easily and accurately, and with less frustration.

The research presented here explores training methods intended to improve listeners’ comprehension of accented speech. For training, listeners heard speech samples of non-native speakers and were asked to perform tasks that encouraged them to concentrate on either the content of the speech or the accented pronunciation
of the words. The primary goal of this work was to determine whether a short, intense training period could improve listeners' comprehension significantly; secondary objectives included observing the durability and transferability of this training and gaining insights into how listeners benefit from paying attention to the pronunciation or meaning of the speech.

Variability in the Speech Signal

Under normal circumstances, the perception of speech is rather amazing; with the additional burden of a foreign accent, the fact that many listeners manage to understand the speaker's message is almost miraculous. Speech perception is the process by which listeners decode spoken messages; it is the human ability to hear the sounds created by articulatory gestures and to understand the message conveyed by these sounds. Decoding the spoken message might appear to be a simple matter of identifying the sections of the speech signal that correspond to given phonemes or words, but variability makes the equation far from straightforward. Phonemes and words are not uttered exactly the same way each time they are spoken, due in part to coarticulatory effects, rate changes, differences in the vocal tracts of speakers, and dialectal or foreign accents. Coarticulation results from overlapping articulatory gestures that cause adjacent segments of speech to influence the articulation of a given segment. More detail on coarticulation, rate changes, and vocal tract differences is given below; unlike these processes that result in variability, which are "normal" elements of speech production and perception, accent marks speech as different. Accent is the deviation of pronunciation patterns in a certain group's
speech from standard pronunciation. Although coming to a consensus about what standard pronunciation is can be difficult (Lippi-Green, 1997), the foreign accents discussed in this dissertation can be defined as resulting from the influence of a speaker's native language on his or her pronunciation patterns in another language. In addition to pronunciation deviations, foreign-accented speech is often distinguished by grammatical deviations as well. Thus, listeners must cope with both pronunciation and grammatical differences in order to understand foreign-accented speech.

Variability abounds in speech, and as a consequence, abstract linguistic units such as phonemes or words do not consistently match exact segments of the speech signal. The acoustic realization of and, for example, may be a canonical combination of the phonemes /æ/, /n/, and /d/, but this canonical pronunciation of each phoneme generally occurs only when the word is said in isolation or when special emphasis is placed on the word; in casual speech, and is more often realized with a reduced or deleted vowel, an unreleased or deleted /d/, or any number of other variations. Yet, the fluent English-speaking listener generally hears and, regardless of how badly the word conforms to the canonical pronunciation. Such acoustic variation results from two primary factors: coarticulation and rate of speech. In general, both coarticulation and speech rate are normal phenomena in normal speech production, and listeners do not consciously attend to these sources of variability unless they are somehow extreme—for example, excessive coarticulation or a rate of speech that is tediously slow.
Because the position of the articulators constantly changes during speech production, their position at any given moment is affected not only by their previous position, but also by their next position. As a result, Handel (1989, p. 147) points out, "With rare exceptions, speech is necessarily coarticulated." One prominent realization of coarticulation is allophonic variation. The stop /t/, for example, has several allophonic variants, including the tap, glottal stop, and aspirated, unaspirated and unreleased forms. In the sentence *I stopped by Patty's house to borrow a button for my coat*, segmental and suprasegmental processes cause the allophonic variation of /t/. At the segmental level, the /t/ in *stopped* is unaspirated because it follows /s/, the glottal stop in *button* comes before a syllabic nasal, and the aspirated /t/ of *to* is syllable initial. Suprasegmental processes are evidenced by the tap in *Patty's*, which occurs between a stressed and an unstressed vowel, and by the unreleased /t/ at the end of *coat*, which is typical of stops in clause final position.

Non-stop consonants are also affected by coarticulation. Ladefoged (1993) enumerates a number of rules for English allophones. Word-final devoicing, for example, is common for voiced fricatives and stops, such as for the /z/ at the end of *dishes*. The lateral /l/ can be a dark, velarized lateral, as in *bell*, /bɛl/, or a bright alveolar lateral, as in *led*, /led/. Approximants, including laterals, may be devoiced after a voiceless, aspirated stop, resulting in *play*, /pʰɛɪl/, and *twin*, /tʰɪn/. Liquids can become syllabic at the end of a word, immediately after an obstruent, as in *tailor*, /ˈteɪlər/. The environments in which certain allophones occur are predictable, as evidenced by allophonic rules that state the contexts in which certain allophones can appear. Listeners also learn to predict the variation, but their awareness, usually
unconscious, of the patterns does not render the variation trivial or irrelevant for speech perception theories. These allophonic processes do not always apply, hence the use of *can* and *may* in stating the rules; their application depends on rate of speech and the individual speaker, and this optionality serves to introduce even more variability into speech (Allen & Miller, 2001).

Compared with consonants, vowels are relatively long segments of the acoustic signal, but they also show substantial variation in speech. The variability of vowels has several sources. For instance, different-sized vocal tracts result in differences in the positions of vowel formants. Formants are overtones that correspond to resonating frequencies, or pitches; the first three formants, F1, F2, and F3, define the harmonics of vowels and give them their characteristic qualities. Although formant frequencies define each vowel, there are differences in these frequencies among speakers. Men naturally have lower-pitched voices than do women, and a typical fundamental frequency, or F0, for men is about 135 Hz, but it is 185 Hz for women (Strange, 1999). The higher fundamental frequency of women means that the harmonics of their vowels are also shifted up so that the formants defining their vowels are higher than the formants of men’s vowels. Speaker normalization is thought to enable listeners to correct for frequency differences and hear consistent vowels across a variety of speakers. However, it is not clear whether such normalization is based on the ratio between F0 and F1, or if higher-order variables, such as patterns of spectral change, may be involved in speaker normalization (Nygaard & Pisoni, 1995; Strange, 1999). In any case, speaker variability introduces inconsistencies in the formant patterns of vowels.
Another source of variability for vowels is coarticulation. In consonant-vowel-consonant syllables, formant patterns often do not reach the steady-state target; for instance, the /e/ in bet has a lower F2 than the /e/ in get because the labial consonant prevents the formant from reaching its normal height (Strange, 1999). Coarticulation can also affect the duration of vowels. English vowels before voiceless stops tend to be shorter than those before voiced stops, so that the /a/ in mop is shorter than that in mob. Vowels in open syllables, such as both /a/s in mama, are longer still, especially in word final position.

Finally, a faster speaking rate may not only alter the quality of vowels, but may delete vowels, or consonants, entirely. Short function words are especially vulnerable to this type of alteration, so that and may be represented by a syllabic nasal in the phrase bread 'n butter, and entire phrases may be rendered with considerable reduction if they are familiar, as in the informal greeting, /'hei 'hæ-jo 'dun/, Hey! How are you doing? Listeners are sensitive to changes in the rate of speech, and the compensatory processes used to decode speech at different rates demand increased attention and time (Nygaard & Pisoni, 1995).

In contrast to the “normal” sources of acoustic variation, the variation caused by accent is, indeed, noticed by listeners. Non-native accents are marked by certain vowels, consonants, and suprasegmentals having a particular non-native quality because the speaker’s native language influences his or her English pronunciation. Characteristics of foreign accents can be defined in terms of their acoustic features. For example, one Portuguese speaker’s father charged her that her Portuguese speech was “so explosive” after spending time in the United States; in terms of acoustic
features, the American English accent that had crept into her Portuguese was a tendency to aspirate voiceless stops /p, t, k/, resulting in measurably longer voice onset times (VOTs) than are typical in Portuguese speech (Sancier & Fowler, 1997). Voice onset time is the moment at which voicing starts relative to the release of the articulators’ closure for a stop. For an English speaker to say the /p/ in pot, the lips close and then open to release the closure, and this release is followed by a puff of aspiration. Only after the aspiration do the vocal cords begin vibrating to make the vowel. The period of aspiration causes a long VOT because there is a lag between when the articulators release and when voicing starts. A Portuguese /p/ has much less aspiration or none at all, so the VOT is considerably shorter than that in English.

Other foreign accents can also be characterized by shorter VOTs, such as French and Turkish, or by longer VOTs, such as Mandarin and German (Arslan & Hansen, 1997).

Arslan and Hansen (1997) considered not only VOT in their description of Mandarin, German, and Turkish accents but also several other acoustic features, including duration of word-final stop closure, duration of voicing in vowels, and formant frequencies in vowels. They found that word-final stop closures, as for the /d/ in would, are significantly longer and that the average duration of voicing for vowels is longer for all three accent types compared to unaccented American English speech. Shifts in the formant frequencies of F2 and F3 were also observed; for the rhotic vowel /ɹ/, as in bird, F3 generally collapses down into F2, but for the non-native speakers, this did not occur, leaving the vowel to sound less r-colored.

Duration of vowels, lexical and phrasal stress, intonation contours, and changes in
intensity (dB) are only some of the other segmental and suprasegmental factors that have also been used to define accent characteristics (Arslan & Hansen, 1997; Johnson, 2000; Magen, 1998). These studies illustrate that foreign accents can be quantifiably described by shifts in pronunciation patterns.

Although naïve listeners may not be able to explain a foreign accent in terms of changes in VOT or fluctuations in vowel formants, they are certainly aware of the impact that the accent has on the speech. In fact, their judgments about the accentedness of speech are consistent enough that listener accent ratings have been used to categorize degrees of foreign accent in a number of studies (e.g., Anderson-Hsieh, Johnson, & Koehler, 1992; Flege et al., 1999; Munro & Derwing, 1994).

Further evidence of listeners’ awareness of foreign accents comes from the affective response accented speech can evoke. Foreign-accented speech can cause listeners to be irritated and to rate the speakers as having lower status: being less educated, poorer, less successful, and less intelligent; the heaviness of the accent and the number of grammatical and lexical errors contribute to the degree of listeners’ irritation and consequent poor ratings (Albrechtsen, Henriksen, & Faerch, 1980; Brennan & Brennan, 1981; Fayer & Krasinski, 1987).

A foreign accent can also cause increases in processing time as listeners work to adjust to the accent. In a study of response times to true or false statements spoken by native English speakers and native Mandarin speakers, Munro and Derwing (1995) found that when the foreign accent of the Mandarin speakers was linked to poorer comprehensibility of the speech, longer processing times were seen. Munro and Derwing point out that accentedness and comprehensibility are not necessarily linked,
and some heavily accented speech may not require more processing time if it remains comprehensible to the listeners. Only when the accent contributes to lower comprehensibility do listeners need more time to process the speech. Although this work demonstrates that accented speech may require additional processing time and effort from the listener, the question that remains is what that processing entails—how listeners perceive speech, accented or unaccented. Insights into the process of speech perception can provide some guidance for developing a program to help listeners better understand accented speech.

*Speech Perception: Nativist, Invariant Feature Approaches*

For many researchers, the theoretical examination of speech perception has been grounded in a nativist assumption of a mental inventory consisting of idealized elements. According to this view, acoustic input maps onto invariant, canonical linguistic features such as phonemes and syllables. Chomsky (1965, p. 4) articulates the nativist position as linguistic theory that “is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior.” A key piece of evidence adduced for Chomsky’s position is the modularity of the brain; among the separate subsystems of the brain is the language faculty. According to the nativist position, then, the production and perception of language is not governed by general pattern-sensitive properties of the brain, but rather by a specific set of faculties innately attuned to the patterns of language.

For speech perception, the nativist assumptions mean two things. First, the actual acoustic or articulatory information must be mapped to innate mental
categories of phonemes, syllables, or words in order to be perceived correctly.

Second, this mapping is facilitated by a specialized language module in the brain which interprets the speech signal by attuning to the relevant information in the signal that matches canonical mental representations of linguistic units. The linguistic units on which nativist approaches concentrate are usually consonants and vowels, but they may include lower-order elements such as voicing or higher-order segments such as syllables. Nativist, invariant feature approaches, then, tend to examine segment perception rather than speech perception, because speech perception encompasses understanding the meanings of words and utterances, in addition to the correct identification of segments. As a result of the emphasis on segment perception, the tenet underlying the research seems to be that bottom-up processing is the fundamental, if not the only, process involved in speech perception. Bottom-up processing is the listener’s perceptual analysis of first the phonetic level, then the semantic and syntactic levels. In contrast, top-down processing of speech involves the use of prior knowledge or linguistic expectations to facilitate word recognition and message comprehension; in this type of processing, semantic and syntactic levels of speech can influence the perception of syllables and words.

The nativist, invariant-feature approach has engendered a number of theories. Some researchers, such as Stevens and Blumstein (1981), have focused on determining the invariant acoustic features that correspond to abstract phonetic categories. Other researchers, such as Liberman (1996), have defined theories based on nativist assumptions of innate modules in the brain.
“One of the major activities in the study of speech has been the search for the acoustic correlates of perceived phonetic distinctions” (Eimas & Miller, 1981, p. 1). According to Stevens and Blumstein (1981), speech perception requires an analysis of the continuous speech signal into distinct phonetic features. The acoustic signal itself contains the properties that correspond uniquely and invariantly to phonetic features. Further, the invariance that corresponds to phonetic categories is not represented by individual components of the signal, such as onset frequencies of certain formants, but rather by integrated acoustic properties that together give the spectrum its shape at, for example, the release of a stop consonant. Because the speech perception system recognizes the invariance of integrated properties in the signal, there is a reasonably direct decoding of sound into a representation based on distinctive features. Basically, the theory posits that bottom-up processing of speech is based almost exclusively on invariance found in the acoustic signal; although Stevens and Blumstein acknowledge that secondary cues may also help listeners find phonemes in fluent speech or in degraded listening conditions, these secondary cues correspond to phonetic dimensions rather than top-down semantic or syntactic information. The theory does not broach the topic of sound-meaning correspondence in speech, instead concentrating on segment perception anchored in distinctive features.

Stevens and Blumstein’s (1981) theory assumes that underlying distinctive features in the acoustic signal, which are based on those proposed by Jakobson (Jakobson, Fant, & Halle, 1963), create a framework for natural classes and the phonology of a language. Evidence for underlying distinctive features comes from an analysis of stop consonants. Stevens and Blumstein created three templates of
synthetic speech for the spectral shapes of labial, alveolar, and velar consonants and then tested the templates against actual speech samples. The synthetic templates are based on theoretical considerations of sound production and the changes in the formant structure caused by physical constrictions in labial, alveolar, and velar places of articulation (see also Ladefoged, 1993).

Stevens and Blumstein (1981) test the appropriateness of their synthetic templates both by comparing the spectral shapes of consonants in consonant-vowel sequences uttered by several speakers to the templates they propose and by examining whether listeners are sensitive to the templates. Although they did not find complete accuracy in matching the templates to the natural data, initial stops and final bursts in all positions fit the templates about 75% to 85% of the time. Listeners also showed some sensitivity to the templates in experiments requiring that they identify synthetically manipulated consonant-vowel sequences in which the consonants gradually changed from diffuse-falling to diffuse-rising to compact. Listeners grouped the syllables into three well-defined groups of formant structures—labial, alveolar, and velar—and the exemplars that closely matched one of the three templates were consistently identified as /b/, /d/, or /g/, respectively, but those spectra whose shapes lacked the gross properties of a template elicited equivocal responses.

In addition to their detailed analysis of stop consonants, Stevens and Blumstein (1981) propose invariant properties for several other phonetic contrasts, including non-stop consonants such as fricatives and nasals. Although these phonetic contrasts are not elaborated as thoroughly as the evidence for place of articulation for stops, Stevens and Blumstein include them to demonstrate that invariant features can
be specified for a number of phonetic contrasts. The existence of such features implies that the auditory system has special feature-detecting mechanisms that extract invariant properties from the speech signal, and Stevens and Blumstein believe that these mechanisms are innate, providing the framework for the organization of language and allowing infants to detect differences between phonetic classes. This assumption of innateness is a fundamental piece of the nativist, invariant-feature view of speech perception, and it is one which the Motor Theory addresses in detail.

The most salient claim of the revised Motor Theory (Liberman & Mattingly, 1985) is that listeners perceive a speaker’s intended phonetic gestures, and therefore speech production and perception share a common processing strategy. For the Motor Theory, the invariants of speech perception are the intended phonetic gestures rather than the invariant acoustic features presented by Stevens and Blumstein (1981). Where the two theories meet, however, is in their assumption of an innate, specialized language module that mediates speech perception. The proposal of such a module is critical to the Motor Theory because it provides the mechanism by which the perception and processing of speech occur, thus establishing a foundation for the theory. Liberman and Mattingly claim that speech perception must be viewed as the specialized perception of phonetic gestures, not as a process governed by general principles that apply to the perception of sounds. This description of speech perception not only emphasizes the role of a specialized language module, or “Black Box,” but also reveals the scope of speech addressed by the theory. The narrow scope of speech perception—perception of phonetic gestures—fits better into the category of *segment perception* than that of *speech perception*. In addition, limiting
the scope in this way largely precludes consideration of the effects of top-down processing.

The specialization of segment perception is essential to the Motor Theory because Liberman and Mattingly (1985) assume that the invariants of segment perception are intended phonetic gestures, and this assumption fits neatly with a modular account of perception. This assumption does not fit well with a non-modular cognitive account. For a cognitive account, the gestures would need to be inferred from the acoustic signal by a general cognitive process, something that Liberman and Mattingly admit would be far-fetched. "It is thus quite reasonable," they explain, "for proponents of a cognitive account to reject the possibility that invariants are motoric and to insist that they are to be found at or near the auditory surface" (p. 29).

However, a modular account avoids such difficulties. Because computing invariant phonetic gestures from the acoustic signal is a task without parallel among other cognitive processes, it is reasonable to posit a special module in the brain designed for the task. Furthermore, if the invariants of segment perception are intended phonetic gestures, the case for a separate linguistic module is more compelling because the domain of speech (or rather, segment) perception is "suitably eccentric." Because the domain of the module is eccentric, it cannot fit within a more general domain of another module; therefore, proposing a separate linguistic module is justified, and indeed, necessary.

In addition to the requirement of an eccentric domain for a module, Liberman and Mattingly (1985) define other characteristics of modules, following Fodor (1983). A module is a piece of specialized neural architecture designed to perform domain-
specific computations. These computations are not cognitive, and therefore are not available to conscious awareness because they are carried out within the module. Given that the module is specialized, it processes and outputs only narrowly definable, domain-relevant information. Operation of the module is not a voluntary choice because of the domain’s ecological importance for the organism.

The characteristics of the perception of segmental information of speech fit the modularity criteria in several ways, according to Liberman and Mattingly (1985). Segment perception involves the processing of relevant information, which is usually acoustic but can also be optical, as the McGurk effect demonstrates (McGurk & MacDonald, 1976; see the section on Direct Realism for a detailed discussion). Irrelevant information in the stimulus is not used, which would be expected given a specialized linguistic module. Such irrelevant stimuli include information about voice quality, which some other module might process, and “extraphonetic” information that makes speech sound natural but is not necessary to comprehension, as evidenced by work with synthetic speech. Noise and distortion are also excluded from processing in the linguistic module: Although they are perceived through other modules, they interfere little with speech perception. Thus, phonetic perception occurs in a specialized module that can operate in competition with a general auditory module. This innate module holds the key to deciphering the invariant phonetic gestures that allow the segmental information of speech to be decoded.

Experimental evidence for existence of this specialized module comes from duplex perception experiments by Liberman and his colleagues (Liberman, 1996). These experiments presented listeners with synthetic speech consisting of syllables
such as /da/ and /go/. The syllables were divided into the “consonant,” an isolated synthetic formant transition, and the “vowel,” a base formant structure that sounded like a vowel. For the listeners, the isolated formant transition was presented in one ear and the base stimulus in the other. What listeners heard was two sounds simultaneously: a nonspeech “chirp” in the ear receiving the isolated formant transition, and the complete syllable in the other ear. In other words, the isolated formant transition is heard in two different ways, as a consonant in a coherent syllable and as a chirp on its own. This phenomenon violates the psychoacoustic principle that listeners can only assign a given element in a sound to one source at a time (Hawkins, 1999). Assuming that this principle operates within modules but not between them, duplex perception demonstrates that speech is processed in a separate module from nonspeech because of “the simultaneous operation of phonetic and auditory modes” (Liberman, 1996, p. 385). Thus, the proposition of a linguistic module, which is theoretically critical to the Motor Theory, may also be justified empirically.

*Speech Perception: Direct Realist Approach*

Similar to nativist, invariant-feature theories, a direct realist approach to speech perception leans heavily on segment perception, and the research used to support its claims are studies of isolated phonemes and syllables. In consequence, much of the direct realist discussion hinges on bottom-up processing of *speech events*, the phonetically structured articulations of a speaker, and how their phonetic structure is perceived by a listener (Fowler, 1986). This discussion falls under my
category of segment perception. However, direct realists acknowledge the representational nature of speech, giving some tentative comments about linguistic events, which are the communicative content of utterances, and the role that top-down processing may play in their perception (Fowler, 1986, 1996). In this way, direct realists touch on speech perception, according to the definition of speech perception as the process of listening to sounds created by articulatory gestures and understanding the message they convey.

The chief claim of Direct Realism is that perception is direct and unmediated by interpretive processes; therefore, the central principle regarding speech perception is that the segmental information of speech is perceived directly from articulatory gestures. Indirect interpretation of segments, through cognitive inference or innate knowledge, implies that sensory data are impoverished or ambiguous, and direct realists reject that implication as untenable (Best, 1995). The direct realist speech perception theory originated from a general perception theory, and it protests the major assertions of the nativist, invariant-feature approach: first, that the acoustic signal contains invariants, and second, that a separate linguistic module is necessary for speech perception.

The direct realist theory of perception, or the ecological theory, was first proposed by Elenor and James Gibson to explain visual perception (see J.J. Gibson, 1979; E.J. Gibson, 1991). Perception is an act of attention, an awareness of the environment and oneself in it, according to J.J. Gibson (1979). For the ecological theory, perception involves three steps or aspects: To begin, there is a distal event, which is some tangible object or occurrence that exists in the environment; next, there
is an informational medium, which provides information about the event to perceivers by stimulating their sense organs; finally, the perceivers actively seek out information that is relevant to their interests and needs. For visual perception, the distal event is an object or image, such as a tree or a drawing. Light reflected off the object is the informational medium that stimulates the eye and provides the eye with information about the properties of the object. The perceiver looks at objects and learns to attend selectively to those aspects of the object that are relevant to his interests. Thus, light is merely an informational medium conveying the properties of the actual distal event, the tree.

Speech perception, like perception in general, consists of events, informational media and active perceivers, according to direct realists. The events of direct perception in speech are the articulatory gestures, and the informational medium is the acoustic speech signal. A gesture is actually a family of movement patterns, such that “phonetic segments are realized as coordinated gestures of vocal-tract structures;” gestures are not the movements of individual articulators or muscles (Fowler, 1986, p. 5). The articulatory gestures cause sound waves that reach the ear as the acoustic signal, but as the light reflected off of a chair is not the chair itself, so the acoustic signal is not the articulatory gestures. Rather, the acoustic signal is an information medium in that it acquires its structure from an event, the activities of the vocal tract, and transmits this structure to the auditory perceptual system, thereby conveying information to a sensitive perceiver. Therefore, as the chair is the real object of visual perception, the articulatory gestures are the real object of speech perception. No one, Fowler asserts, “thinks that, if the objects of visual perception—
that is, trees, tables, people, etc.—do have physical properties, their properties are those of reflected light” (p. 6). By analogy, then, the physical properties of sounds are not those of the acoustic signal, but rather the physical causes of the noises. By describing speech perception in terms of events and informational media, direct realists achieve a tidy parallel between speech perception and other types of perception. A further consequence of this description is that segment recognition, facilitated by bottom-up processing, is fundamental to the discussion of speech perception. As a result, empirical support for Direct Realism comes from studies of segment perception, and direct realists draw both from their own research and from others’.

One piece of evidence Fowler (1986, 1996) invokes to support Direct Realism is the McGurk effect. In McGurk and MacDonald’s (1976) study, children and adult subjects watched a film of a woman saying the syllables /ba/, /ga/, /pa/, and /ka/, but the acoustic signal for these syllables was dubbed to the lip movements of non-corresponding syllables, so that the acoustic signal of /ga/ was paired with the lip movements of /ba/, and the acoustic signal of /ba/ was dubbed with the lip movements of /ga/, and so on. Listeners were able to repeat what they heard the model saying with over 90% accuracy under auditory conditions only, when they were not looking at the video. However, under visual-auditory conditions, error rate was above 50% for children, and 92% for adults. Adults are more strongly influenced by visual information, consistently hearing, for example, /da/ when viewing /ga/ lip movements paired with /ba/ voicing. The McGurk effect clearly indicates that visual
information plays a role in segment perception, and that segment perception does not depend exclusively on the acoustic signal, at least when visual information is available. From a direct realist perspective, the McGurk effect provides evidence that the acoustic signal cannot be the event of segment (or speech) perception, but is simply an informational medium pointing to the event, the articulatory gestures. Listeners may gather information about the articulatory gestures from both visual and auditory sources, and because their perception is influenced by the lip movements they see a speaker making as well as the sounds they hear, the endpoint of speech perception must not be the acoustic signal, but rather the articulatory gestures.

Direct Realism’s account of coarticulation essentially eliminates the problem of variance by invoking the notions of coproduction and direct perception. The account of coarticulation neatly fits the direct realists’ three aspects of perception—events, informational media, and active perceivers (Fowler, 1984, 1986, 1994). First, the event is an articulatory gesture; adjacent gestures are coproduced, overlapping one another. A given gesture starts as a rather weak co-occurrence with another gesture, then becomes the main information of the next time segment, and finally becomes weaker again as another gesture takes prominence. Second, the acoustic signal provides information about the gestures, indicating both where articulatory overlap occurs and which gesture is waxing or waning. The active perceiver is the final of the three aspects of direct perception. Listeners are able to disentangle the different influences from the speech signal, treating certain weak information in one segment of the signal as information about the next segment rather than the current segment. Although it may seem that disentanglement of gestures through means of the speech
signal would involve mental interpretation, direct realists categorize the disentangling process as active perceiving, fitting data into a meaningful schema, rather than as mental interpretation. Innate knowledge or acquired associations, both of which entail mental interpretation, are not necessary for active perceiving, which is in this instance disentangling gestures.

Evidence that listeners factor out coarticulatory information comes from experiments about how listeners segment the speech signal (Fowler, 1984). The argument is that listeners are guided by coarticulation as they segment the speech signal; thus they perceive anticipatory coarticulation as the onset of the new segment it anticipates, and they do not integrate this information with the concurrent information for the preceding phonetic segment. Fowler tested anticipatory coarticulatory effects of vowels on consonants in an experiment involving synthetic manipulations of syllables. The stimuli included /gu/ and /gi/ utterances which were produced by a female speaker, as well as synthetic alterations of these syllables in which a period of silence separated the consonant from the vowel. In the synthetically manipulated segments, the vowel could be either the same as or different from the original vowel, so that an original /gu/ syllable could become either [/g/-silence-/u/] or [/g/-silence-/i/]. Fowler found that listeners responded more quickly and more accurately if the vowel was the same as the original one, and in an identification task in which only the consonant burst was played, listeners showed above-chance vowel identification. These results seem to confirm the hypothesis that listeners do not integrate coarticulatory information into their perception of the segment with which it co-occurs, but rather they recognize it as information about its
own segment. Based on this evidence, Fowler concludes that listeners are able to segment the speech signal in a way that recovers the articulatory gestures that the talkers produce. In a sense, listeners factor out the effects of coarticulation and are left with the invariant, easily recognizable gestures that correspond to phonetic segments.

Although much of the research and evidence invoked in support of Direct Realism has focused on speech perception at the segmental level, its proponents have not avoided consideration of the higher linguistic levels involved in speech. Fowler (1986) gives some tentative comments about how Direct Realism can handle the linguistic encoding of speech which allows a speaker to convey not just phonemes but a message, though no experimental evidence is provided. Perception of the linguistic message is not exactly direct, but Fowler is reluctant to call it indirect. The indirectness of the message would derive from the fact that a listener directly perceives a talker’s vocal-tract activity, but that activity creates words specifying the linguistic content of the utterance. From this perspective, the listener directly perceives only the articulatory gestures, not the linguistic content. However, Fowler suggests an alternative perspective by drawing a parallel between perception of phonemes and perception of semantic content. The scenario goes: If I look at a tree, I see the tree. If I hear a /t/, I perceive the articulatory gestures that created that /t/. However, if you tell me about a tree, what I perceive is selected information about the tree. The selected information serves as the event that is perceived, and it is, by its very nature, secondhand information. The linguistic message, whatever words and sentences you tell me about the tree, serves as an informational medium. I, an active
perceiver, extract from the linguistic message the selected information you have pre-packaged for me. The three-part perception process thus entails no indirectness, although the event—selected information about a tree—is rather more nebulous than a tree or even an articulatory gesture because it exists only in your mind. Furthermore, how you present your information about the tree is governed by social rules instead of physical laws, so your choices of words and grammatical structures may not adequately convey your selected information to me. Fowler concedes that listeners may need to resort to "extraperceptual" guesses in communication, knowing that the talker is not an entirely reliable source of information. By this point, Fowler has described a semantic perception process that seems not only parallel to segmental perception, but independent of it as well. How the two processes of semantic perception and segmental perception combine to allow speech perception remains largely unexplained.

The assumption of direct realists is that "the flow of stimulation provides a rich and reliable source of direct information," which does not need to be interpreted through inferential processes or mappings to invariant units (Best, 1995, p. 175). Although the theory focuses on perception of the segments of speech, some preliminary remarks on speech perception as language perception are included as well. Direct Realism also incorporates the role of learning, in that perceivers actively engaging in exploration of the world gain "increased attunement" to information in the stimulus, which allows them to notice patterns and accurately segment the articulatory gestures. Little more is said about this learning, except that it occurs through general cognitive processes, not through specialized mechanisms. Although
Fowler does not attempt to substantiate the direct realist claim about learning with experimental evidence, a classic study by Gass and Varonis (1984) suggests that listeners’ increased attunement to the patterns of foreign-accented speech can facilitate comprehension. The main finding of the Gass and Varonis study is that familiarity with the topic of speech significantly aided comprehension, but they note trends in three other types of familiarity, including familiarity with nonnative speech, with a particular nonnative accent, and with a particular nonnative speaker, which “all have a facilitating effect on comprehension” (p. 65). They corroborate these claims by comparing comprehension scores of English as a Second Language (ESL) teachers to scores of naïve listeners, and in all cases, “there were fewer errors made by the ESL teachers, who had many years of experience working with nonnative speakers, than there were by the naïve listeners, whose experience with nonnative speakers was more limited” (p. 79). This study, then, demonstrates that listeners seem to learn to notice pronunciation patterns of foreign-accented speech, which in turn enables them to identify more accurately segments and words, and thereby comprehend the message better. Consequently, Gass and Varonis offer some support to the perceptual learning that direct realists posit but do not substantiate.

*Speech Perception instead of Segment Perception*

Segment perception, rather than speech perception, is the focus of the two theoretical approaches to speech perception that have been outlined. These theories concentrate on explaining the perception of phonemes despite low-level coarticulatory influences that modify some aspect of a phoneme, such as formant
frequencies or length of closure for a stop. Coarticulation that causes vowel reduction, syllable deletion, or phoneme substitution lies beyond the scope of segment-oriented theoretical approaches. These more severe cases of coarticulation beg the question of whether the listener actually perceives phonemes in fast, highly coarticulated speech. Furthermore, if the listener does not hear phonemes as the primary information for perception, how do top-down processes clue the listener in to the message? In addition, accented speech renders consonants and vowels differently from unaccented speech due to patterned shifts in pronunciation which presumably affect, at least in some instances, invariant information listeners might recover from the speech signal. Why, then, do the phonemes and the message of accented speech not become completely obscured for most listeners? Instead, most listeners notice the differences and form judgments about the speaker because of them, all the while largely understanding what the speaker is saying. The phenomena of accented speech and coarticulation both suggest that perceiving speech as language and perceiving the segments of speech are not necessarily the same activity. Although listeners are aware to some extent of the segments of speech, they do not depend exclusively on the identification of segments for extracting a message from the speech signal.

At issue is not whether listeners are aware of segmental information, but rather whether segmental information is the primary source needed for speech perception. Listeners are sensitive to features, phonemes and syllables; that such units exist for users of language and not just as theoretical constructs is attested by speech errors, in which segments shift position or are produced incorrectly (Fowler, 1995; Fromkin & Bernstein Ratner, 1998). Speech errors occur at several levels,
including: featural—big and fat rendered pig and vat; phonemic—brake fluid rendered blake fruid and fill the pool rendered fool the pil, and syllabic—Stockwell and Schacter rendered Schachwell and Stockter (Fromkin & Bernstein Ratner). The evidence of speech errors indicates that even though listeners may not consciously notice the lower-level elements of speech, they are nonetheless aware of them.

However, awareness of these elements in speech does not necessarily mean that they are essential to speech perception. Research on temporal induction and the perceptual restoration of missing or masked speech sounds demonstrates that listeners are usually able to understand the message of speech even when some of the speech segments are occluded or absent. Temporal induction occurs when listeners restore a fragment of the acoustic signal which has been obliterated by noise; this restoration can occur for both speech and nonspeech sounds (Bashford, Riener, & Warren, 1992). Numerous studies by Warren and his colleagues have found that temporal induction is quite robust when the missing phoneme or syllable is embedded in the context of a sentence or longer discourse (e.g., Bashford, Riener, & Warren, 1992, Bashford & Warren, 1987; Warren, 1970, 1999; Warren & Obusek, 1971; Warren & Sherman, 1974; Warren & Warren, 1970).

Warren’s early work established that listeners not only perceived phonemes replaced with a cough or other noise, but they were unable to locate the precise position of the noise in the sentence. In two studies (Warren, 1970; Warren & Obusek, 1971), listeners heard the sentence The state governors met with their respective legislatures convening in the capital city. For the experiments reported in Warren (1970), when the first s in the word legislatures was replaced with a cough or
a tone, none of the participants either realized that a phoneme was missing or was able to identify the position of the tone or cough correctly. In fact, the “illusory perception of the absent phoneme was in keeping with the observations of others (graduate students and staff), who, despite knowledge of the actual stimulus, still perceived the missing phoneme as distinctly as the clearly pronounced sounds actually present” (Warren, p. 167). Warren and Obusek (1971) presented the same sentence about legislatures to participants four times, eliciting judgments for each presentation about whether or not a phoneme was replaced by a noise and where the noise occurred. Yet, even with four chances to listen to the sentence and decide if some noise, such as a cough, buzz, or click, replaced a phoneme, participants still incorrectly stated that no phoneme was missing between 70% to 100% of the time by the fourth judgment, depending on which noise replaced the $s$. Furthermore, participants were unable to accurately locate the position of the noise in the sentence, even by the fourth judgment. However, when a short silence replaced the missing $s$, participants accurately located the position of the silence for all but the first judgment; in addition, by the second and succeeding judgments, 65% to 70% of them correctly recognized that a phoneme had been deleted. Warren and Obusek conclude that phoneme restoration is essential for comprehension in noisy listening conditions, which are typical in everyday life. Listeners must be able to extract meaning despite ambient sounds that mask or obliterate some speech sounds. Phoneme restoration, then, may not be just an auditory illusion, but a well-practiced skill listeners must use regularly.
Subsequent work by Warren and his colleagues has sought to determine the scope of phoneme restoration. They have found that silent gaps in a spoken utterance disrupt the perceived continuity of the discourse more than equivalent segments of noise, and also that listeners tolerate much longer interruptions of noise (up to 300 ms) than they do of silence (50 ms) (Bashford & Warren, 1987). However, this increased intelligibility and perceived continuity for noisy interruptions compared to silent interruptions does not hold for word lists: Bashford et al. (1992) found no significant differences between silent and noisy interruptions for lists of monosyllabic words. Based on these results, a two-step process of perceptual synthesis is proposed. First, a basic form of induction causes both speech and nonspeech sounds to appear continuous when interrupted briefly, and the interruptions are undetectable. Second, a knowledge-driven form of induction applies linguistic rules to speech so that the listener can interpolate the missing speech segments and discern the message.

This type of interpolation can be seen in experiments in which missing phonemes create a semantic ambiguity that can only be resolved by contextual information. In one such study reported in Warren and Warren (1970), listeners heard sentences such as:

A. It was found that the [cough]eel was on the orange.
B. It was found that the [cough]eel was on the table.
C. It was found that the [cough]eel was on the axle.

Although the surrounding segments were identical for the eel in all sentences, listeners reported hearing peel, meal, and wheel according to context established by the last word of the sentence. Similar studies mentioned in Warren (1970) and
Warren and Obusek (1971) confirm that words interpolated from ambiguous stimuli can seem quite clear to listeners, leading them to believe they plainly heard the word appropriate to the context. The illusory percepts caused by temporal induction suggest that top-down processing strategies are a common and natural part of speech perception and that accurate recognition of each phoneme or each syllable in an utterance is not essential for speech perception. Top-down processing may be as automatic as bottom-up processing, unnoticed by the listener but nevertheless continually contributing to the listener's perception of speech. The meaning and the sounds of speech work together to allow perception of speech. Speech perception, then, cannot be distilled to segment perception in its essence, nor to meaning perception alone: It is both.

Theoretical Bases for Training Conditions: Speech Perception Research

Several implications for the development of a foreign-accented speech training program can be drawn from research on speech perception. First, speech perception research justifies the endeavor of developing a training program: Direct realists hypothesize that learning is an inherent part of speech perception, and Gass and Varonis (1984) supply empirical support that learning and familiarization do occur and, further, that they allow better comprehension of accented speech. Other studies also show adaptation to accented speech which improves understanding of the speech (Clarke, 2002; Derwing et al., 2002; Weil, 2001). Therefore, a training program designed to familiarize listeners with foreign-accented speech is not merely
possible, it is also likely to be efficacious in enabling listeners to understand accented speech better.

Second, the evidence that speech perception involves top-down processing suggests that training should include casual, spontaneous speech rather than isolated syllables or words. The importance of top-down processing does not preclude segmental information as being necessary for speech perception, as such information clearly provides the building blocks for encoding a linguistic message in the sounds of the acoustic signal. However, coarticulation which deletes or greatly modifies phonemes and syllables is rampant in casual speech and would seem to require top-down processing to enable comprehension, much like the missing or distorted segments in the studies of Warren and his colleagues (e.g., Bashford & Warren, 1987; Warren, 1970; Warren & Obusek, 1971; Warren & Warren, 1970) which were filled in without conscious effort by the listener. Additionally, foreign-accented speech often contains grammatical deviances from unaccented speech as well as pronunciation differences, so the use of spontaneous utterances in examining accented speech perception becomes even more necessary in an attempt to remain true to what listeners typically experience with accented speech. The spoken data used for training in the experiments proposed here were elicited in 10-15 min informal interviews with non-native speakers about their work and their native culture. For the experiments, 2-3 min samples have been extracted from these interviews; as a result, the samples used for training and testing are extemporaneous examples of the type of discourse one might actually encounter when talking with a non-native English speaker.
Third, the theoretical emphasis on segment perception, which contrasts with the types of processes evident in speech perception when longer chunks of data are involved, suggests relevant contrasts for training conditions. Because research has examined primarily the perception of isolated phonemes and syllables, the implication is that bottom-up processes based on the recognition of low-level linguistic units such as features and phonemes drive speech perception. A training condition that focuses participants' attention on the segments of speech rather than on its meaning should be very beneficial. According to the implication of speech perception research and theories, such training should enable listeners to perceive the speech more easily and accurately. By training on the segmental aspects of speech, these listeners should more quickly recognize the shifted phonemes of foreign-accented speech, and because these phonemes are the building blocks of the higher linguistic levels, their improved segment perception should result in improved speech perception and comprehension. On the other hand, the findings of Warren and his colleagues (e.g., Bashford et al., 1992, Bashford & Warren, 1987; Warren, 1970, 1999; Warren & Obusek, 1971; Warren & Sherman, 1974; Warren & Warren, 1970) imply that top-down processes can play an important role in speech perception. The Gass and Varonis (1984) study provides evidence that familiarity with the topic of speech, or the higher-level semantic information, allows better comprehension of foreign-accented speech. Therefore, a training condition that prompts listeners to concentrate on the meaning of speech rather than on the details of its accentedness should help them to understand the speech better. Training on the content of the speech should allow listeners to practice top-down listening strategies, developing
their ability to fill in missing or garbled low-level information more quickly and automatically in order to understand the message with less difficulty.

Theoretical Bases for Training Conditions: Other Research

Because speech perception can be identified as an automatic process, giving listeners facts and information about accented speech is unlikely to improve their ability to perceive this type of speech. To define an automatic process, Schneider and Shiffrin (1977) first explain that long-term memory stores learned sequences of information which can be accessed either through controlled processes or through automatic processes that make few demands on short-term memory. The learned sequences of information may be associative connections, programs for responses, or directions for information processing, but they act as a unit in that all elements are activated when any element is activated. An automatic process, then, is the activation of a learned sequence which always, or nearly always, results in response to particular input, and which does not need active control or attention. Schneider and Shiffrin note that because automatic processes are based on relatively permanent learned sequences, “an automatic process is difficult to suppress, to modify, or to ignore” (p. 2). Conversely, controlled processes are activated through a participant’s control and attention, and they use short-term memory; only one or at most two controlled processes may occur simultaneously due to the capacity limitations of short-term memory.

Speech perception seems to fit the description of automatic processes well. The process of speech perception nearly always results in response to input of the
speech of a language the listener knows, and it would be difficult to modify the content one hears by forcing a reparsing of the acoustic signal and chunking syllables or phonemes differently. Of course, occasionally one may mishear an utterance and misparse it, as a child who interprets a line from “Silent Night” as “Round John Virgin, mother, and child” instead of “Round yon virgin mother and child.” After such a misparsing, listeners may recognize that the utterance is nonsensical and quickly reparse it in a way that makes more sense. Nevertheless, such appearances of chubby Johns instead of distant virgins occur infrequently, so this type of reparsing is relatively rare. Another characteristic of automatic processes is that they do not require active control. Neither does speech perception, as listeners are able to handle over 10 phonemes each second, or 135 words each minute (Warren, 1999), and still eat dinner while socializing. Cleeremans and Jimenez (2002) clarify that automatic behavior is not completely unavailable to consciousness, but that something which does not normally require active control, such as walking, can be raised to conscious awareness when interrupted, as walking would be interrupted if I tripped and stumbled. Similarly, speech perception does not usually require conscious attention, but if something interrupts or changes the acoustic signal, listeners may become more aware of the process and find themselves straining to hear and understand. Such interruption could be caused by static on a cell phone, background conversations at a restaurant, or a heavy foreign accent. In general, however, speech perception is an automatic process that is not consciously attended to or controlled by the listener.

One consequence of the automaticity of speech perception is that implicit learning, rather than explicit learning, is the process by which listeners can modify
how they perceive accented speech. Reber (1989) explains that *implicit learning* is an unconscious process which yields abstract knowledge; Berry and Dienes (1993) clarify that implicit learning involves “learning about the structure of a fairly complex stimulus environment, without necessarily intending to do so, and in such a way that the resulting knowledge is difficult to express” (p. 2). Thus, *implicit knowledge* tends to be inaccessible to consciousness and not easily communicated, whereas *explicit knowledge* is accessible to consciousness able to be communicated or demonstrated on demand; *explicit learning* is “carried out by mechanisms that label the knowledge as knowledge by the very act of inducing it” (Dienes & Perner, 2002, p. 82). Implicit learning has been explored in research on artificial grammar learning (see Dienes, 1993; Reber, 1989). In these studies, participants typically memorize strings of letters, such as MTTV and VXTVRX, and then they are told that the order of the letters is constrained by a set of complex rules. The participants are not told what these rules are, but they are asked to classify novel strings of letters as grammatical or ungrammatical, which they are generally able to do above chance (60-70%). If the participants are requested to state the rules they are using to make their grammaticality decisions, they are able to articulate some rules, but these rules are inadequate to account for the participants’ levels of success, and at times the rules are inaccurate or inconsistently applied. Thus, participants are unable to completely grasp or explain the knowledge they have implicitly learned about the artificial grammar. Similarly, knowledge about accented speech tends to be largely implicit in nature. Most listeners are unable to identify the specific articulatory or acoustic
deviations that characterize a certain accent, and yet they are able to recognize such speech as accented, and sometimes they can even name what type of accent it is.

Although there has been some discussion about just how implicit the knowledge is that participants learn in artificial grammar tasks (e.g., French & Cleeremans, 2002) and what circumstances evoke the use of rules in such tasks (e.g., Healy et al., 2002), Berry and Dienes (1993) note that for artificial grammar tasks, as well as other tasks whose underlying structure is complex or not obvious, verbal instruction does not improve performance unless accompanied by an active practice phase (see also Reber, 1989). It appears that specific information, either about the nature of the task or about how to perform it, is useful only when it is combined with a phase of implicit learning in which the information must be actively employed by the participant; otherwise, performance remains largely unaffected by verbal instruction. Practice alone, however, can significantly improve performance, so although verbal instruction may be beneficial, it is not necessary. This finding has implications for the development of training tasks designed to improve listeners’ comprehension of accented speech, given that speech perception is an automatic process and that the information encoded in speech is implicit in nature. Implicit learning about accented speech would not entail presenting lengthy explanations concerning the characteristics of a particular accent, but rather would involve active participation by listeners, compelling them to notice the characteristics and use them to perform some task without necessarily having to convey what the characteristics are.
A further reason to use a training task requiring the active participation of the listener comes from research on training and skill acquisition, which has asserted that people learn best by active responding to material (Bower & Hilgard, 1981). One example of the advantage of active participation is a study by McNamara and Healy (1995) that examined *generation advantages*, the benefits of having participants generate material by producing their own responses, in skill and knowledge acquisition. In McNamara and Healy's first experiment, participants either solved easy and difficult multiplication problems or copied the answers to the problems. McNamara and Healy found a generation advantage for the difficult problems, such that the participants who solved the problems, generating their own answers, had significantly higher posttest scores and faster posttest response latencies, whereas the participants who read and copied the problems did not show improved scores. The second experiment used a nonce-word learning task in which one group of participants read and copied word-nonword pairs, whereas the other group read and copied the words but attempted to remember and write the nonword associate before checking to see the correct answer. Again, a generation advantage was seen, and the posttest scores of participants who generated an answer before viewing the correct word were higher than the posttest scores of those who had not generated answers. The generation advantage was maintained in a retention test 1 week later. These results are explained in terms of a procedural account of memory (Healy & Bourne, 1995; Healy et al., 2002), which "focuses on the generation process rather than on the nature of the items or the relationship between the items" (McNamara & Healy, p. 162). Because participants were required to generate responses, the learning process
developed cognitive procedures, and these procedures were reinstated, or duplicated, at the tests. *Cognitive procedures* are mental operations that link a stimulus to a response, and generating responses seems to promote the formation of new cognitive procedures.

For speech perception, the mental operations linking the stimulus (the acoustic speech signal) to the response (the listeners’ perception and comprehension of the signal) are largely automatic and implicit. Thus, training intended to help listeners to comprehend accented speech needs to tap implicit learning processes in which the structure of this complex system—the deviations of accented speech—can be acquired without necessarily having to be verbalized. Also important is the active nature of this process, allowing listeners to practice the skill of perceiving accented speech by generating responses. The tasks used in the training experiments presented here were developed to promote active, implicit learning as well as to test differences that arise from focusing on phonetic information, essential to bottom-up processing, or from focusing on semantic information, essential to top-down processing.

*The Training Tasks*

Two training tasks were designed: an imitation task and a paraphrase task. For the imitation task, participants heard short foreign-accented speech samples twice; the second presentation of a given sample was broken into short phrases with pauses after each phrase, and during the pauses, the participants repeated the phrase they heard, imitating the speaker’s pronunciation of the words. The imitation was intended to encourage participants to notice the accented pronunciation and to process
it in order to replicate it. The goal of training was not to teach participants to perfectly reproduce the accented speech, for research suggests that people vary greatly in their ability to authentically mimic accents (Markham, 1999) or to produce phonemes not in their native language (Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Otake & Cutler, 1999; Wang, Spence, Jongman, & Sereno, 1999). Instead, the goal was to promote awareness of the speech segments and shifts in pronunciation.

This awareness could also be fostered by directly teaching participants about the aspects of the accented speech they hear, a paradigm that has been used in a training study by Derwing et al. (2002). In that study, training occurred over eight sessions, each session 1 week apart. Testing took place 1 week before training and 1 week after. Participants in an accent training condition listened to Vietnamese-accented speech, had discussions about cross-cultural issues, and heard lectures about the characteristics of a Vietnamese accent. Participants in a familiarity condition received the same training as those in the accent training condition, except that they were not taught the characteristics of a Vietnamese accent. A control group received no training, only taking the pre- and posttests. Although participants in the accent training condition had significantly greater confidence in their ability to communicate with non-native speakers after training than the familiarity or control group did, their sentence transcription scores were not higher than those in the other two conditions. Their comprehension scores on the posttest were also not significantly higher than those in the other two conditions, according to the 3 x 2 analysis of variance used by the authors. The between-subjects factor was condition (accent instruction, familiarity, control), and the within-subjects factor was time (before or after training).
This analysis gives an omnibus test comparing the three conditions to each other. However, if the comprehension data are analyzed by comparing the difference score (posttest – pretest) of the accent training group \((M = 22)\) with the difference scores of the other groups \((M = 7\) for both), the \(F\) would double to \(F(1, 61) = 4.2\), which would be significant. This comparison of the difference scores involves only one degree of freedom, instead of the two degrees of freedom in the omnibus test, because it combines the non-accent-instruction groups into a single group. Based on this re-analysis, the Derwing et al. study provides some evidence that teaching about pronunciation may aid comprehension by raising awareness of the segmental features of an accent. Thus, awareness of accent characteristics seems to be beneficial to understanding accented speech, but a more effective way of helping listeners gain this awareness, according to research on implicit knowledge and implicit learning, would be through implicit learning rather than explicit learning. In the present study, the imitation task was used so that participants would have the opportunity to discover and implement implicit knowledge about the accented speech rather than passively listening to information without being required to use that information.

The paraphrase task provided a foil for the imitation task. The same material was used for both tasks, so the amount of time spent in training was the same, and the experimenter’s role in both cases was as a facilitator only and not instructor. As in the imitation task, participants in the paraphrase task heard short foreign-accented speech samples twice. During the pauses of the second presentation, the participants paraphrased the phrase they heard, using their own words to express the speaker’s meaning. The requirement to rephrase the speaker’s expressions and come up with
synonyms for the speaker’s words was intended to induce the listener to attend to the meaning and process it well enough to reproduce the content without using the same words. Attention to meaning has been cultivated in other accent training work by giving participants comprehension checks, such as true/false statements or transcription tasks (Clarke, 2002; Weil, 2001). The present research avoided such tests because they would give participants in the paraphrase condition practice on comprehension tasks used in testing, which participants in the imitation condition would not receive. Such practice might enable the participants to perform better on the posttests. Because the results would reflect the practice effect in addition to a training effect, conclusions about the efficacy of training would be difficult to make.

Format of Training Experiments

The basic format of the training experiments, Experiments 2 and 3, involved a pretest, training, and a posttest. Experiment 3 also included a retention test, given 1 week after the completion of training. The training involved listening to several 2-3 min foreign-accented speech samples twice; the second presentation of the samples was a paused version, which had pauses inserted after short phrases. During the pauses in the second presentation of the samples, participants either paraphrased or imitated the preceding phrase, or they remained silent and simply waited to listen to the next phrase. Experiment 2 examined the difference between limited training on three speech samples and more extensive training on five samples. Experiment 3 tested the durability of the training, as well as its transferability to another type of accented speech.
The tests used in Experiments 2 and 3 followed the same format, and they were designed to assess not only participants’ comprehension of accented speech, but also their ability to distinguish different types of accents. In both experiments, the tests had four sections: a) sentence transcription, b) written recall of a paragraph-length speech sample, c) multiple-choice comprehension test over the paragraph-length sample, and d) accent discrimination. For the transcription section, participants transcribed sentences that were 12-16 words long; this section measured participants’ ability to comprehend accented speech with relatively little contextual information. Transcription tasks have been used as a measure of comprehensibility in studies of non-native speech comprehension (e.g., Derwing et al., 2002; Gass & Varonis, 1984; Munro, 1998). The transcription measure was used in the present work because of its sensitivity and because previous research included this type of measure, but because of the limited context available in sentence transcription tasks, two other measures of comprehension, which were more ecologically valid, were also used in the present study.

These other measures, the written recall and the multiple-choice comprehension tests, tested participants’ ability to understand and remember a 2-3 min speech sample. In contrast to the transcription task, listening to a 2-3 min speech sample gave participants more contextual information, so it more closely paralleled real-world listening. To hold a conversation with another person, or even to listen to a lecture and take notes, one does not need to transcribe the speaker’s words verbatim; instead, one needs to comprehend what the speaker says and remember it long enough to respond verbally or to capture the meaning on paper. Free recall is
recognized as providing rich data about what participants remember, but free recall requires more retrieval and generation than does recognition. Therefore, recall can also be a less sensitive measure of memory than a forced-choice recognition test because recognition tends to be an easier task than recall (Zechmeister & Nyberg, 1982). Both free recall and forced-choice tests have been used to examine comprehension in previous research on accented speech (Gass & Varonis, 1984; Major, Fitzmaurice, Bunta, & Balasubramanian, 2002), and both were included in the present training experiments in order to maximize the amount of information available about participants' understanding of the test speech samples.

The final test section, accent discrimination, required participants to decide whether speakers who were reading two short sentences had the same type of accent as the speaker they had heard for the first three sections of the test, or whether the speakers had a different type of accent. None of the speakers in this section of the test were the same as the speaker in the first test sections, nor were they the same as those heard during training. This accent discrimination section of the test assessed two related things: first, whether participants could distinguish one type of accent from another, and second, whether participants could generalize accent characteristics from one speaker to another. This test was designed to examine participants' implicit knowledge about accents by eliciting a decision that required the use of their implicit knowledge.

Before the details of the training experiments are presented, I describe the speech sample collection and rating procedures, as well as a preliminary experiment testing focus of attention during the training tasks. As part of the speech sample
collection process, short excerpts of interviews with non-native speakers were rated for accentedness and intelligibility by linguists who had taught ESL for a number of years. These ratings were used to decide which samples to choose for Experiment 1, which investigated the effectiveness of the imitation and paraphrase training conditions for focusing participants' attention on the pronunciation or meaning of speech, respectively.
Chapter 2:

Speech Sample Collection and Accent Rating

Extemporaneous speech samples were collected from non-native English speakers who were interviewed informally about their native culture and their research or work. Because excerpts of these data were used to train and test native English-speaking participants in the experiments reported here, the interviewer maintained a relaxed atmosphere so that the data would be as natural as possible. The interviews allowed the non-native speakers to use the language as they normally would, not only pronouncing phonemes and words with an accent, but also producing grammatical deviations and lexical variations that would not occur in a native speaker's utterances. As a result, these interviews provided data similar to the type of discourse one might encounter when conversing with a non-native speaker or listening to a lecture by a non-native instructor.

This type of naturalistic data contrasts with the read speech and the isolated phonemes and words which have frequently been used in speech perception and accent comprehension training research. The use of extemporaneous speech in the current research was critical for two reasons. First, relatively long samples of casual speech allow realistic amounts of coarticulation to occur, and, more importantly, they provide sufficient context for invoking top-down processes. As this research intends to look at speech perception, not segment perception, it was necessary to use data that would not preclude the type of variability found in casual speech. Second, the training is designed to help listeners not merely recognize or distinguish
characteristics of accented speech, but actually to understand it better. By presenting extemporaneous speech in training and testing, the experiments produce an experience that closely parallels real-world communication with non-native speakers.

In order to establish the comprehensibility and degree of accentedness of the speech samples, linguists and ESL instructors rated the speech samples on several aspects of pronunciation. These ratings show differences between the groups of non-native speakers who were interviewed, and they confirm that the accent characteristics of the particular speakers who were interviewed conform to the general characteristics expected based on their native languages.

Interviews

Method

Participants

Thirty-six non-native English speakers from four native language backgrounds, Mandarin, Thai, Turkish, and Russian, were interviewed. Nine speakers in each native language group participated. For Mandarin, there were five women and four men; for Russian there were five women and four men; for Thai, there were four women and five men; for Turkish there were five women and four men. In all, 19 women and 17 men participated. All participants were adults over 18 years old, and all were paid $10 for their participation.

According to self-report, the interviewees began learning English when they were 10 years old, on average. At the time of the interview, they were about 30 years old on average, having lived in the United States for an average of almost 4 years and
being about 26 years old when they had arrived in the United States. They reported using English just over 50% of the time in their daily life. These numbers varied slightly for the different native language groups, as shown in Table 1, which also gives the ranges of the responses for each native language group.

Table 1. Language Background Data for Interviewees (all ages given in years).

<table>
<thead>
<tr>
<th>Native language</th>
<th>Mandarin</th>
<th>Russian</th>
<th>Thai</th>
<th>Turkish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age: current</td>
<td>32.0</td>
<td>27.8</td>
<td>30.7</td>
<td>30.1</td>
</tr>
<tr>
<td>Age: began</td>
<td>11.8</td>
<td>9.4</td>
<td>8.0</td>
<td>11.6</td>
</tr>
<tr>
<td>Learning English</td>
<td>11.8</td>
<td>9.4</td>
<td>8.0</td>
<td>11.6</td>
</tr>
<tr>
<td>Age: moved</td>
<td>25.6</td>
<td>23.2</td>
<td>27.3</td>
<td>27.3</td>
</tr>
<tr>
<td>to US</td>
<td>25.6</td>
<td>23.2</td>
<td>27.3</td>
<td>27.3</td>
</tr>
<tr>
<td>Years in US</td>
<td>5.8</td>
<td>3.9</td>
<td>3.3</td>
<td>2.8</td>
</tr>
<tr>
<td>% English use</td>
<td>43.3</td>
<td>52.2</td>
<td>48.9</td>
<td>71.1</td>
</tr>
</tbody>
</table>

Procedure

Before beginning the interview, the interviewer gave participants a language background questionnaire which included questions about their native language and other languages they spoke, the age at which they began learning English, the age at which they moved to the United States, the percentage of L1 and L2 use, and in which circumstances the L1 and L2 were used.

After filling out the questionnaire, participants were given a list of questions to choose from for the interview. The list included two topic areas: culture, and career or academic studies. Both topics had five questions, and participants chose one of the five options under each topic, so that during the interview they talked about two different issues (see Appendix A). The interviewer gave participants time to think
about what they wanted to discuss and to make notes on the questions if they wished to do so. During the interview, the interviewer gave verbal and nonverbal responses to the interviewees in order to maintain a conversational atmosphere. However, the interviewer limited her comments so that the interviews contained primarily the speech of the non-native interviewees. The interviews generally lasted 10-15 min.

The data were recorded on a computer using the Cool Edit program for recording, as well as on a tape recorder for backup. Two microphones were used: for the computer, an Azden SGM-2X dual-barrel professional shotgun microphone, with a frequency response range of 40 to 20,000 Hz; for the tape recorder, a Radio Shack unidirectional dynamic microphone, with a frequency response range of 80 to 15,000 Hz. All interviews took place in the same quiet office.

Accent Rating

Method

Participants

All eight native-English speaking raters had substantial experience communicating with non-native speakers, with at least 14 years of regular contact with non-native speakers. All were experienced at teaching English as a Second Language and evaluating non-native pronunciation, with from 5 to 25 years of teaching experience. They also had advanced degrees (an M.A. or Ph.D.) in linguistics, or were currently pursuing a Ph.D. in linguistics. The raters were paid $25 for their participation.
Procedure

From the 10-15 min interviews of the 36 non-native speakers gathered during the speech sample collection, 1-1.5 min speech samples were extracted from each speaker and used to create compact disks that the raters listened to. The 36 samples were selected so that none included comments or interruptions from the interviewer; the speaker addressed one topic, with the beginning and end of the samples chosen at natural breaks in the discourse. Using Cool Edit, the short segments were cut from the interviews and compiled as audio files to present to the raters.

The speech samples were presented in two orders. Samples were arranged into nine blocks of four samples, with one sample from each of the native language groups in each block. The order of native language groups within the blocks was counterbalanced using a balanced Latin Square; the sequence of four blocks created by the Latin Square was repeated twice, and the ninth block had the order of language groups of the first block in the Latin Square sequence (see Table 2).

<table>
<thead>
<tr>
<th>1st sample</th>
<th>2nd sample</th>
<th>3rd sample</th>
<th>4th sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>Mandarin</td>
<td>Thai</td>
<td>Turkish</td>
</tr>
<tr>
<td>Block 2</td>
<td>Thai</td>
<td>Mandarin</td>
<td>Russian</td>
</tr>
<tr>
<td>Block 3</td>
<td>Turkish</td>
<td>Russian</td>
<td>Thai</td>
</tr>
<tr>
<td>Block 4</td>
<td>Russian</td>
<td>Turkish</td>
<td>Mandarin</td>
</tr>
</tbody>
</table>

The blocks were presented in two orders, half of the raters hearing one order and half hearing the other (see Table 3). The second order was arranged so that blocks that were adjacent in the first order were not adjacent in the second order. Raters were given a 5 to 10 min break in the middle, between samples 18 and 19.
Table 3. Accent Rating: The Two Presentation Orders for the Speech Samples.

<table>
<thead>
<tr>
<th>Blocks</th>
<th>Presentation Order 1</th>
<th>Presentation Order 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

Raters judged the speech samples on 10 aspects of pronunciation: vowels, consonants, vowel reduction, pausing, linking, stress, rhythm, intonation, overall pronunciation, and overall intelligibility (see Appendix B). Criteria from the pronunciation assessment in *Accurate English: A Complete Course in Pronunciation* (Dauer, 1993) guided the development of the rating sheet and the selection of aspects to include in the assessment.

To evaluate the vowels and consonants, raters listed the problematic vowels and consonants, using IPA to indicate the target vowels and consonants and how the speakers rendered these vowels and consonants. For the areas of vowel reduction, pausing, and linking, raters indicated on a seven-point scale whether the pronunciation was *inappropriate* (1) or *native-like* (7), and if there was too much or too little vowel reduction, pausing, or linking. For the other five areas, raters used a seven-point scale to indicate whether the pronunciation was inappropriate or heavily accented, or if it was appropriate or unaccented.

Results

The analysis of the rating data does not include the raters' evaluation of the problematic consonants and vowels, but only the eight items in the rating assessment that were evaluated on a seven-point scale. These items are: 1: vowel reduction; 2:
pausing; 3: linking; 4: stress; 5: rhythm; 6: intonation; 7: overall pronunciation; 8:
overall intelligibility. The means of rating scores for each native language group on
these eight aspects of pronunciation are presented in Table 4. The columns in the
table are ordered by overall pronunciation scores from lowest to highest (most
accented to least accented)

Table 4. Mean Ratings for Aspects of Non-Native Pronunciation, by Native
Language Group.

<table>
<thead>
<tr>
<th>Item</th>
<th>Thai Mean</th>
<th>Mandarin Mean</th>
<th>Turkish Mean</th>
<th>Russian Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. vowel reduction</td>
<td>4.35</td>
<td>4.86</td>
<td>5.91</td>
<td>5.50</td>
</tr>
<tr>
<td>SD</td>
<td>0.38</td>
<td>0.74</td>
<td>0.71</td>
<td>0.61</td>
</tr>
<tr>
<td>2. pausing</td>
<td>4.52</td>
<td>4.90</td>
<td>5.25</td>
<td>5.70</td>
</tr>
<tr>
<td>SD</td>
<td>0.58</td>
<td>0.53</td>
<td>0.86</td>
<td>0.80</td>
</tr>
<tr>
<td>3. linking</td>
<td>4.17</td>
<td>4.43</td>
<td>5.04</td>
<td>5.71</td>
</tr>
<tr>
<td>SD</td>
<td>0.55</td>
<td>0.71</td>
<td>0.80</td>
<td>0.62</td>
</tr>
<tr>
<td>4. stress</td>
<td>4.04</td>
<td>4.67</td>
<td>5.23</td>
<td>5.61</td>
</tr>
<tr>
<td>SD</td>
<td>0.61</td>
<td>0.66</td>
<td>0.42</td>
<td>0.86</td>
</tr>
<tr>
<td>5. rhythm</td>
<td>3.82</td>
<td>4.26</td>
<td>4.81</td>
<td>5.31</td>
</tr>
<tr>
<td>SD</td>
<td>0.63</td>
<td>0.67</td>
<td>0.67</td>
<td>0.86</td>
</tr>
<tr>
<td>6. intonation</td>
<td>3.95</td>
<td>4.59</td>
<td>5.01</td>
<td>5.59</td>
</tr>
<tr>
<td>SD</td>
<td>0.76</td>
<td>0.62</td>
<td>0.69</td>
<td>0.85</td>
</tr>
<tr>
<td>7. overall pronunciation</td>
<td><strong>3.66</strong></td>
<td><strong>3.93</strong></td>
<td><strong>4.78</strong></td>
<td><strong>5.35</strong></td>
</tr>
<tr>
<td>SD</td>
<td><strong>0.67</strong></td>
<td><strong>0.87</strong></td>
<td><strong>0.79</strong></td>
<td><strong>0.83</strong></td>
</tr>
<tr>
<td>8. overall intelligibility</td>
<td><strong>5.26</strong></td>
<td><strong>5.23</strong></td>
<td><strong>6.46</strong></td>
<td><strong>6.67</strong></td>
</tr>
<tr>
<td>SD</td>
<td><strong>0.59</strong></td>
<td><strong>1.12</strong></td>
<td><strong>0.46</strong></td>
<td><strong>0.35</strong></td>
</tr>
</tbody>
</table>

A reliability analysis of the data was performed to determine the extent to
which the raters agreed in their ratings. In order to avoid the problem of unequal
variances for different judges, the ratings were converted to z-scores before
calculating the intraclass correlations. Intraclass correlations provide an index of the
degree of nonindependence between data points, in this case, judges' ratings: A
positive value, at most +1, indicates that scores within a group are similar on average;
a negative value indicates dissimilarity on average within a group (Judd & McClelland, 1989).

The single measure intraclass correlations were positive but tended to be rather low; however, the high average measures (Cronbach’s alphas) indicate that there is good reliability of the average judgments of the eight judges (see Table 5).

Table 5. Intraclass Correlations of Raters for Pronunciation Aspects.

<table>
<thead>
<tr>
<th>Item</th>
<th>Intraclass Correlation</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single Measure</td>
<td>Average Measure (Cronbach’s α)</td>
<td></td>
</tr>
<tr>
<td>1. vowel reduction</td>
<td>.36</td>
<td>.82</td>
<td></td>
</tr>
<tr>
<td>2. pausing</td>
<td>.43</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td>3. linking</td>
<td>.45</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>4. stress</td>
<td>.51</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td>5. rhythm</td>
<td>.46</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>6. intonation</td>
<td>.54</td>
<td>.91</td>
<td></td>
</tr>
<tr>
<td>7. overall pronunciation</td>
<td>.66</td>
<td>.94</td>
<td></td>
</tr>
<tr>
<td>8. overall intelligibility</td>
<td>.39</td>
<td>.84</td>
<td></td>
</tr>
</tbody>
</table>

In other words, the agreement between a given judge and another judge was not particularly good because the spread of the ratings was broad, but averaged together the eight judges establish a reliable measure of ratings of accentedness.

Contrasts between the ratings of the language groups were made using a MANOVA with a step-down adjustment to allow for nonnormality and multiple testing of the data (Westfall, Randall, Rom, Wolfinger, & Hochberg, 1999). The mean rating scores for each assessment item were used in these analyses, as justified by the high alphas in the intraclass correlation analyses, which indicate that the mean rating scores represent a stable and reliable measure of accentedness.
Three contrasts were examined; these contrasts were chosen because the data from the Mandarin and Thai native speakers seemed more heavily accented than the data from the Turkish and Russian native speakers (as shown in Table 4). The three contrasts were: Mandarin and Thai versus Turkish and Russian; Mandarin versus Thai; Turkish versus Russian. Table 6 gives the step-down adjusted $p$ values for these contrasts for each item. There were no significant differences on any of the items between the Mandarin and Thai native speaker groups nor between the Turkish and Russian native speaker groups. However, the contrasts between the Mandarin and Thai pair compared to the Turkish and Russian pair were significant for all items. Importantly, the final two items, overall pronunciation and overall intelligibility, strongly indicate a difference between the pairs of language groups, but not within the pairs. Overall, then, it is clear that this group of Mandarin and Thai speakers have heavier accents and are more difficult to understand than the Turkish and Russian speakers.

Table 6. Adjusted $p$ Values for Contrasts Between Mean Ratings of Language Groups.

<table>
<thead>
<tr>
<th>Item</th>
<th>Adjusted $p$ Values for Contrasts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mandarin &amp; Thai vs. Turkish &amp; Russian</td>
</tr>
<tr>
<td>1. vowel reduction</td>
<td>.015</td>
</tr>
<tr>
<td>2. pausing</td>
<td>.029</td>
</tr>
<tr>
<td>3. linking</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>4. stress</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>5. rhythm</td>
<td>.003</td>
</tr>
<tr>
<td>6. intonation</td>
<td>.003</td>
</tr>
<tr>
<td>7. overall pronunciation</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>8. overall intelligibility</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
The results of the speech sample rating indicate that the Russian- and Turkish-accented speech samples form a group that is more intelligible and less heavily accented than the Mandarin- and Thai-accented speech samples. Despite differences in Russian and Turkish accents, the speakers represented in this sample have similar levels of accentedness and intelligibility, and the same is true for the Mandarin and Thai native speakers.

Deviations from target vowels and consonants that were noted by the raters confirm that the segmental aspects of the accents in these speech samples conform to the typical segmental deviations of Mandarin, Russian, Thai, and Turkish accents. Table 7 summarizes general patterns in the pronunciation of consonants and vowels for the four accent types. Although the table does not give an exhaustive account of the qualities of these accents, it presents the most common and persistent segmental problems and is based on the description of these accents in Swan and Smith (2001), in conjunction with the raters' judgments of the speakers in this particular dataset.

Consonants that are problematic for all four accent types include the fricatives /v, θ, δ/ and the approximant /j/; vowels that are difficult for all accent types except Thai are /i/ and /a/. Mandarin and Russian accents share some common features, such as vowel confusions of /i/ and /u/, and substitutions of other vowels for /æ/. Both Mandarin and Russian accents have devoicing of voiced stops, although in a Russian accent the devoicing tends to occur only word finally. Mandarin and Thai accents exhibit similar devoicing of /z/, and final consonants tend to be dropped; in a Thai

Discussion

The results of the speech sample rating indicate that the Russian- and Turkish-accented speech samples form a group that is more intelligible and less heavily accented than the Mandarin- and Thai-accented speech samples. Despite differences in Russian and Turkish accents, the speakers represented in this sample have similar levels of accentedness and intelligibility, and the same is true for the Mandarin and Thai native speakers.

Deviations from target vowels and consonants that were noted by the raters confirm that the segmental aspects of the accents in these speech samples conform to the typical segmental deviations of Mandarin, Russian, Thai, and Turkish accents. Table 7 summarizes general patterns in the pronunciation of consonants and vowels for the four accent types. Although the table does not give an exhaustive account of the qualities of these accents, it presents the most common and persistent segmental problems and is based on the description of these accents in Swan and Smith (2001), in conjunction with the raters' judgments of the speakers in this particular dataset.

Consonants that are problematic for all four accent types include the fricatives /v, θ, δ/ and the approximant /j/; vowels that are difficult for all accent types except Thai are /i/ and /a/. Mandarin and Russian accents share some common features, such as vowel confusions of /i/ and /u/, and substitutions of other vowels for /æ/. Both Mandarin and Russian accents have devoicing of voiced stops, although in a Russian accent the devoicing tends to occur only word finally. Mandarin and Thai accents exhibit similar devoicing of /z/, and final consonants tend to be dropped; in a Thai

Discussion

The results of the speech sample rating indicate that the Russian- and Turkish-accented speech samples form a group that is more intelligible and less heavily accented than the Mandarin- and Thai-accented speech samples. Despite differences in Russian and Turkish accents, the speakers represented in this sample have similar levels of accentedness and intelligibility, and the same is true for the Mandarin and Thai native speakers.

Deviations from target vowels and consonants that were noted by the raters confirm that the segmental aspects of the accents in these speech samples conform to the typical segmental deviations of Mandarin, Russian, Thai, and Turkish accents. Table 7 summarizes general patterns in the pronunciation of consonants and vowels for the four accent types. Although the table does not give an exhaustive account of the qualities of these accents, it presents the most common and persistent segmental problems and is based on the description of these accents in Swan and Smith (2001), in conjunction with the raters' judgments of the speakers in this particular dataset.

Consonants that are problematic for all four accent types include the fricatives /v, θ, δ/ and the approximant /j/; vowels that are difficult for all accent types except Thai are /i/ and /a/. Mandarin and Russian accents share some common features, such as vowel confusions of /i/ and /u/, and substitutions of other vowels for /æ/. Both Mandarin and Russian accents have devoicing of voiced stops, although in a Russian accent the devoicing tends to occur only word finally. Mandarin and Thai accents exhibit similar devoicing of /z/, and final consonants tend to be dropped; in a Thai

Discussion

The results of the speech sample rating indicate that the Russian- and Turkish-accented speech samples form a group that is more intelligible and less heavily accented than the Mandarin- and Thai-accented speech samples. Despite differences in Russian and Turkish accents, the speakers represented in this sample have similar levels of accentedness and intelligibility, and the same is true for the Mandarin and Thai native speakers.

Deviations from target vowels and consonants that were noted by the raters confirm that the segmental aspects of the accents in these speech samples conform to the typical segmental deviations of Mandarin, Russian, Thai, and Turkish accents. Table 7 summarizes general patterns in the pronunciation of consonants and vowels for the four accent types. Although the table does not give an exhaustive account of the qualities of these accents, it presents the most common and persistent segmental problems and is based on the description of these accents in Swan and Smith (2001), in conjunction with the raters' judgments of the speakers in this particular dataset.

Consonants that are problematic for all four accent types include the fricatives /v, θ, δ/ and the approximant /j/; vowels that are difficult for all accent types except Thai are /i/ and /a/. Mandarin and Russian accents share some common features, such as vowel confusions of /i/ and /u/, and substitutions of other vowels for /æ/. Both Mandarin and Russian accents have devoicing of voiced stops, although in a Russian accent the devoicing tends to occur only word finally. Mandarin and Thai accents exhibit similar devoicing of /z/, and final consonants tend to be dropped; in a Thai

Discussion

The results of the speech sample rating indicate that the Russian- and Turkish-accented speech samples form a group that is more intelligible and less heavily accented than the Mandarin- and Thai-accented speech samples. Despite differences in Russian and Turkish accents, the speakers represented in this sample have similar levels of accentedness and intelligibility, and the same is true for the Mandarin and Thai native speakers.

Deviations from target vowels and consonants that were noted by the raters confirm that the segmental aspects of the accents in these speech samples conform to the typical segmental deviations of Mandarin, Russian, Thai, and Turkish accents. Table 7 summarizes general patterns in the pronunciation of consonants and vowels for the four accent types. Although the table does not give an exhaustive account of the qualities of these accents, it presents the most common and persistent segmental problems and is based on the description of these accents in Swan and Smith (2001), in conjunction with the raters' judgments of the speakers in this particular dataset.

Consonants that are problematic for all four accent types include the fricatives /v, θ, δ/ and the approximant /j/; vowels that are difficult for all accent types except Thai are /i/ and /a/. Mandarin and Russian accents share some common features, such as vowel confusions of /i/ and /u/, and substitutions of other vowels for /æ/. Both Mandarin and Russian accents have devoicing of voiced stops, although in a Russian accent the devoicing tends to occur only word finally. Mandarin and Thai accents exhibit similar devoicing of /z/, and final consonants tend to be dropped; in a Thai

Discussion

The results of the speech sample rating indicate that the Russian- and Turkish-accented speech samples form a group that is more intelligible and less heavily accented than the Mandarin- and Thai-accented speech samples. Despite differences in Russian and Turkish accents, the speakers represented in this sample have similar levels of accentedness and intelligibility, and the same is true for the Mandarin and Thai native speakers.

Deviations from target vowels and consonants that were noted by the raters confirm that the segmental aspects of the accents in these speech samples conform to the typical segmental deviations of Mandarin, Russian, Thai, and Turkish accents. Table 7 summarizes general patterns in the pronunciation of consonants and vowels for the four accent types. Although the table does not give an exhaustive account of the qualities of these accents, it presents the most common and persistent segmental problems and is based on the description of these accents in Swan and Smith (2001), in conjunction with the raters' judgments of the speakers in this particular dataset.

Consonants that are problematic for all four accent types include the fricatives /v, θ, δ/ and the approximant /j/; vowels that are difficult for all accent types except Thai are /i/ and /a/. Mandarin and Russian accents share some common features, such as vowel confusions of /i/ and /u/, and substitutions of other vowels for /æ/. Both Mandarin and Russian accents have devoicing of voiced stops, although in a Russian accent the devoicing tends to occur only word finally. Mandarin and Thai accents exhibit similar devoicing of /z/, and final consonants tend to be dropped; in a Thai
accent the dropping of final consonants occurs particularly with consonant clusters, where only the first consonant of a cluster is retained.

Table 7. Segmental Aspects of the Accents of the Native Mandarin, Russian, Thai, and Turkish Speakers in These Speech Samples.

<table>
<thead>
<tr>
<th>Native Language</th>
<th>Problematic Target Consonant</th>
<th>Typical Substitution</th>
<th>Problematic Target Vowel</th>
<th>Typical Substitution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td>b, d, g v θ, δ z j</td>
<td>b, d, g f or w s, z or t, d s l (distinction unclear) oo or dropped dropped</td>
<td>i i æ e a Λ or ɔ or ɑ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>final 1 final consonants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td>pʰ, tʰ, kʰ b, d, g η v θ, δ tʃ</td>
<td>p, t, k or b, d, g p, t, k n or nk w (distinction unclear) s, z s j various substitutions</td>
<td>i i æ e or ɑ a ɔ or ɔɔ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>final b, d, g v θ, δ tʃ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>initial tw, tj, pr, dʒ bʒ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thai</td>
<td>v θ, δ j z l consonant clusters</td>
<td>w t or s, d or t or s tʃ s l reduced</td>
<td>ɛt various substitutions</td>
<td></td>
</tr>
<tr>
<td>Turkish</td>
<td>v θ, δ j</td>
<td>w (distinction unclear) t, d various substitutions word final devoicing</td>
<td>i a u ı oo ɔ, ʌ</td>
<td></td>
</tr>
</tbody>
</table>
A unique feature of the Mandarin accent is the tendency for final /l/ to be dropped or substituted with an /ou/ vowel. Unique features of the Russian accent include lack of aspiration on voiceless stops, which can sound like their voiced counterparts as a result, and the substitution of an alveolar nasal /n/ for the velar /ŋ/.

Russian and Thai accents are marked by confusion of the fricative /ʃ/ and the affricate /tʃ/, but in a Russian accent the fricative is substituted for the affricate, whereas in a Thai accent the affricate is substituted for the fricative. The mid front vowel /eɪ/ is problematic for native Thai speakers, with a variety of front or mid vowels substituted for it, and the high back vowel /u/ causes difficulty for native Turkish speakers. The Turkish accent also has word final devoicing of voiced stops and of the affricate /dʒ/.

Thus, despite the commonalities shared by the four types of accents, each one has unique segmental traits which distinguish it from the others.
CHAPTER 3:  
EXPERIMENT 1—FOCUS OF ATTENTION

This experiment tested some of the materials and methods that would be used in the training experiments to ensure their viability. For the experiment, participants listened to two 2-3 min accented speech samples twice and answered multiple-choice questions about the content and pronunciation of the speech. Participants were randomly divided into four conditions involving either silence or verbal repeating tasks during the second presentation of the speech samples.

It was not a training experiment, although it employed the two training tasks, imitation and paraphrase. The imitation task, repeating the speaker’s exact words and pronunciation, aimed to promote listeners’ awareness of the speaker’s shifts in pronunciation. In contrast, the goal of the paraphrase task, paraphrasing the content of the speaker’s utterance by using different words, was to focus listeners’ attention on the meaning of the speech. This experiment investigated whether these tasks prompted the types of attentional focus that they were expected to promote, that is, a focus on pronunciation in the imitation task and a focus on content in the paraphrase task.

Besides examining attentional focus for the imitation and paraphrase tasks, this experiment served a number of other purposes. First, a language background questionnaire elicited information about the types of foreign accents to which participants had been exposed; this information was used to guide the selection of the foreign-accented speech used in subsequent experiments. Second, the participants’
ability to follow the instructions about the tasks in each condition tested the appropriateness of these instructions. Third, the experiment demonstrated differences in participants’ abilities to remember the content or the features of accented speech, at least given the experimental paradigm that was used.

Method

Participants

Thirty-two native-English speakers participated in this experiment; seven were men and twenty-five were women. All were students in an introductory linguistics course at the University of Colorado and received course credit for participating. None reported having speech or hearing dysfunctions. Most participants (30) reported having some contact with non-native English speakers, most commonly (26 participants) with Spanish speakers. Other native language groups many participants reported contact with include: French—12, German—10, Japanese—7, Russian—7, Arabic—5, and Chinese—4. Many participants (27) had traveled abroad, primarily to Central and South America (14), but also to European countries (including the British Isles—10, France—10, Germany—8, Spain—8, Italy—7, Switzerland—4). Some participants had traveled outside of Europe, with one participant each reporting travel to China, Hong Kong, Japan, New Zealand, North Africa, Taiwan, and Thailand.
Materials

Using Cool Edit, two 2- to 3-min speech samples were excerpted from two of the interviews with non-native English speakers which were gathered in the speech sample collection phase. The Russian-accented sample was 2 min, 25 s long; the female speaker discussed her work as a professional piano player, describing a project she was involved in to create a CD of Israeli prayers set to traditional Israeli music (see Russian, Appendix C). The Turkish-accented sample was 2 min, 40 s long; the male speaker explained research he was conducting on speech recognition software, noting challenges that this research faces (see Turkish A, Appendix C).

In addition to the regular versions of these samples, versions with pauses after short phrases were also created using Cool Edit. Pauses were inserted at natural phrase breaks in the syntax, and the length of phrases was between 4-13 words. For the Russian sample, the average length of phrases was 8.1 words, and with the addition of the pauses, this version of the sample was 10 min, 54 s long. For the Turkish sample, the average length of phrases was 7.2 words, and the version of the sample with pauses was 11 min, 8 s long. The length of each pause varied according to the length of the phrase preceding it, but all of the pauses were approximately 4 s longer than their preceding phrases.

For the purpose of demonstrating to participants what they were to do for the different conditions of the experiment, two short speech samples were developed. These were taken from a native Hindi speaker who had been interviewed as part of the original speech sample collection process, but whose data were not used for training in any of the experiments. The first sample was 16 s long, and the second
with the Russian- and Turkish-accented samples, versions of these demonstration samples with pauses were also created, and these were 37 and 35 s long for the first and second samples, respectively.

Multiple-choice comprehension and pronunciation tests were created for the Russian- and Turkish-accented samples. Each of the four tests contained 10 questions with 4 options each; participants were told to circle the letter of the best answer. The comprehension questions asked about details of the content of the speech samples. For example, one of the questions about the Russian-accented sample was:

1. Where did the speaker first play professionally in the United States?
   a) at a children’s party
   b) at a friend’s wedding
   c) at an elementary school
   d) in a church

For the pronunciation tests, 10 words were selected from each speech sample, and 4 pronunciation options were given for each word. The options were spelled in an eye dialect, using the type of letter-sound representations commonly found in dictionaries rather than the International Phonetic Alphabet (IPA) under the assumption that very few undergraduate students are familiar with IPA. The instructions indicated: *The underlined section is the part of the word that is accented, and whatever is not underlined was not pronounced with an accent.* Participants were encouraged to ask the experimenter about the intended pronunciation of any option for which they were
not sure of the pronunciation. From the Turkish-accented pronunciation test, one question was:

2. that
   a) dat
   b) deht
   c) theht
   d) thad

**Design and Procedure**

Participants were tested individually. First, participants filled out a brief questionnaire about their language background, including information about any foreign countries they had visited, and the extent of their contact with non-native English speakers. Next, participants were told that they would be listening to two short speech samples twice, the first time hearing the entire sample, and the second time hearing the sample broken into short phrases with pauses after the phrases. Participants were informed that they would be tested on the speech samples, but they were not told the nature of those tests.

The order of presentation of the two speech samples was counterbalanced so that half of the participants heard the Russian-accented sample first and half heard the Turkish-accented sample first. In addition, the experiment included four conditions: silence, repetition, imitation, and paraphrase. Each participant was in a control condition (either silence or repetition) and an experimental condition (either imitation or paraphrase); the order of the control and experimental conditions was
counterbalanced using a Latin Square. The reason for having two control conditions was that the silence condition involved no secondary task, so participants would presumably be able to listen most freely and naturally in this condition; the repetition condition, however, did involve a secondary task, though it was not as demanding as the tasks in the experimental conditions. Thus, the repetition condition served as a control for the influence of a secondary task on comprehension and attention to pronunciation patterns.

In all conditions, participants simply listened to the first presentation of each sample. Immediately following the first presentation, the version of the sample with pauses was played. In the silence condition, participants listened quietly again to the second presentation. In the repetition condition, participants repeated each phrase aloud during the pauses following the phrases. In the imitation condition, participants were instructed to imitate as closely as possible the phrases they heard, using the same words and the same pronunciation that the speaker used. In the paraphrase condition, the participants were told to paraphrase each phrase during the pauses after the phrases, using their own words to express the meaning of the phrase.

Before the participants heard the first speech sample, the experimenter provided demonstrations of the task in each condition, playing the short Hindi-accented samples for the participants. According to the participant’s experimental condition, the experimenter was silent, repeated, imitated, or paraphrased the phrases during the pauses of the second version of the samples. Similar instructions and demonstrations were given before the second sample, as participants were all in different conditions than they had been in for the first sample.
After listening to each speech sample, the participants took the comprehension and pronunciation tests for that sample. The order of presentation of these tests was counterbalanced so that half of the participants in each condition received the pronunciation test first, and half received the comprehension test first. To counterbalance the order of test presentation within participants, each participant took one set of tests in one order (e.g., pronunciation then comprehension) and the other set in the opposite order (e.g., comprehension then pronunciation).

Results

The data were entered as first and second comprehension scores and first and second pronunciation scores. Separate analyses were computed for each combination of score type (comprehension and pronunciation) and test position (first and second). Each analysis was a 2 x 4 (language x condition) analysis of variance. There was a significant main effect of language only for the second comprehension scores, $F(1, 24) = 4.36, p = .048$; the scores for the Russian-accented sample ($M = 6.06$) were higher than those for the Turkish-accented sample ($M = 5.00$). The means for the test scores in each condition are shown in Table 8; although there was no significant main effect of condition, the silence and paraphrase conditions tend to have the best scores on comprehension, but the pronunciation scores do not seem to follow a clear pattern.
Table 8. Experiment 1: Mean Test Scores by Condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Comprehension</th>
<th>Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
</tr>
<tr>
<td>Silence</td>
<td>Mean</td>
<td>6.00</td>
</tr>
<tr>
<td></td>
<td>SEM</td>
<td>0.68</td>
</tr>
<tr>
<td>Repetition</td>
<td>Mean</td>
<td>5.50</td>
</tr>
<tr>
<td></td>
<td>SEM</td>
<td>0.42</td>
</tr>
<tr>
<td>Imitation</td>
<td>Mean</td>
<td>4.75</td>
</tr>
<tr>
<td></td>
<td>SEM</td>
<td>0.73</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>Mean</td>
<td>6.00</td>
</tr>
<tr>
<td></td>
<td>SEM</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Planned paired-comparison <i>t</i>-tests revealed significant differences between the pronunciation and comprehension scores for both the first sets of tests, <i>t</i>(31) = 5.82, <i>p</i> < .001, and the second sets of tests, <i>t</i>(31) = 5.35, <i>p</i> < .001. Comprehension scores were significantly higher in both cases than pronunciation scores, as can be seen in Table 8. The first and second sets of scores were also compared with planned paired-comparison <i>t</i>-tests, which showed no significant difference between the first (<i>M</i> = 5.56) and second (<i>M</i> = 5.54) sets of comprehension scores, but a marginally significant difference between the first (<i>M</i> = 3.13) and second (<i>M</i> = 3.74) sets of pronunciation scores, <i>t</i>(31) = -1.94, <i>p</i> = .061. Therefore, the first and second comprehension scores are comparable; however, the second set of pronunciation scores tend to be higher than the first set, indicating that participants’ performance improved slightly on the pronunciation scores from the first to second speech sample.

Because the first and second sets of comprehension scores were not different according to the paired-comparison <i>t</i>-test, they were combined to be analyzed in separate planned one-tailed <i>t</i>-tests. By these tests, the advantage for the paraphrase condition (<i>M</i> = 6.06) relative to both the repetition condition (<i>M</i> = 5.19) and imitation
condition \((M = 5.00)\) was significant, \(t(30) = 1.76, p = .045\), and \(t(30) = -1.93, p = .032\), respectively.

**Discussion**

The degrees of accent for the two speech samples are rated similarly by the expert raters, so a difference was not expected between the test scores of participants who listened to Russian-accented speech and those who listened to Turkish-accented speech. No difference was seen, except in the case of the second comprehension scores, where the Russian-accented scores were higher than the Turkish-accented scores. This discrepancy may be due to differences between the content of the samples and the practice effect of taking a comprehension test on the first sample. After taking the first multiple-choice comprehension test, participants were able to anticipate the type of comprehension test that would be given, and this anticipation helped those participants listening to the Russian-accented sample second to score significantly higher than those listening to the Turkish-accented sample. Perhaps the more technical information of the Turkish-accented sample, which dealt with research on speech recognition systems, prepared participants to listen closely to the details of the second Russian-accented sample, and combined with the knowledge of what the comprehension test was like, the practice allowed those participants to score better on the Russian-accented comprehension test than the other participants taking the Turkish-accented comprehension test second. In contrast, participants who first heard the Russian-accented sample about a professional pianist and her CD project were not obliged to pay close attention to technical details, so even though they could predict
what the test for the second sample would be like, this boost was mitigated by the less intuitively accessible information in the Turkish-accented sample.

Overall, comprehension scores did not improve from the first to the second sample, as indicated by the paired-comparison \( t \)-tests. However, pronunciation scores did tend to improve. In addition, the pronunciation tests were much more difficult than the comprehension tests, as implied by the fact that the pronunciation scores were significantly lower than the comprehension scores. These effects may be influenced by various factors. Listening to speech in order to understand the speaker’s message is the whole point of speech perception, so comprehension tests on the content of a speech sample check the material the listener naturally attended to in listening to the sample. On the other hand, listeners do not usually pay close attention to the pronunciation patterns of another person’s speech because holding a meaningful conversation does not entail noticing or defining pronunciation patterns. Testing about the speakers’ pronunciation is, therefore, a more difficult task and one for which listeners do not have years of practice. Furthermore, the literature on speech perception demonstrates that listeners are influenced by the speaker’s pronunciation patterns and accent, but listeners are generally unable to describe the segmental and suprasegmental aspects of that pronunciation. The pronunciation tests in this experiment demanded a certain amount of accuracy in the listeners’ representations of the speakers’ accented pronunciations of words, and this accuracy was largely beyond most listeners’ ability, especially as they were required to recall and reconstruct the accented words from memory. Interestingly, though, after the participants took the pronunciation test on the first sample and were aware that they
would be similarly tested after the second sample, they were able to concentrate more on pronunciation, and their test scores tended to improve. This result suggests that listeners can be encouraged to attend to pronunciation, and that when they do, they are better at noticing details and patterns.

Finally, the *t*-tests on the combined first and second comprehension scores showed a significant difference between the comprehension in the paraphrase condition and the imitation conditions, as well as between the paraphrase and repetition conditions. Even with the added burden of a secondary paraphrasing task, participants in the paraphrase condition did not comprehend worse than those in the silence condition—in fact, the mean score was numerically slightly better for the paraphrase condition (*M* = 6.06) than for the silence condition (*M* = 5.94) for the combined first and second comprehension scores. More importantly, the fact that comprehension scores in the paraphrase condition were better than those in the repetition condition seems to provide evidence that the additional processing required in the paraphrase task enabled listeners to focus on the content and meaning of the speech. Both the repetition and paraphrase conditions involved a secondary task as the participants listened to the accented speech, but the repetition task was neutral, whereas the paraphrase task demanded that participants find synonyms for the speaker’s expressions in order to restate the phrases in their own words. With the effect of a secondary task controlled, it is apparent that paraphrasing did indeed encourage participants to attend to meaning, resulting in improved comprehension scores. The difference between the comprehension scores in the two experimental conditions, with the combined first and second comprehension scores in the imitation
condition worse than those in the paraphrase condition, may imply that participants in
the imitation condition were unable to comprehend the speech samples as well
because their attention was focused elsewhere, presumably on the pronunciation of
the samples rather than the meaning. This interpretation would be supported by better
pronunciation scores in the imitation condition, but because of the already mentioned
difficulties of measuring pronunciation awareness, this result is not apparent in these
data. Anecdotal reports from several participants in the imitation condition did,
however, confirm that participants felt that they were concentrating on pronunciation
in that condition, and they commented that they felt more comfortable and competent
in taking the pronunciation test after having imitated the speech.

The information gleaned from the Experiment 1 had bearing on the
development of materials for the experiments described here as well as for their
design. The language background questionnaire indicated that many participants (7)
had contact with native Russian speakers, and some (4) had contact with native
Chinese speakers. Only two participants reported contact with native Thai speakers,
and only one reported contact with native Turkish speakers. Therefore, in order to
avoid the more commonly contacted native languages of undergraduate students at
the University of Colorado, Turkish- and Thai-accented speech would be better to use
in training than Mandarin or Russian.

The instructions and demonstrations were understandable to these
participants, and the participants were able to perform the required tasks more or less
successfully. Some participants commented that the imitation and paraphrase tasks
were difficult and made them think, but the tasks were not beyond their abilities. Thus, these tasks seemed appropriate to use during training in future experiments.

From the results of the Experiment 1, it is apparent that listeners more easily focus on the content of speech than on the pronunciation, but that they are able to pay attention to pronunciation when they have some impetus, such as knowing that they will be tested on pronunciation. Participants were not informed how they would be tested in this experiment or what they would be tested on. When they realized that a pronunciation test was part of the experiment, they seemed better able to pay attention to pronunciation, as would be expected based on studies showing that participants develop test-appropriate memory strategies when they know what information will be tested and how it will be tested (see Zechmeister & Nyberg, 1982).

Using the paraphrase condition to train listeners to attend to the message of speech and the imitation condition to train listeners to attend to the pronunciation of speech seemed to be justified by Experiment 1. The better performance on comprehension tests in the paraphrase condition clearly indicates that this condition focuses listeners' attention on the meaning of the speech. The poorer comprehension of those in the imitation condition may not provide as clear evidence that participants in this condition attended to pronunciation, but it does indicate that participants were not attending well to comprehension. Moreover, anecdotal reports from participants in the imitation condition confirm that they felt their awareness of pronunciation patterns improved by imitating the accented speech, suggesting that they were focusing on the pronunciation even though their pronunciation scores do not reflect this increased attention.
This first training experiment established the effect of training on the comprehension of accented speech and compared the consequences of limited training versus extensive training. Two training tasks, imitation and paraphrase, combined with two training amounts, limited and extensive, to create four conditions: limited/imitation, limited/paraphrase, extensive/imitation, extensive/paraphrase. Because Turkish-accented speech appeared to be the least familiar type of accent to University of Colorado undergraduates, according to Experiment 1, it was used for training and testing. A pretest and posttest were given to participants in all conditions.

The primary issue addressed by this experiment was what effect, if any, training would have on participants' understanding of Turkish-accented speech. Various questions fell under this issue. First, would training improve comprehension, as measured by the tests? Second, would training improve participants' ability to discriminate between different accents? Third, would the two training tasks affect test scores differently? The evidence from Experiment 1 indicates that the paraphrase task encouraged a focus on the content of speech whereas the imitation task did not. Therefore, in Experiment 2, scores on comprehension measures were expected to improve more for participants who paraphrased the content during training than those who imitated the accent. Even though Experiment 1 did not provide statistical evidence that the imitation task focused participants' attention on pronunciation,
anecdotal evidence suggested that this might be the case. If so, the scores on the
discrimination test should improve more for participants practicing the imitation task
than those practicing the paraphrase task.

A secondary issue that the experiment dealt with was the impact of amount of
training on test scores. By varying the number of speech samples presented during
training, the experiment measured whether a limited amount of training was sufficient
to familiarize participants with foreign-accented speech and whether more extensive
training including five speech samples was beneficial or if it instead caused fatigue.

Method

Participants

Ninety-six native English speakers participated; all were undergraduate
students at the University of Colorado who received course credit for their
participation. None of the participants reported being bilingual or having speech or
hearing dysfunctions. Most participants (76) reported having some contact with non-
native English speakers, primarily with native Spanish speakers (76), but also
commonly with French (18), German (17), Italian (11), Japanese (9), and Chinese (7)
speakers. No one reported contact with native Turkish speakers. Most participants
(70) also indicated that they had traveled abroad. The most popular travel
destinations included Central and South America and Europe, but some participants
had also traveled in Asia, Africa, the Caribbean, and Australia. None, however,
reported travel to Turkey or Cyprus.
Materials

The training phase of this experiment presented speech samples from the interviews with nonnative English speakers which were conducted during the collection of speech samples. Using Cool Edit, five 2- to 3-min segments were extracted from the interviews of five different native Turkish speakers (see Appendix D). Three or five of these samples were presented during training, depending on the training condition. These samples had a regular version and a paused version; a paused version was created by inserting pauses at natural phrase breaks. The paused versions were 7 to 10.5 min long (see Table 9 for details).

Table 9. Experiment 2: Training Samples Information

<table>
<thead>
<tr>
<th>Training Sample</th>
<th>Speaker Gender</th>
<th>Total # of Words</th>
<th>Mean # of Words per Phrase</th>
<th>Unpaused Version</th>
<th>Paused Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>male</td>
<td>308</td>
<td>7.9</td>
<td>2 min 43 s</td>
<td>8 min 35 s</td>
</tr>
<tr>
<td>2</td>
<td>female</td>
<td>228</td>
<td>7.3</td>
<td>2 min 23 s</td>
<td>7 min 01 s</td>
</tr>
<tr>
<td>3</td>
<td>female</td>
<td>249</td>
<td>7.8</td>
<td>2 min 35 s</td>
<td>7 min 34 s</td>
</tr>
<tr>
<td>4</td>
<td>male</td>
<td>330</td>
<td>7.9</td>
<td>2 min 40 s</td>
<td>9 min 09 s</td>
</tr>
<tr>
<td>5</td>
<td>female</td>
<td>388</td>
<td>7.6</td>
<td>2 min 32 s</td>
<td>10 min 21 s</td>
</tr>
</tbody>
</table>

The testing materials included speech samples from the interviews described in the Speech Sample Collection and Accent Rating chapter as well as sentences from an accented speech database which was collected by researchers at Duke University (Arslan & Hansen, 1997). The speech in the Duke University corpus was sampled at 8 kHz, using either microphone or telephone input into a computer database, and it consisted of words and phrases read by informants. Two tests were developed, both with four sections: a) transcription, b) written recall, c) multiple-choice
comprehension, and d) accent discrimination (see Appendix F for a sample test).

Two of the Turkish-accented interviews were the source for the speech samples presented in the first three sections of the tests; Test A presented excerpts from a male speaker and Test B presented excerpts from a female speaker (see Appendix C). The Duke University database provided the sentences used in the final accent discrimination section of the test.

For the transcription section, participants transcribed word-for-word five sentences which had been cut from the interviews using Cool Edit. These sentences were 12-15 words long, with a total of 67 words in each set of transcription sentences (see Appendix E).

The comprehension sections of each test, both the written recall and multiple-choice sections, were based on a short speech sample. The male speech sample in Test A was 279 words, 2 min 34 s long; it was about the speaker’s research on computer speech recognition. The female speech sample in Test B was 307 words, 2 min 40 s long, and it concerned the college entrance examinations and the college admission process in the speaker’s native country. For the written recall section, participants listened to one of the speech samples and then wrote down what they remembered about its content; afterward, they took a multiple-choice comprehension test on the same sample. There was a multiple-choice test with 10 four-option questions for each speech sample.

The final section of each test assessed participants’ ability to discriminate a Turkish accent from other types of accents. The discrimination section consisted of 10 different speakers saying the sentences *This is my mother* and *Where are you*
Participants indicated whether or not the speaker of the sentences had the same accent as the speaker they had just heard in the first three parts of the test. None of the 10 speakers were the same as the male and female speakers in the first parts of the test; four were native Turkish speakers, and six were native speakers of other languages, including German, Japanese, English, Hindi, Thai, and Mandarin. Although these six accent types were used in both the A and B versions of this test section, none of the speakers were the same in the two test versions, and the order of presentation of the accent types in A was different from the order in B.

The test questions were presented on a computer screen in an Excel file with protected cells so that the participants could mark only in the appropriate answer spaces. For the transcription and written recall sections of the tests, participants typed their answers in text boxes. For the multiple-choice and discrimination sections, participants used the mouse to click the check box next to their selected answer.

Design and Procedure

Participants were tested in groups of 2 – 14 participants. All were tested in the same computer laboratory, and each participant had a Macintosh computer and a set of headphones. Before the experiment began, they filled out a questionnaire about their experience in traveling and in conversing with non-native English speakers. The experiment had three phases: pretest, training, and posttest. The entire procedure took 2 – 2.5 hr.

The pretests and posttests involved listening to the speech samples described in the Materials subsection and typing in responses in an Excel document on the
computer. Participants were instructed how to open the audio files and toggle into the Excel file window in order to listen to the transcription sentences and then type the sentences. They used the RealOne audio player on the Macintosh to listen to the files through their headphones. They were told to listen to the entire comprehension speech sample, and then type what they could remember in the textbox provided for Part 2 of the test. Their instructions were to write as if they were the speaker giving the same speech to a new audience, giving the gist of the speech and as many details as they could remember. For the discrimination section, they opened a new audio file and again toggled back to the Excel document to mark their answers as they listened to the speech samples.

The presentation order of samples for the pretests and posttests was counterbalanced, with half of the participants in each condition receiving Test A for the pretest and Test B for the posttest, and the other half receiving Test B for the pretest and Test A for the posttest.

During the training phase, half of the participants in each condition (48 total) received limited training, listening to three Turkish-accented speech samples, and half received more extensive training, listening to five Turkish-accented samples. In both types of training, participants took a 5 min break after the second training sample before finishing the training and taking the posttest. The extensive training lasted about 1 hour, including the break; the limited training lasted about 40 min, including the break. The first presentation of each sample was the regular version, which had no pauses, and the second presentation was the paused version, which was divided.
into phrases, with pauses after each phrase. The order of presentation of the speech samples was the same for all participants.

After all the participants in a group had finished the pretest, they received instructions as a group about what to do during training. The two tasks for the training phase were imitation and paraphrase. Half of the participants (48) performed each type of task. Thus, there were 24 participants in each training condition: limited/imitation, limited/paraphrase, extensive/imitation, extensive/paraphrase. In a given experiment group, all participants were in the same training condition.

For the imitation task, participants listened to the first presentation of the speech samples, and during the pauses in the second presentation, they were told to imitate aloud as closely as possible the phrases they heard, using the same words and the same pronunciation that the speaker used. For the paraphrase task, the participants also listened to the first presentation, and then they were instructed to paraphrase aloud each phrase during the pauses after the phrases, using their own words to express the meaning of the phrase. The experimenter remained in the room during the entire experiment in order to resolve any technical difficulties the participants had and in order to monitor the participants during training to ensure that they completed the training tasks as instructed.

Results

Scoring Procedures

The different sections of the tests were scored in the following ways. For the multiple-choice and discrimination sections of the tests, answers were correct if the
box next to the correct answer was checked; answers were incorrect if the correct box was not checked, or if more than one box was checked, even if the correct option had been selected as one of the checked boxes.

For the transcription section, an exact word match technique was used. The number of points possible for a given sentence was equal to the number of words in the original sentence. The participants’ transcriptions were compared to the original sentences, and one point was deducted if a word was missing or if an additional word had been inserted. Also, one point was deducted for substitutions of a new word for something in the original sentence, such as transcribing valley for the original word goalie. No points were deducted for misspelled words or for homophones, such as flower for the original word flour. See Appendix G for a complete account of the transcription scoring rubric.

Finally, the written recall scores were generated by Latent Semantic Analysis (LSA), a computational technique that represents the semantic content of a text as a vector in a multi-dimensional semantic space. The method and rationale behind LSA are summarized in Landauer, Foltz, and Laham (1998); for the analysis of the written recall, a one-to-many comparison was used, with the original transcript compared to the many recall paragraphs. Basically, the similarity between the original transcript and a recall paragraph is expressed as the cosine of the angle between the two vectors representing the semantic content of each text. An exact match produces a cosine of 1.00, and the less similarity there is between texts, the smaller the cosine. The semantic space for the comparisons was a database of general reading up to first year college, with 300 dimensions. So that the data could be appropriately scored by LSA,
the written recall was spell-checked, abbreviations were converted to their full-word form, and numbers were spelled out. Incorrect spellings which did not correspond obviously to an English word were left without correction.

**Pretest Comparisons**

In order to determine whether pretest scores were comparable across the groups of participants in the four conditions, a 2 x 2 analysis of variance compared the means of pretest scores between conditions. The two between subjects factors were training amount (limited, extensive) and training task (imitation, paraphrase). There were no significant main effects, and neither was the interaction between training amount and training task significant. On the pretest, then, participants in a given condition did not perform significantly better or worse than those in the other conditions.

However, according to unpaired t-tests, there were significant differences between the pretest scores of Test A and those of Test B for the transcription, $t(94) = 7.02, p < .001$, multiple-choice $t(94) = 4.59, p < .001$, and accent discrimination, $t(94) = -2.86, p = .005$, sections, as well as a marginally significant difference for the recall section $t(94) = 1.90, p = .060$ (see means in Table 10). Therefore, participants scored lower when listening to the speech samples of the male speaker in Test A than when listening to the female speaker in Test B.
Table 10. Experiment 2: Raw Pretest Score Means.

<table>
<thead>
<tr>
<th>Test Section: Raw Score Computation</th>
<th>Test A (male speaker)</th>
<th>Test B (female speaker)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SEM</td>
</tr>
<tr>
<td>Transcription: number correct out of 67</td>
<td>45.58</td>
<td>0.95</td>
</tr>
<tr>
<td>Recall: cosine between 0 and 1</td>
<td>.65</td>
<td>0.02</td>
</tr>
<tr>
<td>Multiple-choice: number correct out of 10</td>
<td>4.27</td>
<td>0.22</td>
</tr>
<tr>
<td>Accent Discrimination: number correct out of 10</td>
<td>6.58</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Unlike the transcription, multiple-choice, and recall sections, the accent discrimination scores were higher for Test A than Test B. The accent discrimination section was not based on the same speakers heard in the transcription, multiple-choice, and recall sections. Therefore, the disanalogy between the discrimination section and the other three sections is understandable. Because Test A and Test B had been counterbalanced as pre- and posttests, the discrepancy in test difficulty did not imply that differences between pre- and posttest scores could be due to inherent test difficulty. Nevertheless, this discrepancy did introduce noise into the data because the means and variances of Test A and Test B were different, so the scores were converted to z-scores for the statistical analysis. The z-scores created a mean of zero for all sections of the pretest for both Test A and Test B, and they also equalized the variances of the scores.

Analyses of Pre- and Posttest Z-scores

The pre- and posttest z-scores of each of the test sections were analyzed separately with a 2 x 2 x 2 repeated measures analysis of variance. The between
subjects factors were training amount (limited, extensive) and training task (imitation, paraphrase). The within subjects factor was test type (pretest, posttest).

For the written recall comprehension measure, there was a main effect of test type, $F(1, 92) = 9.63, \ p = .003$, such that $z$-scores on the posttest ($M = .40$) were higher than $z$-scores on the pretest ($M = .00$) across all conditions. The multiple-choice comprehension measure also suggested a trend of higher posttest $z$-scores ($M = .23$) than pretest $z$-scores ($M = .00$), although the effect was only marginally significant, $F(1, 92) = 3.04, p = .085$. See Figure 1 for a summary of the $z$-score means on both comprehension measures. Planned paired-comparison $t$-tests indicated a significant difference between pre- and posttest recall $z$-scores in the limited paraphrase condition, $t(23) = -2.77, p = .011$, and a marginally significant difference in the limited imitation condition, $t(23) = -1.85, p = .078$. No significant differences were found between pre- and posttest recall $z$-scores in either the extensive paraphrase or extensive imitation conditions; multiple-choice pre- and posttest $z$-scores were not significantly different in any of the conditions according to the planned paired-comparison $t$-tests.
Figure 1. Experiment 2: Z-score means for written recall and multiple-choice comprehension measures on the pre- and posttests across conditions.

The transcription z-scores showed a main effect of test type, $F(1, 92) = 6.05$, $p = .016$, with posttest z-scores ($M = .28$) higher than pretest z-scores ($M = .00$) on average (see Figure 2). In addition, there was a significant interaction between test and task, $F(1, 92) = 5.06$, $p = .027$, such that pretest z-scores were similar for the two task types (imitation $M = -.04$; paraphrase $M = .04$), but the posttest z-scores were higher for the imitation task ($M = .49$) than for the paraphrase task ($M = .07$).

Planned paired-comparison $t$-tests comparing pre- and posttest scores indicated higher transcription posttest z-scores in the limited imitation condition, $t(23) = -2.87$, $p = .009$, and a trend for higher transcription z-scores in the extensive imitation condition, $t(23) = -1.72$, $p = .099$. Based on the planned paired-comparison $t$-tests, there were
no significant differences between pre- and posttest transcription z-scores in the limited or extensive paraphrase conditions.

![Graph showing z-score means for transcription measure on pre- and posttests across conditions.](image)

Figure 2. Experiment 2: Z-score means for the transcription measure on the pre- and posttests across conditions.

The accent discrimination test z-scores had a significant interaction between training amount and task, $F(1, 92) = 5.42, p = .022$ (see Figure 3). Overall test z-scores for limited training were below average for the imitation task ($M = -0.13$), but above average for the paraphrase task ($M = 0.22$). For extensive training, the opposite effect was seen, with z-scores below average for the paraphrase task ($M = -0.07$) and above average for the imitation task ($M = 0.35$). Planned paired-comparison $t$-tests indicated a significant difference between pre- and posttest z-scores for the extensive
imitation condition, $t(23) = -2.29, p = .031,$ which had the highest posttest $z$-scores of all the conditions.

Figure 3. Experiment 2: $Z$-score means for the accent discrimination measure on the pre- and posttests across conditions.

Discussion

The results of this training experiment demonstrate that the imitation and paraphrase training tasks are effective in enabling participants to better comprehend non-native accented speech. The differences in results on the various test measures in the four conditions additionally provide insights into the nature of the training and testing methodologies.

The general success of the training for improving comprehension is evidenced by the main effect of test type for the recall and transcription measures, as well as a marginal main effect of test type for the multiple-choice measure. These main effects
indicate that posttest scores were higher than pretest scores across conditions, which implies that participants’ understanding of accented speech increased as a result of training. However, not all conditions were equally effective in promoting better comprehension, as the paired-comparison t-tests and the interaction for transcription indicate. For the written recall measure, the significant difference between pre- and posttest scores in the limited paraphrase condition suggests that this condition provided the best training in the comprehension of long speech samples. The written recall tested participants’ ability to understand and remember a speech that was 2 to 3 min long. The paraphrase task was designed to direct participants’ attention to the meaning of the speech samples during training, thus encouraging them to practice listening for the message of the speech. It was therefore not surprising that participants who performed a limited version of the paraphrase task during training improved significantly on the recall comprehension measure. Nevertheless, a trend toward improvement in the limited imitation condition may mean that the imitation task also enabled better comprehension and higher recall scores.

Although transcription has been used to measure comprehension in previous studies (Derwing et al., 2002; Weil, 2001), it emphasizes the ability to distinguish discrete segments with little context and de-emphasizes the ability to recall the gist of a speech, unlike the written recall measure. For the transcription measure, paired-comparison t-tests confirmed what the significant interaction between test and task pointed to: Posttest scores were higher than pretest scores for participants who had performed the imitation task during training, but not for those who had performed the paraphrase task. The imitation task was intended to prompt participants to pay
attention to segmental differences in the pronunciation of accented speech, which would allow participants to recognize more easily individual words with little or no context. The transcription test seems to measure this ability to identify more or less isolated segments of speech, and as a result the imitation task proved successful in enabling better performance on the transcription section of the posttest.

The accent discrimination test scores did not show a significant effect of test type, but there was an interaction between training amount and task, and a paired-comparison $t$-test confirmed that the posttest score was significantly higher than the pretest score in the extensive imitation condition. In order to perform well on the accent discrimination test, participants had to be able to recognize when the characteristics of a speaker’s accent were similar to the characteristics of another speaker’s accent. The amount of speech that participants heard in order to make this judgment was quite limited—only two short sentences. Nevertheless, participants in the extensive imitation condition were able to do well on the accent discrimination posttest. Imitating the speech seems to have heightened their awareness of the characteristics of a Turkish accent, and the exposure to a variety of native Turkish speakers, along with the subtle differences in accents between these speakers, allowed participants to score well on the accent discrimination posttest when they were tested on the speech of speakers they had not previously heard.

Finally, the general lack of improvement of test scores for the extensive paraphrase condition, as evidenced by the non-significant paired-comparison $t$-test results, suggests that participants may have become fatigued by the paraphrase task when they had extensive training. For the paraphrase task, participants were
instructed to think of synonyms for all of the speaker’s words if possible, and this task seems to have been more difficult than the imitation task, in which participants simply repeated the exact phrases the speaker used. Thus, participants in the extensive imitation condition did not seem to suffer from fatigue in the same way that those in extensive paraphrase condition did.
Experiment 3 was designed to examine the durability of training and to find evidence for the transferability of training. Experiment 2 indicated that the imitation and paraphrase training tasks improved comprehension of accented speech, but if this improvement quickly disappeared after the end of training, then it would not ultimately be useful. If the training is to have practical implications and applications, the improvement in comprehension must be durable. As a result, the principal object of the present training experiment was to test the durability of training.

The other consideration of this experiment was whether participants would be able to transfer the implicit knowledge and skills they gained through training to understanding a different type of accented speech. Like durability, transferability is a concern for practical reasons: It would be impractical, if not impossible, to train people on all the different types of accents they might encounter in their lives. For training to be successful in helping people understand accented speech in general, not just one particular type of accent, it must be transferable. However, previous research indicates that transferability is unlikely because implicit knowledge often fails to transfer to a new situation, even when the participant is aware of the parallels between the original and transfer tasks (Dienes & Berry, 1993). Also, durability is usually linked to limited generalizability, so that training tasks which result in durable retention also tend to produce highly specific skills or knowledge (Healy & Bourne,
In other words, durable implicit knowledge has two strikes against its likelihood of being transferable.

The design of this experiment was similar to that of Experiment 2, using the same Turkish-accented speech samples for training and the same format for the tests. Participants were trained only on Turkish-accented speech, but the test versions included Thai-accented speech samples as well as Turkish-accented samples in order to test transferability. The Thai-accented speech had been rated as less comprehensible and more accented; in consequence, any transfer that was exhibited would be from a less difficult task to a more difficult task, thus working against the null hypothesis. Because participants seemed to become fatigued in the extensive paraphrase condition of Experiment 2, training for the present experiment was divided into two sessions, each 1 week apart. Finally, a retention test, given 1 week after the second training session, provided evidence for durable retention of the skills and knowledge gained during training.

Method

Participants

Seventy-two native English speakers participated; all were undergraduate students at the University of Colorado who received course credit for their participation. None of the participants reported having speech or hearing dysfunctions. Most participants (61) reported having some contact with non-native English speakers, primarily with native Spanish speakers (51), but also commonly with French (16), German (15), Italian (11), Chinese (11), Korean (9), Russian (8),
and Japanese (7) speakers. One participant reported contact with native Turkish
speakers, and two reported contact with native Thai speakers. Most participants (57)
also indicated that they had traveled abroad. The most popular travel destinations
included Central America and Europe, but some participants had also traveled in
South America, Asia, Africa, the Caribbean, and Australia. However, none reported
travel to Thailand or Cyprus, and only one had been to Turkey.

Materials

The five training samples in this experiment were the same as those used in
the extensive training of Experiment 2. Thus, all the samples of the Turkish-accented
speech had the same regular and paused versions as before (see Table 9 in
Experiment 2 and Appendix D).

The testing materials had the same format as the materials used in Experiment
2, but additional tests were developed so that the durability and transferability of
training could be addressed. Like the tests in Experiment 2, the additional tests
included speech samples from the interviews described in the Speech Sample
Collection section as well as sentences from the Duke University corpus (Arslan &
Hansen, 1997), and they had four sections: a) transcription, b) written recall, c)
multiple-choice comprehension, and d) accent discrimination (see Appendix F for a
sample test). One of the new tests used Turkish-accented speech samples from a
male speaker in the first three test sections; the other three additional tests used Thai-
accented speech samples from one male and two female speakers in the first three test
sections (see Appendices C and E). Altogether, including the new samples and those
presented in Experiment 2, there was a total of three Turkish speakers and three Thai speakers (see Table 11). The Duke University database provided the sentences used in the final accent discrimination section of the test.

The tasks for the four test sections were identical to those in Experiment 2, namely, transcribing sentences, writing the content of a longer speech sample from memory and taking a multiple-choice test on that sample, and deciding which of 10 unfamiliar speakers had the same accent as the speaker in the first three test sections. For the transcription section, the sentences were 11 – 16 words long and had a total 65 – 67 words for a given set of sentences. The speech samples for the recall and multiple-choice comprehension sections of the test had 279 – 329 words and were between 2 and 3 min long (see Table 11 for details). As in Experiment 2, the accent discrimination section of the additional Turkish-accented test consisted of the sentences *This is my mother* and *Where are you going?* spoken by four native Turkish speakers and six native speakers of other languages, including German, Japanese, English, Hindi, Mandarin, and Thai. The accent discrimination sections for the Thai-accented test versions were similar, except that there were four native Thai speakers and six native speakers of other languages, including German, Japanese, English, Hindi, Mandarin, and Turkish. None of the speakers in any of the accent discrimination test sections were the same as the speakers in training or testing; the speakers in one accent discrimination test were never used in another accent discrimination test.
Table 11. Experiment 3: Characteristics of the Speech Samples Used for Comprehension Measures.

<table>
<thead>
<tr>
<th>Test—Topic of Sample</th>
<th>Accent Type</th>
<th>Speaker Gender</th>
<th>Number of Words</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>A—speech recognition (also used in Experiment 2)</td>
<td>Turkish</td>
<td>male</td>
<td>279</td>
<td>2 min 34 s</td>
</tr>
<tr>
<td>B—college entrance exams (also used in Experiment 2)</td>
<td>Turkish</td>
<td>female</td>
<td>307</td>
<td>2 min 40 s</td>
</tr>
<tr>
<td>C—highway expansion project</td>
<td>Turkish</td>
<td>male</td>
<td>307</td>
<td>2 min 43 s</td>
</tr>
<tr>
<td>D—photography</td>
<td>Thai</td>
<td>male</td>
<td>299</td>
<td>2 min 38 s</td>
</tr>
<tr>
<td>E—computer network security</td>
<td>Thai</td>
<td>female</td>
<td>329</td>
<td>2 min 28 s</td>
</tr>
<tr>
<td>F—computer-animated books</td>
<td>Thai</td>
<td>female</td>
<td>302</td>
<td>2 min 51 s</td>
</tr>
</tbody>
</table>

As in Experiment 2, the test questions were presented on a computer screen in an Excel file with protected cells so that the participants could mark only in the appropriate answer spaces.

**Design and Procedure**

Participants attended three sessions for this experiment, each session 1 week apart. In the first session, they took a pretest and began training, in the second session they finished training and took a posttest, and in the final session they took a retention test. The first two sessions lasted about 1 hr 45 min, and the last session lasted about 45 min. All sessions took place in the same computer laboratory, and participants were tested in groups of 9 – 16 participants. Each participant had a Macintosh computer and a set of headphones. Before the first session began, they filled out a questionnaire about their experience in traveling and in conversing with non-native English speakers.

The same testing procedure was used as in Experiment 2, except that participants took two tests each, one Turkish-accented and one Thai-accented, for the
pre-, post-, and retention tests. The presentation order of the tests was
counterbalanced with a Latin Square so that one-third of the participants received a
given test for a pre-, post-, or retention test. The Turkish- and Thai-accented tests
were presented in fixed pairs: A/F, B/D, and C/E; however, the order of presentation
for these pairs was counterbalanced, with half of the participants taking the Turkish-
accented test first and half taking the Thai-accented test first.

During training, all participants listened to five speech samples. They listened
to the first two samples during the first session and the last three samples during the
second session. As in the training for Experiment 2, the first presentation of each
sample was the regular version, which had no pauses, and the second presentation
was the paused version, which was divided into phrases, with pauses after each
phrase. The order of presentation of the speech samples was the same for all
participants.

After all the participants in a group had finished the pretest, they received
instructions as a group about what to do during training. The three conditions for the
training phase were imitation, paraphrase, and control, and one-third (24) of the
participants were in each condition. The imitation and paraphrase tasks were
identical to those in Experiment 2, and participants received the same instructions
about how to perform the tasks. For the control condition, participants simply
remained silent during the pauses in the second presentation of the speech samples;
they were told to listen to the samples and to wait silently during the pauses for the
next phrase to be spoken. In a given experiment group, all participants were in the
same training condition.
Results

Scoring Procedures

The scoring system employed in Experiment 2 was also used for the various sections of these tests, including the exact word match technique for the transcriptions and the use of LSA for the recall paragraphs. The transcription scores were converted to proportion correct because the total points possible varied slightly from one set of transcription sentences to another.

Pretest Comparisons

In order to determine whether pretest scores were comparable across the groups of participants in the three conditions, 2 x 3 repeated measures analyses of variance compared the means of pretest scores between conditions separately for each test section. The within subjects factor was accent type (Turkish, Thai), and the between subjects factor was condition (control, imitation, paraphrase). There were no main effects of condition or interactions between accent type and condition for any of the test sections except the recall section, which had a main effect of condition, $F(2, 69) = 4.02, p = .022$. The pretest recall scores for the control condition ($M = .74$) were higher than those for the imitation condition ($M = .71$), and the scores for the paraphrase condition were the lowest of all ($M = .68$). In other words, participants in a given condition did not perform significantly better or worse on the pretest than those in the other conditions, except in the case of the recall scores.

However, according to 2 x 3 repeated measures analyses of variance, there were significant differences for some of the pretest sections between the different
versions of the tests (see means in Table 12). The within subjects factor was accent type (Turkish, Thai), and the between subjects factor was test version (preA/F, preB/D, preC/E), which represented the pair of tests a participant took for the pretest. Significant interactions between accent type and test version were found for the multiple-choice, $F(2, 69) = 6.87, p = .002$, and recall sections, $F(2, 69) = 4.90, p = .010$. These interactions indicated that there were differences between the pretest scores on the three test versions, but the differences did not follow the same direction for the Turkish-accented and Thai-accented tests within the pairs. For the Turkish-accented tests, it appears that participants scored lowest on the multiple-choice sections of Test A, higher on Test B, and highest on Test C. The recall scores have a different order, with scores lowest on Test B, higher on Test C, and highest on Test A. For the Thai-accented multiple-choice section, the ordering of scores from lowest to highest is Test E, Test F, and Test D; the ordering from lowest to highest for the recall sections is Test F, Test E, and Test D.

The transcription section had a main effect of test, $F(2, 69) = 12.64, p < .001$, such that the differences among the three test versions were similar for Turkish- and Thai-accented tests that composed each test pair. The preA/F pair received the lowest scores, the preB/D pair had middle scores, and the preC/E pair received the highest scores. No significant differences were seen between the accent discrimination pretest scores.
Table 12. Experiment 3: Raw Pretest Score Means.

<table>
<thead>
<tr>
<th>Test Section: Raw Score Computation</th>
<th>Transcription: Proportion Correct</th>
<th>Recall: Cosine Between 0 and 1</th>
<th>Multiple-choice: Number Correct out of 10</th>
<th>Accent Discrimination: Number Correct out of 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>A— Turkish</td>
<td>Mean .67</td>
<td>.74</td>
<td>4.67</td>
<td>6.33</td>
</tr>
<tr>
<td></td>
<td>SEM .03</td>
<td>.02</td>
<td>.36</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>Variance .02</td>
<td>.01</td>
<td>3.19</td>
<td>2.84</td>
</tr>
<tr>
<td>B— Turkish</td>
<td>Mean .71</td>
<td>.70</td>
<td>6.08</td>
<td>5.92</td>
</tr>
<tr>
<td></td>
<td>SEM .04</td>
<td>.02</td>
<td>.39</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>Variance .05</td>
<td>.01</td>
<td>3.73</td>
<td>2.78</td>
</tr>
<tr>
<td>C— Turkish</td>
<td>Mean .86</td>
<td>.71</td>
<td>6.46</td>
<td>6.38</td>
</tr>
<tr>
<td></td>
<td>SEM .02</td>
<td>.01</td>
<td>.40</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td>Variance .01</td>
<td>.01</td>
<td>3.83</td>
<td>2.16</td>
</tr>
<tr>
<td>D— Thai</td>
<td>Mean .74</td>
<td>.74</td>
<td>6.29</td>
<td>5.75</td>
</tr>
<tr>
<td></td>
<td>SEM .02</td>
<td>.02</td>
<td>.44</td>
<td>.33</td>
</tr>
<tr>
<td></td>
<td>Variance .01</td>
<td>.01</td>
<td>4.56</td>
<td>2.54</td>
</tr>
<tr>
<td>E— Thai</td>
<td>Mean .80</td>
<td>.68</td>
<td>5.21</td>
<td>6.46</td>
</tr>
<tr>
<td></td>
<td>SEM .02</td>
<td>.02</td>
<td>.35</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td>Variance .01</td>
<td>.01</td>
<td>2.87</td>
<td>1.56</td>
</tr>
<tr>
<td>F— Thai</td>
<td>Mean .69</td>
<td>.68</td>
<td>5.46</td>
<td>6.13</td>
</tr>
<tr>
<td></td>
<td>SEM .03</td>
<td>.02</td>
<td>.41</td>
<td>.23</td>
</tr>
<tr>
<td></td>
<td>Variance .02</td>
<td>.01</td>
<td>4.08</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Although not all sections of all the test versions had scores significantly different from each other, the raw scores were converted to z-scores for the statistical analyses in order to equalize the variances and create a mean of zero for all versions of the pretest. This conversion follows the precedent set in Experiment 2.

**Analyses of Pre-, Post-, and Retention Test Z-scores**

The pre-, post- and retention test z-scores of each of the test sections were analyzed separately with 3 x 3 x 2 repeated measures analyses of variance. The between subjects factor was training task (imitation, paraphrase, control). The within
subjects factors were test type (pretest, posttest, retention test) and accent type (Turkish, Thai).

For the multiple-choice comprehension measure, an interaction between test type and accent was evident, but not significant, $F(2, 138) = 2.38$, $p = .097$. The trend suggested that pretest scores were lowest for both the Turkish- and Thai-accented tests ($M = .00$ for both). However, for the Turkish-accented test, the retention test scores ($M = .28$) were higher than the posttest scores ($M = .03$), whereas the opposite was seen with the Thai-accented scores, with higher posttest scores ($M = .27$) and retention test scores ($M = .06$). See Figure 4 for a summary of the multiple-choice z-score means.

![Figure 4. Experiment 3: Z-score means for the multiple-choice comprehension measure on the pre-, post-, and retention tests across conditions, with scores on the Turkish-accented tests separate from scores on the Thai-accented tests.](image)

Planned paired-comparison $t$-tests indicated a marginally significant difference between pre- and posttest z-scores for the Thai-accented multiple-choice
section in the imitation condition, \( t(23) = -1.94, p = .065 \), thus somewhat confirming the trend suggested by the analysis of variance.

There were no significant differences among recall scores, as can be seen by the lack of discernable meaningful patterns for the means presented in Figure 5.

Figure 5. Experiment 3: Z-score means for the recall comprehension measure on the pre-, post-, and retention tests across conditions, with scores on the Turkish-accented tests separate from scores on the Thai-accented tests.

The transcription z-scores showed a main effect of test type, \( F(2, 138) = 4.27, p = .016 \), with retention test z-scores \((M = .36)\) higher than posttest z-scores \((M = .06)\), which were higher than pretest z-scores \((M = .00)\) averaged across conditions (see Figure 6). There was also a marginally significant interaction between accent and test, \( F(2, 138) = 2.39, p = .096 \), such that pretest z-scores were the same for the two accent types \((M = .00)\), but the z-scores for the Thai-accented tests tended to improve slightly across the posttest \((M = .13)\) and retention test \((M = .25)\), whereas the z-
scores for the Turkish-accented test did not improve on the posttest ($M = -.01$), but they showed a large improvement on the retention test ($M = .47$). Planned paired-comparison $t$-tests comparing pretests to posttests and retention tests indicated higher transcription $z$-scores on the retention test in the imitation condition for the Turkish-accented tests, $t(23) = -3.48$, $p = .002$, and a trend toward higher transcription $z$-scores on the retention test for Thai-accented tests in the both the imitation condition, $t(23) = -1.77$, $p = .090$, and the paraphrase condition, $t(23) = -2.03$, $p = .054$. According to the planned paired-comparison $t$-tests, there were no significant differences between transcription test scores in the control condition.

![Figure 6](image.png)

**Figure 6.** Experiment 3: $Z$-score means for the transcription section on the pre-, post-, and retention tests across conditions, with scores on the Turkish-accented tests separate from scores on the Thai-accented tests.
The accent discrimination test z-scores had a significant interaction between test type and condition, $F(4, 138) = 2.53, p = .043$ (see Figure 7). Averaging across accent type, z-scores in the paraphrase condition were higher on the posttest ($M = .14$) than on the pretest ($M = -.02$), but were lowest on the retention test ($M = -.33$). In contrast, z-scores in the control condition were lower on both the posttest ($M = -.16$) and retention test ($M = -.12$) than they were on the pretest ($M = .15$). In the imitation condition, z-scores decreased on the posttest ($M = -.37$) compared with the pretest ($M = -.13$), but increased on the retention test ($M = .04$). Planned paired-comparison $t$-tests indicated a significant difference between pre- and retention test z-scores for the Thai-accented tests in the paraphrase condition, $t(23) = 2.31, p = .030$, signifying a decrement in performance from the pretest to the retention test.

![Figure 7](image_url)

Figure 7. Experiment 3: Z-score means for the accent discrimination section on the pre-, post-, and retention tests across conditions, with scores on the Turkish-accented tests separate from scores on the Thai-accented tests.
Discussion

Durability and Transferability of Training

This experiment offers evidence that even a short amount of training, just over 1 hour long, can improve listeners’ comprehension of accented speech, and that this improvement is durable. Also, the results may indicate that training on one type of accent can transfer to another type of accent and improve comprehension for that accent as well.

The durability of training is demonstrated by the main effect of test type for transcription scores, with retention test scores higher than posttest scores, which were higher than pretest scores. This main effect shows that the higher post- and retention test scores occurred across conditions and across accent type, suggesting a general, durable benefit of training on participants’ ability to understand accented speech, at least in the limited context of a transcription task. Nevertheless, a marginally significant interaction between accent and test, as well as paired-comparison t-tests, imply that the benefit of training was not equal across conditions and accent types. The numerically highest mean transcription z-score occurred for the Turkish-accented retention test in the imitation condition, and a paired-comparison t-test verified that this mean retention test z-score was significantly higher than the mean pretest z-score. Trends for higher retention test scores were also seen for Thai-accented speech in the imitation and paraphrase conditions, according to the marginally significant interaction between test and accent, and marginally significant paired-comparison t-tests. Thus, although comprehension of Thai-accented speech seemed to benefit from
either type of training task, comprehension of Turkish-accented speech clearly improved for the imitation task.

An indication of improved comprehension in more naturalistic contexts comes from the multiple-choice comprehension scores. Unlike the transcription task, which measured participants' ability to understand isolated sentences and write them down verbatim, the multiple-choice test measured participants' memory of specific information from a relatively long speech sample. For participants to remember this information, they first had to understand it, but they did not necessarily need to understand every word of the speech in order to grasp the information it conveyed. The marginally significant interaction between test type and accent implies that comprehension of both Turkish-accented and Thai-accented speech improved, although the retention test z-scores were the highest for Turkish-accented speech and the posttest z-scores were the highest for Thai-accented speech. Training on Turkish-accented speech thus seems to have benefited comprehension not just for Turkish-accented speech, but for Thai-accented speech as well, although that transfer benefit was not durable.

_Evidence for Implicit Learning from the Accent Discrimination Scores_

The significant interaction between test type and condition for the accent discrimination test z-scores indicates that scores across accent type did not improve in the control condition, but in the paraphrase condition, they improved on the posttest, though they dropped on the retention test, and in the imitation condition, scores dropped on the posttest but improved on the retention test. This interaction may
result from the combination of the training task and the test, which differentially affected participants in the different conditions.

For the control condition, no differential effect on post- and retention test scores was seen, and no benefit of training was evident. In fact, exposure to accented speech may have caused a decrement in performance for participants in the control condition, perhaps because they were consciously attempting to judge the accent characteristics, and this conscious attention interfered with any implicit knowledge they had gained about accented speech during training.

However, in the paraphrase condition, performing the paraphrase task immediately before taking the posttest apparently enabled participants to do better on the accent discrimination test, although this improvement was not significant by paired-comparison $t$-tests. Perhaps these participants had acquired an implicit sense of accented speech characteristics, and because their recent training had encouraged them to focus on content rather than accent, they were less inclined to invoke explicit knowledge and judgments for the discrimination test. At any rate, whatever advantage these participants had from the combination of their paraphrase training condition and the task of discriminating between accents seems to have disappeared on the retention test.

Participants in the imitation condition, on the other hand, did not exhibit any immediate benefit of training, but their scores did improve on the retention test, although, again, this improvement was not significant by paired-comparison $t$-tests. It may be that they gained implicit knowledge about accented speech characteristics by performing the imitation task, but the accent discrimination test prompted them to
think consciously and explicitly about the accents of the speakers they were hearing, as they had just finished focusing on accent characteristics in order to imitate them.

By 1 week later when these participants took the retention test, the effect of the imitation training task, directing their conscious attention to the accent characteristics, seems to have gone away, but the implicit knowledge they gained during training may have enabled their improved performance on the retention test. The benefit of training on ability to discriminate accent types thus appears to be more durable for the imitation condition than it is for the paraphrase condition.

*Written Recall Measure*

The scores for the written recall test section did not yield any predictable pattern, which could be due to a couple of factors. First, free recall has been recognized as a potentially less sensitive comprehension measure than recognition tests (Zechmeister & Nyberg, 1982), and the differences between what participants understood and remembered from the speech samples may have been too subtle for the written recall to reflect. Second, the scores generated by LSA may not have sufficiently captured differences in quality between the different participants’ recall paragraphs. LSA calculates a vector for each text, which represents the text’s semantic content, and then compares the vectors for two texts and produces a cosine representing the similarity of both texts’ semantic content. Using a propositional analysis to score the texts instead of LSA might provide a more detailed picture of information contained in a given recall paragraph and allow differences between recall scores to become evident.
Measuring comprehension, whether of texts or of spoken utterances, entails some inexactness in determining how much of the content was truly understood. The recall and multiple-choice comprehension measures in the present research tested not just understanding of the speech sample, but also ability to remember the content. The two functions are obviously related, in that it is difficult to remember what one has not understood, but people generally comprehend more than they can remember. Free recall demonstrates the information that the participant can remember well enough to articulate, but not the information the participant remembers well enough only to recognize. Forced-choice tests account for information the participant can recognize, but they also may not be exhaustive enough to capture all that the participant remembers.

To some extent, these problems were avoided by the transcription test, as the amount of information participants had to remember was quite small, and their memory was not taxed because they typed each sentence immediately after listening to it. Minimizing the role of memory is one reason the transcription task should be more sensitive than the other comprehension measures to the effect of accent on listeners' ability to perceive speech. Another reason for the greater sensitivity of the transcription task is that there are fewer contextual clues about content in single, isolated sentences than in a discourse of two or three minutes. Listeners are therefore less equipped in a transcription task to decipher the meaning based on syntactic and
semantic context. Of course, these sources for the sensitivity of the transcription test also detract from its viability as a measure of comprehension: Spoken communication generally occurs in a rich context of semantic and syntactic information, and participating in a conversation or listening to a lecture entails remembering what a speaker says for more than a few seconds in order to effectively follow the speaker’s train of thought.

Because of the issues described here, a final experiment was designed to determine how well participants could understand unaccented versions of the speech samples used in the training experiments. A native English speaker read transcripts of the speech samples, thus eliminating the accent but preserving the lexical choices, grammatical patterns, and content of the samples (see Appendices C and E). Participants in the native English experiment took the same tests used Experiments 2 and 3, and these scores were compared to the scores from Experiments 2 and 3. The results of the native English experiment show what the maximum score would be if training were perfectly successful in enabling listeners to understand accented speech as well as they could understand unaccented speech. In addition, the comparison of the native English scores to the scores from the two training experiments gives insight into how much the accent, rather than other linguistic factors, contributed to the comprehensibility of the accented speech, as measured by the tests, and the extent to which training aided comprehension.
Method

Participants

Twenty-four native English speakers participated in Experiment 4, the native English experiment, none of whom had participated in the first three experiments. All were undergraduate students at the University of Colorado who received course credit for their participation. None reported speech or hearing dysfunctions, and most (19) reported some contact with non-native English speakers, mainly with native Spanish speakers (18), but also commonly with German (8), French (6), and Japanese (5) speakers. No participants reported contact with native Turkish or Thai speakers. Most participants (19) also indicated that they had traveled abroad. The most popular travel destinations included Central and South America and Europe, but some participants had also traveled in Asia, Africa, the Caribbean, and Australia. However, none reported travel to Thailand, Cyprus, or Turkey. Thus, the background and experiences of the participants in this experiment were similar to those of participants in Experiments 2 and 3.

Materials

The testing materials were the same as the six tests used in Experiment 3, with two exceptions. First, the transcripts of the original accented speech samples were read by a native English speaker (see Appendix D). The reader was a woman who had a neutral General American accent; she had lived in the Midwest and West most of her life, including Arizona, Colorado, Minnesota, Nevada, North Dakota, and Wyoming, so her speech was not marked by regional dialect characteristics. A Telex
M-560 ISB microphone was used to record her speech in Cool Edit on a Compaq computer.

The second difference from the tests in Experiment 3 was that the participants took only the first three sections of the tests, involving the transcription sentences speech samples and the long comprehension passage speech sample. The fourth section of the tests, the accent discrimination section, was not given to the participants in this experiment. (It would have been impossible for these participants to take the accent discrimination section, which required a comparison of the accented speech from the first three sections of the test to the accents of the speakers in the accent discrimination test. These participants heard unaccented speech for the first three sections of the test and therefore could not make such comparisons.)

To ensure a naturalistic reading of the speech, the reader rehearsed each sample before recording it. The transcripts included all the original disfluencies and hesitation words, such as *um* and *uh*, as well as grammatical errors, such as lack of subject-verb agreement. Because the sentences for the transcription tests had originally been extracted from the continuous discourse of the interview, the reader read transcripts of these sentences which included a sentence or two of context before and after the target sentence. The target sentences were then cut out using Cool Edit, similarly to how the accented versions of the sentences had originally been cut out of their larger context. This step was taken to encourage the reader to produce the sentences with the fluency of contextualized utterances rather than with carefully enunciated speech, which is typical of isolated words and sentences.
Design and Procedure

Participants were tested in groups of 12 – 14 participants. All testing took place in the same computer lab, and each participant had a Macintosh computer and a set of headphones. Before the experiment began, they filled out a questionnaire about their experience in traveling abroad and in conversing with non-native English speakers. The experiment lasted about 2 hours, with one 5-min break in the middle.

This experiment involved no training. Participants took the six versions of the tests used in the training experiments, but because they did not participate in training, the tests could not be designated as pre-, post-, or retention tests. Instead, the tests are considered according to their version letter. Details on these versions are given in Experiments 2 and 3, but the accent of the original speaker and the topics of the long comprehension passages in each version are repeated here: Turkish accented tests: A—speech recognition, B—college entrance exams, C—highway expansion project; Thai-accented tests: D—photography, E—computer network security, F—computer-animated books.

As in the training experiments, participants typed their responses to the test questions in Excel documents on the computer. Participants received the same instructions that had been given during the training experiments about how to navigate between the Excel document and the audio files, and they were told the same instructions about how to take the different sections of the tests. All participants took all six tests, but the order of presentation was counterbalanced across participants using a Latin Square, resulting in six orders of presentation with each test occurring once in each position (first, second, third, etc.).
Results

Native English Data

Separate one-way analyses of variance were performed on the scores of each test section to compare the scores for the six versions of the test. Scoring procedures followed the same guidelines as those of the two training experiments, Experiments 2 and 3. For the multiple-choice and transcription sections, scores represented the proportion correct. The recall scores were generated by LSA as the cosine of two angles representing the semantic content of a text.

Significant differences between the scores of the various test versions were found for the multiple-choice section, $F(5, 115) = 6.95, p < .001$, and for the transcription section, $F(5, 115) = 7.34, p < .001$, but not for the recall section. The test version means for all three sections are presented in Table 13, including the recall section despite the lack of significant differences between the recall scores. The multiple-choice scores, in ascending order, were versions A, E, C, B, F, and D. The transcription scores were nearly perfect, with proportions correct at or above 95%. These scores also had very little variance, and in part due to the lack of variance, even the small numerical differences between the test means were significant. For the transcription section, Test C had the lowest mean score, followed by Test A and Test B; Tests D, E, and F shared the highest mean score.
Table 13. Experiment 4: Native English Mean Test Scores for Each Version of the Test.

<table>
<thead>
<tr>
<th>Test version</th>
<th>Accent of Original Speaker</th>
<th>Multiple-Choice Mean</th>
<th>Multiple-Choice SEM</th>
<th>Recall Mean</th>
<th>Recall SEM</th>
<th>Transcription Mean</th>
<th>Transcription SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>A—speech recognition</td>
<td>Turkish</td>
<td>.50</td>
<td>.03</td>
<td>.69</td>
<td>.03</td>
<td>.96</td>
<td>.01</td>
</tr>
<tr>
<td>B—college entrance exams</td>
<td>Turkish</td>
<td>.68</td>
<td>.03</td>
<td>.73</td>
<td>.01</td>
<td>.97</td>
<td>.01</td>
</tr>
<tr>
<td>C—highway expansion project</td>
<td>Turkish</td>
<td>.65</td>
<td>.04</td>
<td>.73</td>
<td>.02</td>
<td>.95</td>
<td>.01</td>
</tr>
<tr>
<td>D—photography</td>
<td>Thai</td>
<td>.73</td>
<td>.04</td>
<td>.76</td>
<td>.03</td>
<td>.98</td>
<td>.00</td>
</tr>
<tr>
<td>E—computer network security</td>
<td>Thai</td>
<td>.60</td>
<td>.03</td>
<td>.73</td>
<td>.02</td>
<td>.98</td>
<td>.01</td>
</tr>
<tr>
<td>F—computer-animated books</td>
<td>Thai</td>
<td>.70</td>
<td>.03</td>
<td>.74</td>
<td>.02</td>
<td>.98</td>
<td>.00</td>
</tr>
</tbody>
</table>

Native English Compared to Experiment 2

In Experiment 2, only test versions A and B were used, so the comparison of scores from Experiment 2 and the native English experiment (the “native” scores) only involves these two tests. The pre- and posttest scores from Experiment 2 are included in the analyses. Separate 3 x 2 repeated measures analyses of variance were done for the three test sections: multiple-choice, recall, and transcription. The between-subjects factor, group (native, preA, preB), designated the tests according to whether the participants had listened to the native English speech in the native English experiment (native), or whether they had taken Test A as a pretest and Test B as a posttest (preA) or Test B as a pretest and Test A as a posttest (preB) in Experiment 2. The within-subjects factor was test version (A, B).

Multiple-choice section. For the multiple-choice scores, a main effect of group, $F(2, 117) = 30.26, p < .001$, indicated that participants who took Test A as a
pretest (preA, $M = .43$) scored lowest on the multiple-choice tests, participants who took Test B as a pretest (preB, $M = .62$) scored the highest, and those who listened to native English speech (native, $M = .59$) scored a little lower than the preB group.

An interaction between group and test version, $F(2, 117) = 8.62, p < .001$, showed that participants who listened to native English speech scored the highest on the Test B multiple-choice section, but, interestingly, they did not score the highest on Test A (see Figure 8). Instead, participants in Experiment 2 who took Test A as a posttest after training (the preB group) scored the highest on the multiple-choice section of Test A. Those in the preA group scored the lowest on both versions of the test.

![Graph showing multiple-choice scores for test versions A and B, according to the group of participants (native, preA, preB).](image)

**Test Version**

Figure 8. Experiments 2 and 4 comparison: Multiple-choice scores for test versions A and B, according to the group of the participants (native, preA, preB).

**Recall section.** The recall scores also showed a main effect of group, $F(2, 117) = 4.82, p = .010$, which had the same pattern as the multiple-choice scores: the
preA group scored the lowest ($M = .67$), the preB group scored the highest ($M = .72$),
and the native group scored a little lower than the preB group ($M = .71$).
Furthermore, an interaction between group and test version $F(2, 117) = 5.06, p = .008$, also presented a pattern of results similar to those for the multiple-choice scores
(see Figure 9). The native group scored the highest on Test B, the preB group scored
the highest on Test A, and the preA group scored the lowest on both test versions.

![](image)

Figure 9. Experiments 2 and 4 comparison: Recall scores for test versions A and B,
according to the group of the participants (native, preA, preB).

**Transcription section.** A main effect of group was seen for the transcription
scores, $F(2, 117) = 269.58, p < .001$, showing that the preA group scored the lowest
($M = .681$), the preB group had higher scores ($M = .831$), and the native group scored
the highest ($M = .965$). There was no significant interaction.
Native English Compared to Experiment 3

The pretest and retention test scores from Experiment 3 were compared separately to the native English experiment scores. The pretest scores give a picture of how well Experiment 3 participants performed on these tests without training; participants in the native English experiment received no training, but they had the advantage of listening to unaccented versions of the samples. Therefore, the comparison of pretest scores to native scores gives a sense of how much accent affected the comprehensibility of the speech and participants’ ability to succeed on the tests. The retention test scores, on the other hand, represent the performance of Experiment 3 participants a week after finishing training. By comparing these retention test scores to the native English scores, the benefit of training relative to the benefit of hearing unaccented speech can be seen.

A mixed 2 x 2 x 3 repeated measures analysis of variance was performed for the multiple-choice, recall, and transcription sections of the test. The within-subjects factor was accent type (Turkish, Thai), determined by the accent of the original speakers of the speech samples used in the tests. One between-subjects factor, group (native, accented), designated the tests according to whether the participants had listened to the native English speech (in Experiment 4), or whether they had listened to accented speech (in Experiment 3). The other between-subjects factor was test version (A/F, B/D, C/E), indicating which set of Turkish- and Thai-accented tests Experiment 3 participants had taken as either a pretest or a retention test. Only the main effects and interactions that include group are presented here, as it is the factor of interest in these comparisons.
Multiple-choice section. For the pretest, the multiple-choice scores showed a main effect of group, $F(1, 138) = 7.76, p = .006$, as well as an interaction between accent type and group, $F(1, 138) = 5.45, p = .021$. Scores for the native group ($M = .64$) were higher overall than scores for the accented group ($M = .57$). The interaction between accent type and group showed that although native group scored higher than the accented group on both tests, the two groups differed greatly on the Thai tests but only modestly on the Turkish tests (see Figure 9).

For the retention test multiple-choice scores, there was a marginally significant main effect of group, $F(1, 138) = 3.70, p = .057$, suggesting that scores were again higher for the native group ($M = .64$) than for the accented group ($M = .59$). As for the pretests, there was significant interaction between accent type and group, $F(1, 138) = 12.48, p < .001$. Unlike the for pretests, however, the native group scored higher than the accented group only on the Thai tests; on the Turkish tests, the accented group scored higher (see Figure 10). This reversal for the Turkish tests is due to the fact that the accented group’s Turkish test scores increased sufficiently from the pretest to the retention test to surpass the native group’s scores for these tests. The mean of the Thai-accented test scores remained the same as it had been on the pretest.
Recall section. The pretest recall scores had a marginal interaction between accent type and group, $F(1, 138) = 3.76, p = .054$, and a significant interaction between accent type, group, and test, $F(2, 138) = 3.49, p = .033$. Mean scores for the accented and native groups were about the same for the Turkish tests ($M = .72$ for both), but the accented group scored lower on the Thai tests ($M = .70$) than the native group ($M = .74$). The interaction between accent type, group, and test reveals that although the accent group generally scored lower on the recall tests than the native group, they scored considerably higher on Test A than the native group (see Table 14). No significant effects involving group were found for the comparison of Experiment 3 retention tests to the native English tests, although a marginal effect of
group, $F(1, 138) = 2.79, p = .097$, may suggest that the native group’s scores ($M = .73$) were higher than the accented group’s scores ($M = .71$) overall.

Table 14. Recall Test Means for Experiment 4 (Native) and Experiment 3 Pretest (Accented Pretest) for Each Test Version.

<table>
<thead>
<tr>
<th>Group</th>
<th>Turkish recall test means</th>
<th>Thai recall test means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A$</td>
<td>$B$</td>
</tr>
<tr>
<td>native</td>
<td>.69</td>
<td>.73</td>
</tr>
<tr>
<td>accented pretest</td>
<td>.74</td>
<td>.70</td>
</tr>
</tbody>
</table>

*Transcription section.* The pretest transcription scores showed a main effect of group, $F(1, 138) = 310.67, p < .001$, such that the native group ($M = .97$) scored much better than the accented group ($M = .74$). An interaction between group and test version, $F(2, 138) = 13.15, p < .001$, indicated that the accented group scored highest on the C/E test set ($M = .83$), lower on the B/D test set ($M = .72$), and lowest on the A/F test set ($M = .68$). In contrast, the native group scored the highest on the B/D test set, ($M = .98$), and slightly lower on the A/F and C/E test sets ($M = .97$).

Another interaction between group, accent type, and test version, $F(2, 138) = 3.19, p = .044$, emphasizes that the pattern of scores for the native group was different from that of the accented group (see Figure 10). For the native group, test scores in descending order were: Test D, E, F, B, A, and C. In contrast, the accented group’s scores in descending order were: Test C, E, D, B, F, and A. The scores for the top three tests of the native group, D, E, and F, were nearly equivalent, and all rounded to $M = .98$. The test with the lowest mean for the native group, Test C, had the highest mean for the accented group, and one of the high-scoring tests of the native group,
Test F, had one of the lowest means for the accented group. Thus, tests that were easy for one group were not necessarily easy for the other group.

For the retention test transcription scores, there was again a main effect of group, $F(1, 138) = 452.62, p < .001$, with the native group still scoring higher ($M = .97$) than the accented group ($M = .79$) overall. An interaction between group and accent, $F(1, 138) = 11.76, p < .001$, showed that the accented group scored higher on the Turkish tests ($M = .80$) than on the Thai tests ($M = .78$); in contrast, the native group scored higher on the Thai tests ($M = .98$) than on the Turkish tests ($M = .96$).

As in the comparison of the native English scores to the pretest scores, there was an interaction between group and test version, $F(2, 138) = 30.99, p < .001$, with the same ordering of the tests, although the scores for the accented group had generally improved from the pretest. As with the pretest, the accented group scored highest on the C/E test set ($M = .86$), lower on the B/D test set ($M = .82$), and lowest on the A/F test set ($M = .70$). The native group’s scores, as given above, were highest on the B/D test set ($M = .98$) and lower on the A/F and C/E test sets ($M = .97$ for both).

The final interaction between group, accent type, and test version, $F(2, 138) = 3.04, p = .051$, was marginally significant (see Figure 11). The native group’s scores in descending order are given above; the accented group’s retention test transcription scores had a similar order to the scores for the pretest, though not exactly the same.

In descending order, the accented groups scores were: Test C, Test B, Test E, Test D, Test F, Test A. As for the pretest, Test C had the highest score, and Tests F and A had the lowest scores.
Figure 11. Transcription scores for all test versions according to the group of the participants in Experiment 4 (native), or in Experiment 3 for the pretest (accented pre) and the retention test (accented retention).

Discussion

Comprehensibility of Speech Samples

Participants who listened to native English renditions of the speech samples in Experiment 4 had significant differences between scores on the six test versions for the multiple-choice and transcription sections. These results suggest that at least some of the differences in comprehensibility for the speech samples stem from the content or grammar of the samples rather than the accentedness of the speech. Furthermore, the lack of significant differences between the recall scores may imply that either the recall task was not sensitive enough to reflect the differences in
comprehension that the multiple-choice tests demonstrated, or that the LSA-generated scores for the recall paragraphs did not adequately reflect the quality of the recall paragraphs, and hence that the LSA scores did not adequately reflect participants' comprehension.

**Insights Based on the Comparison of Experiments 2 and 4**

The analysis comparing the native English scores and the scores from Experiment 2 indicated that the native speaker group scored the highest on the multiple-choice and recall measures for Test B, but not for Test A; the highest scores for Test A came instead from the group of participants who listened to the accented version of the speech samples and took the test as a posttest after training. Although familiarity with the test format might contribute to the higher posttest scores of participants in Experiment 2, familiarity alone cannot explain why the Experiment 2 posttest scores were higher than the Experiment 4 scores for Test A. Experiment 2 participants took only one test before taking the posttest. In contrast, Experiment 4 participants took an average of three tests before taking Test A because the order of the six test versions was counterbalanced across participants. It seems unlikely, therefore, that the ability of trained participants to perform better on the multiple-choice and recall measures than untrained participants listening to unaccented speech stems from familiarity with the test format. Instead, this difference in performance seems to come from a combination of the difficulty of the speech sample and the effect of training.
The long speech sample in Test A, which was used for the multiple-choice and recall sections, appears to have been the most difficult to understand of all the test versions, or at least the multiple-choice section was the most difficult, because the multiple-choice pretest scores were the lowest for Test A in the two training experiments. The content itself, not just the accented speech, seems to have contributed to the difficulty, as even participants listening to the native English reading scored lower on the multiple-choice section of Test A than on any other test version. Yet, when participants went through training in Experiment 2, they were able to do quite well on Test A, surpassing the native English group. Training seems to have aided comprehension of this difficult-to-understand speech sample, as reflected in the multiple-choice scores as well as the recall scores.

This training benefit does not occur with Test B, which was one of the easier tests, and the participants listening to an unaccented version of the test speech sample retained their advantage over those listening to the accented version after training as well as before training. Training, then, appears to be maximally beneficial if a speech sample is hard to understand, whether because of its content or because of its accentedness.

The benefit of training on the comprehension of a relatively long, contextualized speech sample, allowing trained participants to surpass even those listening to an unaccented reading of the sample, does not hold for the comprehension of short, decontextualized speech samples. The near-perfect transcription scores of the participants listening to native English readings of the transcription sentences were much higher than either of the Experiment 2 groups. Not only do these results
provide evidence that the transcription test measured different listening skills than the multiple-choice and recall tests, they also signify that participants’ ability to transcribe accented speech did not reach the potential maximum performance shown by the native English group’s transcription scores.

Insights Based on the Comparison of Experiments 3 and 4

The comparison of the native English group to the participants in Experiment 3 leads to further insights about the tests and about how participants’ performance was affected by training. As with Experiment 2, the multiple-choice scores from Experiment 3 are better than the scores of the native English group in some instances. Compared with the pretest scores from Experiment 3, the native English group consistently performed better than the Experiment 3 group, which had listened to accented speech. This finding was expected because unaccented speech is presumably easier to understand than accented speech, and higher test scores reflect better comprehension. However, the superior performance of the native English group was not maintained when their scores were compared with the Turkish-accented retention tests, and the Experiment 3 group numerically outperformed the native English group on these tests. This result is similar to the effect seen in Experiment 2; for that experiment, the phenomenon of trained participants performing better than untrained participants was hypothesized to result from the combined difficulty of the speech sample and the effect of training. The same explanation seems appropriate in the case of Experiment 3. The content of the Turkish-accented samples is apparently more difficult to understand than the Thai-
accented samples, given that the native English group scored lower on the Turkish-accented samples than the Thai-accented samples. Nevertheless, after a little over an hour of training on Turkish-accented speech, Experiment 3 participants were able to deal effectively with not just the accentedness of the speech, but also with the more complex content, thereby scoring higher on the multiple-choice test than participants who heard unaccented speech but had no training.

The Thai-accented retention test scores, in contrast, remained as far below the native English group’s scores on the retention test as they had been on the pretest. The lack of change is probably due in part to the fact that Experiment 3 participants did not train on Thai-accented speech. The Thai-accented samples were rated as more heavily accented than the Turkish-accented samples, and the decrement in comprehension due to accent is evident from the disparity between the very low multiple-choice test scores for the accented group and the very high test scores for the native English group.

For the recall section of the tests, the lack of a significant main effect of group in the comparison of the Experiment 3 pretest scores and the Experiment 4 scores, combined with the significant interaction between accent type, group, and test, indicates that the native group did not consistently perform better than the Experiment 3 groups who listened to accented speech. Furthermore, retention test scores had no significant effects, only a marginal main effect of group. These ambivalent results, coupled with the lack of significant differences between recall scores in Experiment 4, point to the need for a more accurate way of scoring the recall paragraphs, or for a more sensitive test measure than free recall.
At any rate, the transcription scores do demonstrate differences between the native English scores and those of Experiment 3. The near-perfect native English scores were significantly higher than the scores of the Experiment 3 groups, just as they had been higher than the scores of the Experiment 2 groups.

**General Implications**

In sum, the comparison of native English scores to the scores of Experiments 2 and 3 for the transcription, multiple-choice, and recall sections of the test points to some useful information. The native scores establish the quality of scores attained when comprehensible, unaccented speech is heard. This native speech essentially eliminated the accentedness of the original speech samples, but it retained other aspects of the samples, including the content and any non-native lexical or grammatical infelicities (see Appendices C and E). The similarities and differences between scores for the accented and unaccented speech also have important implications.

First, although training for a little more than 1 hour improves participants’ ability to transcribe accented speech, it does not allow them to transcribe as well as participants who listen to unaccented renditions of the speech. Perhaps more training would enable participants to transcribe accented speech with greater accuracy, coming closer to the native English group’s nearly exact transcriptions.

Second, training may help participants understand longer speech samples and remember them better than if they were listening to unaccented speech but had no training. If this is the case, training clearly must be accomplishing more than
neutralizing the effect of accent on speech perception: It must be giving participants skills in concentration and memory, for example, which allow them to understand difficult discourse and remember its content. However, more evidence, including a re-scoring of the recall paragraphs using a more accurate scoring method, such as propositional analysis, is necessary to determine if training truly enhances listening skills.

Third, the low Thai-accented multiple-choice test scores in Experiment 3, which did not change from pretest to retention test, raise a question. How much effect would training on Thai-accented speech have on participants’ ability to comprehend this speech, which is more heavily accented than the Turkish-accented speech? Training appears to improve comprehension of the moderately accented Turkish-accented speech samples about complex topics that are difficult to understand. Would training be more or less effective for heavily accented speech with simple topics that are easy to understand? If the training actually improves listeners’ perception of speech, as it seems to, then a more dramatic effect of training would be expected for speech that is more highly accented.

Finally, the transcription section of the test seems to measure something quite different from the recall and multiple-choice sections. Groups that performed well on the transcription section did not necessarily perform as well on the recall or multiple-choice sections. It could be that the discrepancies between the transcription scores and the scores of the multiple-choice and recall sections stem from the different testing material: The transcriptions, after all, involved sentences on different topics than the longer speech sample used for the multiple-choice and recall sections. But,
the fact that transcription scores of participants listening to accented speech did not come close to attaining the performance levels of those listening to unaccented speech, yet the recall and multiple-choice scores did, supports an interpretation that the transcription section was testing different skills, not simply different material, than the recall and multiple-choice sections.
Chapter 7:
Conclusion

Improving Accented Speech Comprehension

The central goal of this research was to design training methods which would enable listeners to comprehend foreign-accented speech better. One premise of the research was that speech perception involves more than just segment perception: To perceive speech, a listener must go beyond merely identifying segments of speech, such as phonemes or syllables, to fitting those segments into a meaningful schema of semantics and syntax. The processes of identifying segments and extracting meaning are mutually informative, with the consequence that speech perception depends on both processes together. As a result, training and testing in this research involved utterances with a fair amount of context rather than decontextualized phonemes or words. The use of extemporaneous speech allowed the examination of listeners’ ability to comprehend naturalistic speech, with all its variability and coarticulation (Handel, 1989; Ladefoged, 1993; Nygaard & Pisoni, 1995). This variability can make segment identification more difficult. However, the semantic and syntactic information inherent to conversational speech enables listeners to fill in gaps in the acoustic signal and comprehend speech despite lack of segmental clarity (Bashford et al., 1992, Bashford & Warren, 1987; Warren, 1970, 1999; Warren & Obusek, 1971; Warren & Sherman, 1974; Warren & Warren, 1970). By basing this research on naturalistic speech, instead of isolated phonemes and words, the findings more closely address real world speech perception, and the benefits for comprehension that
result from training are more likely to parallel actual benefits that might be found outside a carefully controlled experimental setting.

Because the central goal of the research was to improve comprehension of foreign-accented speech, a critical finding is that a short amount of training on accented speech does indeed improve participants' ability to perceive speech. This improvement is attested by the main effects of test for Experiment 2: Posttest scores were higher than pretest scores for the recall and transcription sections of the test, and there was a trend of higher posttest scores for the multiple-choice section. For Experiment 3, a main effect of test for the transcription section shows that participants' scores improved from the pretest to the posttest, and from the posttest to the retention test.

This improvement is critical because of practical concerns. Eventual applications of the training in business or academic settings would be more realistic and convenient if training can be kept compact; multiple sessions of training spread over weeks or months are not likely to be feasible, for instance, for a university to use in helping freshmen understand their instructors who are not native English speakers. However, a university might be able to implement one or two sessions of training during freshman orientation, or as part of a course during the first week of a semester. The improvement of test scores after about an hour of training in the present research suggests that training may be reasonably easy to implement in practical applications.
Theoretical Bases for Training Tasks

The training tasks exploited theoretical differences between bottom-up and top-down processing. Speech perception theories are built on research that principally examines the perception of isolated phonemes and syllables, with the implication that bottom-up processes drive speech perception (Fowler, 1984, 1986, 1994, 1996; Liberman, 1996; Liberman & Mattingly, 1985; Stevens & Blumstein, 1981). These bottom-up processes are based on the recognition of low-level linguistic units such as features and phonemes. The imitation task tested the benefit of attending to low-level linguistic units; for this task, participants repeated the exact words the speaker said and imitated the pronunciation of those words as closely as possible. In contrast, experimental evidence for the role of top-down processes in speech perception prompted the development of the paraphrase task (Bashford et al., 1992, Bashford & Warren, 1987; Warren, 1970; Warren, 1999; Warren & Obusek, 1971; Warren & Sherman, 1974; Warren & Warren, 1970). For this task, participants repeated the content of the speaker’s utterances, but they used their own words to do so, finding synonyms for the speaker’s words and paraphrasing as much as possible.

That the imitation and paraphrase training tasks directed listeners’ attention to either the pronunciation or the content of the speech samples, respectively, is verified by Experiments 1 and 2. From the results of Experiment 1, which tested participants’ focus of attention during the imitation and paraphrase tasks, it is clear that the paraphrase task helped participants to focus on the meaning of the speech samples they heard. The multiple-choice comprehension test scores were better for participants who paraphrased the utterances than for those who simply repeated them.
or for those who imitated them. Although the comprehension test scores were worse for the group imitating the speech than for the other groups, the pronunciation test scores were not better for the imitation group. The poor comprehension performance of the imitation group, combined with anecdotal reports that they felt that they were concentrating on the pronunciation, implies that this group was at least not focusing on the content of the speech, and may have been focusing on pronunciation.

Experiment 2, the first training experiment, supports the claim that listeners focused on pronunciation during the imitation task, in that participants in the extensive imitation condition improved on the accent discrimination section of the posttest, but participants in the other conditions did not. The accent discrimination section tested participants’ ability to distinguish between the accent of the speaker in the first sections of the test, which for Experiment 2 was always a Turkish accent, and a variety of other accents. After imitating Turkish-accented speech during training, participants in the extensive imitation condition were better able to determine which speech samples in the accent discrimination test section were marked by Turkish accents and which had some other accent. Also, in Experiment 2, participants in the limited paraphrase condition improved on the recall section of the posttest, and participants in the other conditions did not. Based on these results from Experiments 1 and 2, the training tasks apparently accomplished their intended aim of directing participants’ attention to pronunciation, in the imitation task, or content, in the paraphrase task.

Despite the evidence from Experiment 2 that listeners attend to pronunciation during the imitation task, a follow-up study on the focus of attention during the
imitation task would be useful. The tests used to examine knowledge of accent characteristics in Experiment 1 were difficult, and they required participants to indicate how the speaker had pronounced 10 words by choosing between four pronunciation options for each word. Thus, participants had to remember the distinct pronunciations of specific words in order to succeed on the pronunciation tests in Experiment 2. This type of detailed, concrete information does not fit well with the definition of implicit knowledge as abstract, inaccessible to consciousness, and difficult to communicate (Berry & Dienes, 1993; Reber 1989). As a result, these tests may have demanded use of explicit knowledge of the accent rather than encouraging participants to exploit implicit knowledge to make judgments. To test implicit knowledge of pronunciation more effectively, a follow-up study could include a variation of the accent discrimination tests that were used in Experiments 2 and 3. This type of test would allow participants to make decisions about accent characteristics without needing to invoke a specific memory of a particular word’s pronunciation. By invoking implicit knowledge rather than explicit knowledge, the follow-up study might lead to more definitive results about the participants’ attention to pronunciation and accent characteristics.

Effectiveness of Imitation Training Task

Assuming that the imitation training task promotes attention to the pronunciation of speech, it thereby encourages practice of bottom-up processes. According to speech perception research and theories, with their emphasis on bottom-up processes (Fowler, 1984, 1986, 1994, 1996; Liberman, 1996; Liberman &
Mattingly, 1985; Stevens & Blumstein, 1981), the imitation training condition should have greatly enabled listeners to perceive speech more easily and accurately. In fact, some benefit was found. This improvement, however, mostly occurs for the transcription test section, which arguably measures something more like segment perception than speech perception. Therefore, these training experiments intimate that practicing bottom-up processing by imitating speech enables segment perception, but the improvements in segment perception do not necessarily correspond to similar improvements in speech perception.

Focusing on the pronunciation of the speech does not predictably improve comprehension of longer conversational speech samples, but it does improve perception of fairly short samples with little context. This claim is supported by the results of Experiment 2; according to an analysis of variance and paired-comparison t-tests, the transcription scores improved for participants who did the imitation task, but not for those who did the paraphrase task. A marginal improvement of the recall scores, based on paired-comparison t-tests, for those in the limited imitation condition may indicate that imitation training can also improve comprehension of longer speech passages, but it does not have this effect consistently, as recall scores did not improve for the extensive imitation condition. Experiment 3 also supports the benefit of imitation for the perception of speech samples having little context. For the imitation condition, retention test transcription scores were significantly higher than pretest scores for Turkish-accented speech, and there was a similar trend for Thai-accented speech. Multiple-choice posttest scores also improved in the imitation condition, but only for Thai-accented speech, suggesting that, as in Experiment 2, imitating can
have some benefit on comprehension of longer speech samples as well as shorter samples.

The improvement on the transcription section for the group focusing on pronunciation, as well as the general lack of comprehension improvement for longer passages of speech, is somewhat surprising in light of the Derwing et al. (2002) training study. In the Derwing et al. study, accent familiarization training on Vietnamese-accented speech took place once a week for eight weeks; testing occurred a week before training began and a week after training was completed. Testing involved listening to English passages read by native Vietnamese speakers, and comprehension measures included a transcription task as well as a multiple-choice comprehension test. The three training conditions were: control, familiarity, accent instruction. Participants in the control condition received no training, simply taking the pre- and posttests. In the familiarity condition, participants listened to Vietnamese-accented speech and discussed cross-cultural issues during training; in the accent instruction condition, participants received the same training as those in the familiarity condition, with the addition of lectures on Vietnamese accent characteristics. Transcription scores improved on the posttest across conditions, including the control condition, with no evidence that accent instruction enabled better transcription scores for the accent instruction group. Therefore, despite explicit instruction about Vietnamese accent characteristics during training, the accent instruction group did not improve more than the other groups on the transcription test. These results contrast with results from the present training experiments, which
indicate that transcription scores benefited from implicit, active learning during training, especially in the imitation condition.

Further, the accent instruction group in the Derwing et al. (2002) study apparently improved more than the other groups on the multiple-choice comprehension test. Although the authors do not claim this improvement in comprehension, a re-analysis of their results, as described in the Introduction, indicates that the listeners who received accent instruction improved more on the posttest than the control or familiarity groups. If explicit instruction about accent characteristics enabled better comprehension, implicit learning should have a similar, if not more dramatic, effect because implicit learning should be a more effective way of training the automatic and implicit process of speech perception (Berry & Dienes, 1993; Reber, 1989; Schneider & Shiffrin, 1977). Yet, the present training experiments do not unequivocally support a comprehension benefit from learning about accent characteristics. Instead, the present results lean the opposite direction of the Derwing et al. study, suggesting that learning about accent characteristics is not a very effective way to improve comprehension of long speech passages.

To clarify the benefits of learning about accent characteristics, whether implicitly or explicitly, more work needs to be done. The results of the present training experiments are not completely unambiguous, perhaps in part due to problems with the comprehension measures, as discussed below. A more lengthy multiple-choice test or a re-scoring of the recall paragraphs might clarify the comprehension results. In addition, contrasting explicit and implicit learning about accent characteristics could shed light on the different benefits the two modes of
learning may have on comprehension of accented speech. Finally, a follow-up of the Derwing et al. (2002) study, using the same methodological paradigm that they used, could be done to verify their results.

**Effectiveness of Paraphrase Training Task**

In contrast to the benefit of imitation training, paraphrase training mainly appears to help participants comprehend relatively long speech samples, although it may have some benefit also for perception of short speech samples. Training on the content of the speech in the paraphrase condition was expected to require listeners to practice top-down listening strategies, and, as a result, develop their ability to fill in missing or garbled low-level information more automatically in order to understand the message with less difficulty. For this training condition also, some of the expected benefit was found.

In Experiment 2, scores improved on the recall posttest for participants in the limited paraphrase condition, according to paired-comparison $t$-tests. Transcription scores did not show improvement for participants who paraphrased speech in Experiment 2; however, in Experiment 3, there was a trend for higher transcription scores on the retention test for the paraphrase condition, but only for Thai-accented speech.

The results of the two training experiments generally indicate that imitating speech enables perception of speech with limited contextual cues, and that paraphrasing speech helps the comprehension of speech with an abundance of semantic and syntactic information. This differential effect of training is not
completely orthogonal. Nevertheless, it appears that the paraphrase task may enable comprehension of contextualized speech better than the imitation task, which facilitates comprehension of decontextualized speech. Because the type of speech most common in daily interactions is contextualized discourse, the paraphrase task may ultimately be the most effective training method for practical applications.

This proposal is rather unexpected in view of Derwing et al.’s (2002) research, as well as in consideration of speech perception theories that are grounded in bottom-up perceptual processes (Fowler, 1984, 1986, 1994, 1996; Liberman, 1996; Liberman & Mattingly, 1985; Stevens & Blumstein, 1981). If speech perception necessarily entails accurate segment perception, then imitation training should reliably improve comprehension, regardless of how comprehension is measured. Yet, the two training experiments reported here imply that gaining familiarity with the shifts in pronunciation of accented speech improves performance on tasks demanding recognition of discrete speech segments, but that it is not as effective in improving comprehension of speech itself. In other words, gaining skills in perceiving discrete segments does not translate unequivocally into gaining the skill of perceiving speech. Top-down processes are critical for speech perception, and practicing these processes during training by paraphrasing speech seems to effectively enable accented speech comprehension. However, more information about the testing methods, discussed below, needs to be gathered before the imitation task can be rejected as being an effective way of improving speech comprehension.
Measuring comprehension has its challenges, as addressed in the discussion section of Experiment 4. The three comprehension measures used in these experiments were: transcription, multiple-choice, and written recall. The transcribing task assesses fairly discrete, isolated perception of words and phonemes. It is a sensitive measure, but not one that necessarily reflects real-world comprehension, the kind that is needed to hold a conversation or listen to a lecture. As suggested in the Discussion section of Experiment 4, the recall measure may not be sensitive enough to capture improvement in speech perception abilities. However, before that judgment is made, the recall paragraphs need to be re-scored using propositional analysis, which may prove a more accurate way of scoring. Better accuracy in scoring would allow the recall measure to be more sensitive. Developing multiple-choice tests can also be problematic, and their difficulty is likely to vary unless the tests are independently verified. So, the comprehension measures that get at real-world comprehension are problematic and messy, but it is worth including them because transcription is not a fully adequate way to examine comprehension.

Training as Implicit Learning

The training conditions entailed implicit, active learning. Explicit knowledge is accessible to consciousness and able to be communicated, but implicit knowledge tends to be not accessible to consciousness and not easily communicated. Because speech perception is an automatic process, it is largely beyond conscious control, so an intervention based on explicit instruction is unlikely to have much effect on a
participant’s ability to perceive speech (Berry & Dienes, 1993). Furthermore, listeners are usually quite aware of an accent, but they are generally unable to articulate the characteristics of an accent in precise linguistic or acoustic terms. The automatic and implicit nature of speech perception suggested a training approach which relied on implicit learning and required the listeners’ active participation (Bower & Hilgard, 1981; Reber, 1989; Schneider & Shiffrin, 1977).

During training, therefore, participants were not given explicit information about the accent characteristics of the speech they were listening to, and neither were they instructed about the technical aspects of perceiving speech. The two training tasks instead required that participants reproduce the speech they heard, concentrating either on the content or on the pronunciation of the speech. These tasks were intended to help listeners comprehend accented speech by tapping into implicit learning processes in which the complex system of accented speech could be acquired without necessarily having to be verbalized. Also important was the active nature of this process, allowing listeners to practice the skill of perceiving accented speech by generating responses. Thus, the tasks used in the training experiments presented here promoted active, implicit learning. Implicit learning has been used successfully to foster the acquisition of implicit knowledge in artificial grammar tasks (e.g., Dienes, 1993; Reber, 1989), and active learning has shown a benefit in knowledge and skill acquisition (Berry & Dienes, 1993; Bower & Hilgard, 1981; McNamara & Healy, 1995). The success of the active, implicit approach to training in the present experiments is attested by the higher test scores after training than before for at least some of the test sections in both training experiments.
Further evidence for the implicit nature of the knowledge gained during training comes from Experiment 3. As already mentioned, imitating speech improved participants’ ability to distinguish between accent types in Experiment 2: Posttest scores for the accent discrimination section of the test improved for participants in the extensive imitation condition. Paraphrasing did not improve recognition of the characteristics that contribute to a certain accent in Experiment 2. In contrast, Experiment 3 did indicate that paraphrasing improved accent discrimination test scores, though only on the posttest and not on the retention test. Imitating speech also improved accent discrimination scores, but only on the retention test and not on the posttest. In the discussion of Experiment 3, these results were explored in terms of explicit and implicit knowledge. The lack of improvement on the discrimination posttest for the imitation condition may indicate that explicit attention to accent characteristics was detrimental to implicitly recognizing accent characteristics in the short term yet helpful in the long term. The opposite seemed true for the paraphrase condition: Implicit learning about accent characteristics while participants focused on meaning in order to paraphrase helped accent discrimination in the short term but not in the long term.

**Durability, Transferability, and Other Insights about Training**

Based on the results of Experiment 3, which tested the durability and transferability of training, training seems to be durable and possibly transferable. Scores for the transcription test section improved not only on the posttest, immediately after training, but also on the retention test, which occurred a week after
training. There is also some evidence that the training is transferable, with trends for improvement on Thai-accented transcription tests occurring in both training conditions. Participants received training only on Turkish-accented speech, so improvement on Thai-accented tests may indicate that the skills they gained during training transferred to a different type of accent. The marginal interaction between test and accent for the multiple-choice scores imply improvement on the posttest for the Thai-accented speech, and improvement on the retention test for Turkish-accented speech. The posttest improvement for the Thai-accented speech is corroborated by a paired-comparison t-test indicating higher posttest scores in the imitation condition. Furthermore, the lack of a significant main effect of accent type provides no evidence that the Thai-accented tests were not improving when the Turkish-accented ones were.

However, the results on durability and transferability need to be explored and confirmed. To examine transferability, future training experiments could involve training on only Thai-accented speech, as well as on only Turkish-accented speech as these experiments did, or training on more than one accent type. Training on multiple accent types would be expected to promote transfer to new types of accents not included in training. Another index of transfer would be a comparison of listeners trained on the same speaker to those trained on different speakers with the same accent. Speaker normalization, becoming familiar with a particular speaker’s voice, has been shown to facilitate comprehension (Nygaard & Pisoni, 1995), so training and testing on the same speaker would probably yield greater improvement in comprehension than training on multiple speakers, but the training might be less
transferable to understanding other speakers with the same accent or to comprehending speakers with different accents. To determine the extent of durability, the delay between training and retention test could be varied, and longer delays compared to shorter ones.

In general, training may have the unexpected result of making people better listeners, not just better perceivers of accented speech. The comparison of scores for the training experiments to the native English experiment suggests this improvement in listening skills. If trained participants can actually beat untrained participants on comprehension measures, regardless of the fact that the untrained participants have the advantage of listening to comprehensible, unaccented input whereas the trained participants hear accented speech, then training may have improved general listening skills, such as paying attention and remembering what the speaker said, and not just improved ability to understand accented speech. In other words, this training is not just about eliminating the negative effect of accent on perception; it also may help people listen better on the whole.

Nevertheless, the training does not seem to have reached its maximum potential for improving accented speech perception, as the transcription scores for those who listened to native speech were considerably higher than the scores for those who listened to accented speech, even after training. More work on how much training would be maximally beneficial is necessary to see how long it would take for people to understand accented speech just as well and as easily as they do unaccented speech, or if this desired ultimate attainment could ever be reached.
REFERENCES


APPENDIX A:

INTERVIEW QUESTIONS

During your interview, you will speak about two different topics. One topic will be related to your native culture; the other topic will address academic or work-related issues. Below, you are given questions about culture and school/work. Please choose one question to answer from each topic. You will have a few minutes to think about or make notes on the two questions you choose, and then the interviewer will ask you those questions and you will discuss them aloud.

CULTURE:
1. Describe in detail how people celebrate one of your native country's holidays.
2. How do you prepare one of your favorite foods?
3. Explain how to play one sport that is popular in your native country.
4. What is a good book you have read recently? What was it about and why did you like it?
5. Choose a special celebration, such as a wedding, birthday party, graduation, or funeral. Describe what this special occasion is like in your country and how people celebrate it.

SCHOOL or WORK:
6. Why and how did you become interested in what you are studying now, or what you studied in college?
7. Have you taught or taken a class recently? Talk about the topics you covered.
8. If you are doing research, please summarize the research you are doing.
9. If you have a job, explain in detail one of the projects you are working on, or one of the projects you have worked on in the past.
10. What do you plan to do with what you have learned in school? Explain specific jobs or projects or research you would like to do and how they relate to something you have learned.
# Appendix B: Rater Assessment Form

## Pronunciation Assessment

Instructions: Please answer the following questions as thoroughly as possible.

1. **Consonants** (use IPA)
   List the consonants or consonant clusters that are problematic (incorrectly produced or omitted) for this speaker.
   - Target consonant
   - Speaker's consonant

2. **Vowels** (use IPA)
   List the vowels that are problematic for this speaker.
   - Target vowel
   - Speaker's vowel

For the following areas, please rate the speaker's pronunciation on a scale of 1-7 by circling the appropriate number.

3. **Vowel reduction**: assess reduction of vowels in unstressed syllables and function words
   - 1 2 3 4 5 6 7
   - Constant
   - Native-like

4. **Pausing**: assess appropriateness of pausing between words and phrases
   - 1 2 3 4 5 6 7
   - Constant
   - Native-like

5. **Linking**: assess appropriateness and smoothness of linking between words
   - 1 2 3 4 5 6 7
   - Choppy / Slurred
   - Native-like (Smooth)

6. **Stress**: assess word-level stress placement
   - 1 2 3 4 5 6 7
   - Incorrect
   - Correct
   - Stress (Heavy accent)
   - Stress (Native-like)

7. **Rhythm**: assess sentence-level stress and placement
   - 1 2 3 4 5 6 7
   - Incorrect stress
   - Correct stress
   - Native-like (Heavy accent)

8. **Intonation**: assess appropriateness of word- or phrase-level pitch patterns (rises and falls in pitch)
   - 1 2 3 4 5 6 7
   - Incorrect
   - Correct
   - Native-like

9. **Overall pronunciation**: How heavily accented is this person's speech?
   - 1 2 3 4 5 6 7
   - Heavy
   - No foreign accent
   - Foreign accent

10. **Overall intelligibility**: To what extent can you comprehend (understand) this person's speech?
    - 1 2 3 4 5 6 7
    - Some
    - Most
    - All
APPENDIX C:

COMPREHENSION TEST SAMPLES

Russian
349 words, 2 min 25 s

(Sigh) When I just came here eight years ago, was very surprise. One of uh parents at McKenna school, where I started to work, um, they knew I am I am a professional pianist, and they invited me to play for children. And I was glad to do that. And em it was just between thirty forty minutes maybe music for party, and I was very surprise how they enjoy music. And I started uh to think about I can do that, if people really will enjoy these pieces. And uh I started to play more and more around, for my friends’ weddings (laugh). And I explained them I am playing very different music, but they liked it, more classical music and Russian pieces. And when I started to show my last project, with the Israeli Cycle, was very surprise. Lot of people cried, and uh they thought they uh heard this music on the heart, and uh they really want to have the CD. And I have a a lot of people who wants to have it, but its not rel-ready yet, but I hope I will finish it. And um, I started to work on this project about ten years ago. And I um I didn’t think it will be very serious. Uh, I just worked couple hours a month for a culture center, it was Israel and America’s culture center in Tashkind. And just asked me uh to help a little bit and play for a couple embassies. And I thought, “Okay,” and uh, I had a choir, about thirty five, forty children. They really liked it. And I started to harmonized very very old Israeli uh songs. Actually, it’s two very important things for Israeli people. It’s like um all this lives they’ve had a lot of prayers, and um, and uh melodies, and when they put it together, it started to be very beautiful songs. And I just think I I will finish this pro-just project finally, and I will put it on a CD. And I hope people will enjoy it.

Turkish, Test A
280 words, 2 min 34 s

I’m doing research on speech recognition and spoken dialogue systems. So, our aim is to improve the system performance and make these systems practical and available, widely available for people usage. Uh, you can either use a cell phone or a regular telephone. And you’ll dial up a system and you’ll just say, “Okay.” The system will say, “How can I help you?” And you will talk the system as you are talking to a human. Actually we are trying to beat human beings. Like, we must develop a system better than humans. But it’s far, we are very far from that. The obstacles are, yeah, we don’t know how f—actually we don’t know how we recognize the speech. We have clues, but it’s still the brain, it—that puzzle has not been solved. Let me give you an example. Take a one second speech. When we input the, that, into our system, in some cases it takes to understand that ten seconds. One second speech needs ten seconds processing. So, ten times slower than human beings. And, if you look at the technology behind that, it’s huge. We have huge models, we have very
goes together. It’s not like a traditional project, the design is completed before the construction starts. So, these two go together. And, uh, the design builder has to conform with the requirements of RTD and CDOT. And the only way you can control this project is auditing, is done by inspection like the like it, you know in a you do in a conventional uh project. So, auditing is basically just uh take samples from the work. You cannot audit the whole work, you just you know go uh for example when they are pouring the concrete. You cannot uh inspect the whole concrete, but you can uh the auditors uh basically be present there for the specific period of work. And basically just uh observe the construction activities uh just to make sure that the the design-builder, the contractor, is uh constructing the project according to these requirements. So, uh I’m an auditor in this project, uh auditing the the light rail work. But right now I spend most of my time reading the the specifications for the seven? maintenance facility of RTD because the project is going to start the like the twenty sixth of August. So, um, eh I guess I’ll be there full time. So, I’m just being prepared for reading the specifications to get familiar with the requirements.

Thai, Test D
303 words, 2 min 38 s

Uh, actually, I just went to hike, uh I just went to Rocky Mountain National Park last week. And, imagine that you are in the same place as your friend. You have the same view, same mountain, same waterfall, same uh tree. But, you are, the point of view of you and your friends may be different. So, that’s what make me like taking pictures because it’s the way that I want people to see the picture, from my eyes, through my eyes. And using the camera, sometime we can make the picture look better than what it is, right, because we can see in the books, in the pictures books that, uh for example, this same waterfall taking from somebody might look uh very soft to you. The water may be like uh, stream of water may be look soft and long, but taking the picture from somebody might look just a simple picture of when you use the simple camera to take a picture. So, and also sometime taking picture from the camera, maybe uh those that may help you improve the what what you see. Uh, for example, if you open the, oh if if you use the longer speed of your shutter, you may see the picture in the different way. For example, in the at night when you use shuttered speed at one hour, so the the result of your picture you you are going see the movement of the sky, of the s-star in the s-sky. It’s not only dark, it’s the light that move around. And, most of the time it’s going to be in round shape because the earth circle in the way of the orbit move is circle or oval maybe. Yeah. So, using that you are going to see the picture that you cannot see from your eyes.
Thai, Test E
340 words, 2 min 28 s

I’m working in the computer science and in networking-related area. And the specific organization that I will work with, which is the national, statistical national office. And currently they are just only like it’s small office, so if I came home I would be the only one doctor in that field. So I think I can do a lot of thing in that particular organization. Like, uh, I’m currently working on a project which is about a security attack, that they call intrusion tolerance. For, uh, like, when someone enter in, attack your computer or your network do not, you do not want to shut down your whole computer just to solve that problem because the other organ—I mean the other member of your organization may need to use that network also. So what I can do what my research is to exploring that. How can you just suspend that, uh, susp- uh I mean the suspicious host of computer uh for a while. We do not get rid of them at once, but we just suspend them in the suspended period of time. And then after we sure that that member, that host computer is um attack is truly attack and is harmful for the whole network then we get rid of them from the whole network. During the in the suspended period we can still work on our normal job, like regular job, without harming the whole system. I’m collaborate with my professor, and I’m we are trying to find more collaboration from the I mean from the other university. And, next week I going in to a conference in Washington D.C. I think can find cooperation there. I think especially this year I travel I mean almost every month since January. Like, on spring break I went to Utah. And again I love Utah a lot so I have in May I went to Utah again. I think if someone invite me to Utah again, I will go there.

Thai, Test F
319 words, 2 min 51 s

Um, right now I working on the reading project. Um, this project is about uh is about um create the book that have include the adjacent animation that will help kids reading. Um, I think it’s fun it’s very fun because last two week I went to school and I have to install program and see how kid play the game and, you know, and they will have, it’s like, wonderful idea. Something like um in the game when you play, I will have the animation that will is like the progress. That suppose like when you play the game you will have the the bear, uh bear. Um, it’s femal- female bear and male bear. If you come closer, every time when you answer the question, and then when they come closer they will dance together, you know. And kids kids um suggest me that it’s it’s good, but they like um when they come closer and the female bear should be can pick can pick the male bear that they like, you know. And I think, “Oh, it’s it’s cool idea!” Right now I’m trying to um improve the reading book um because right now it’s more it’s very complicated, you know. It’s difficult for kids to to play around because it it’s have many button on there. I think I I maybe in the future I want to make a change something, like the interface that more easier and not look more uh look complicated. Because sometime I I know that when kid play it it need to have
teacher to suggest that how to play that. And, there are some bugs that we need to figure out because the book is very complicated because they have the many component inside. Its have the animation. We suppose that there no teacher in the room and kid can read the book with the interactive character.
Okay, in my country there are uh two kinds of holidays
One is uh religious holidays,
and the second one is kind of patriotic holidays.
I don’t know what to called.
Uh, we have to patriotic holidays, they are festivals.
Uh, one of them is for the independence, like independence day.
And the other one is for Turkish children.
We had a leader called Attaturk?, and
he set this festival for children
to celebrate for their funs, their joys.
Then, uh, for the religious holidays we have two,
and one of them, it’s called Sugar Festival,
and the other one is Sacrifice Festival.
In the Sacrifice Festival,
we sacrifice a sheep or cow to eat its meat.
I know it’s not, uh,
doesn’t sound good for Indians or
for other religions but we do this.
And, in Sugar Festival,
we, it’s not, we don’t, we don’t sacrifice anything,
we just serve sugar, this is a tradition.
Then, in both of these,
uh, in compared to the patriotic festivals,
the difference is we visit our relatives, our friends,
in f- three or four days.
But, in patriotic festivals, we celebrate the festival or holiday
uh, as a public celebration. This is the difference.
And, it’s becoming, the religious festivals are becoming different
as the years pass because
uh, the number of friends or relatives that you visit is decreasing, and
if my grandfa- grandfather or grandmother was living,
uh, they wouldn’t like this at all.
And, it’s very common for religious festivals
to buy a new clothes for the children,
and you put your clothes, new clothes, under your pillow
before the night of the, before the, one night before the festival,
and, because you want to wear them as soon as possible.
So, you keep them under your pillow
uh for them not to be stolen by anybody.
I want to talk about my study.
Uh, I'm interested in drawing the places.
Yes, I'm interested in planning the building.
I'm a student of the faculty of architecture.
Uh, I love my studies a lot
because I like to be a new things.
Uh, I like to draw something all time.
I like to see new places,
so I'm here.
Uh, I wonder how is your buildings, house.
I wonder this so, I came here to my sisters, yes, and uh.
In architecture, first you learn how to measure the things,
and then how to measure the place,
and you can start drawing, planning
some houses or buildings about museum.
Uh, here, all houses have a
maybe little, but maybe big, they have a garden.
We don't have any garden.
Yes, and uh, you have a very big road, not narrow,
but we have narrow road, and uh
like this. But your houses is very small.
Room is very small
than our houses. We have a very big and
um very comfortable house and room.
And you have very beautiful and wonderful hotels.
And, I'm studying here to
uh, all day, I went to do,
not all day, three days of a week,
uh, I go to do Denver Art Museum.
To my entir- internship, yes.
Uh, and, I write them about their um, materials.
Um, I’m working on polymer, in the area of polymer uh chemistry. Uh, polymers are eh the general or polymer is the name given uh to the class of substances such as plastics, nylons, and other uh similar uh uh substances. Uh, uh I was working on uh polymers which are used in uh as biomaterials. So, the one project that I uh I worked was uh uh on uh preparing polymers uh new polymers which can be used uh in- in the body uh uh in the place of cartilage uh so in order to do that we have to synthesize, uh prepare, new uh chemicals that can be used uh to prepare cartilage-type substances. Uh, or uh the same substances can be used to uh in the tissue engineering field to prepare uh or to grow cells of different tissues and use those polymers as support materials for uh or um as the media to grow these cells. So, the main the main su- uh topic that I worked on is the polymerizations uh or the formation of polymers uh under the, using light as a as a source for the uh for the synthesis of them. And, uh, so, one, one other project related to this it was the uh investigation of the eh structure of polymer that will be formed by the, uh by using again light to understand how they form and how the formation conditions are affecting their physical and mechanical properties.
After I graduated I apply to a job and found a job in Denver. So now I'm working as a programmer in Denver. Uh, we I'm doing more uh like software for the internet, business to business software for companies, uh. The one of the things like I do is like writing locator for like uh clients. So, like, for instance this tool is uh you know you go to like IBM's website and you want to buy an IBM computer. You, you know, put your address or some search criteria, and it, you know, it brings you the IBM resalers on your, in your neighborhood. So, like one of the my favorite project I finished a couple of months ago is this profiler we did for some big client in in the US. It was a multilingual uh application. So it was, with Veno. This company has like offices all over the world and like and different, and [mumble] everybody speaks different languages. So we wrote this application so that you know they put their profiles and all their information with this tool in the language they want. So, they could put Chinese, and and, you know German or Polish or whatever language. We supported like twelve different languages. And, they could like switch languages in the application. So, they put like Sw- Polish and then they can put in Russian too. And, for the customers, that way you go to like their website again, and you do the search based on the country, city or some key words. And they bring you all these resalers that they profile you know themselves in different languages. So, you do the search in different languages; so it says like you have these companies they profile themselves in English, but you can like click on this [mumble] and the whole site changes to Chinese. And, so, all this thing is dynamic, and it was a very you know uni- uh unique experience.
So when two people decide to get married, um, of course, the man asks the woman for most of the time uh to get married. And then, um they tell about their decision to their parents. And then the first thing to do is the man’s parents uh visit the g- uh girl’s uh parents at the girl’s parents’ house, and they ask uh for the permission of girl’s father for the marriage. They see each other and you know judge if it is a good uh match or not. Um, man’s uh father ask uh girl’s father that our son wants to get married with your daughter. Do you accept it or not? They say, “Okay,” or not, I don’t know.

So, once they decide about the wedding date, uh, they start to uh get uh ready and make the plans and decide where to make it, where- and when to make it, who will be invited, and what will be served as food and things like that. And, when the wedding date comes, uh most of the time, uh, the bride goes to a uh hair salon and has um professional hair-do, which continues like hours. And then also professional make up, and she gets uh dressed and she gets ready uh for the wedding. And then, one traditional thing is that the bride goes to uh her parents’ house when she’s ready, and then uh the uh groom ’s uh parents, family, relatives, friends, they come to uh bride ’s uh parents’ house just in huge crowd like in maybe ten cars and you know, just to take her and uh take her to the r- uh wedding place. And then they go to the wedding place. It is, actually, dependent on people, how you want to do that. You can do the wedding at one place, you can do the reception and dinner at another place,
or you can do both of them at the same place. This is actually what we did. We did both of them at the same place. So, we went to the wedding place, and we had the official wedding there, and then we had our dance and then uh, we had our dinner, which, you know, everybody definitely liked. And they celebrated us, and then, and then, you are done.
APPENDIX E:

TRANSCRIPTION TEST SENTENCES

Test A. (Turkish; 67 words)
1. But the same is valid for the other team; as time passes, they get tired.
2. I like to make defense, don’t let people to come to our goalie.
3. It’s an environment for people to discharge, rather than discharging on each other [laugh].
4. They can’t come back very quickly, and y-you might be in trouble.
5. But the best team is one which keeps their structure throughout the whole game.

Test B. (Turkish; 67 words)
1. They don’t taste so good, but [laugh] I don’t think lentils should be sweet.
2. Take for example the cover of the orange and then put it inside.
3. This is the most boring part because it takes it so long to [laugh] start boiling.
4. And you can use that powder to cook anything you like, like flour.
5. You might also add uh some tomato pa- paste, and that might be one tablespoon.

Test C. (Turkish; 67 words)
1. Somebody who is in authority give the speech about the importance of that day.
2. Everybody wakes up early in the morning and then (uh) we go to (to) the mosque.
3. When I was at the elementary school, that day was a holiday.
4. At night, when the weather is clear, there is a firework show.
5. We visit the relatives, and the tradition is the younger ones visit the older ones.

Test D. (Thai; 67 words)
1. People come from all over the world to have a water fight.
2. We have the theatre to show the traditional style of dancing to the foreigner.
3. The people who are elder than them will give them the good word.
4. They just catch the bus, catch the train just to go back to see their parents.
5. When it’s hot, people just want to cool each other down, right?

Test E. (Thai; 66 words)
1. After that, you just mix its together and put it in a plate.
2. But the other one that you usually saw in the TV show is spicy.
3. We have a lot of mountain, but it’s in the tropical area.
4. You choose a papaya that is not ripe yet, and then you chop it.
5. We use another kind of utensil, which I do not have it here.

Test F. (Thai; 65 words)
1. If you use another kind of sugar, that will make it soft sweet.
2. But at the first time, it’s not successful, and my friend cannot eat it [laugh].
3. When he want me to update the page, I have no time to do that.
4. It’s hard to explain because it have a lot of component.
5. I don’t like American food because its has a lot of calorie.
APPENDIX F:

SAMPLE TEST

Test B

I. Sentence Transcription Sample B

*Transcription*: In the box below, please transcribe the 5 sentences that the speaker says. Press the return key after each sentence so that new sentences begin on a separate line.

II. Comprehension Sample B

*Recall*: In the box below, please write what the speaker said, as much as you can remember, as if you were the speaker repeating the speech to another person.

III. Comprehension Sample B

*Test*: According to the speech sample you heard, please check the box next to the correct answer by clicking on it with the mouse.

1. Admission to college in the speaker's native country is based on
   □ a) high school grades.
   □ b) one's score on a test taken after high school.
   □ c) a combination of grades and one's post high school test score.
   □ d) an application essay.

2. Students apply to universities by
   □ a) paying a fee and submitting their applications.
   □ b) sending their applications to the universities they are interested in.
   □ c) applying to specific departments.
   □ d) providing a central system with their academic preferences.
3. In their applications, students must specify two things:
   □ a) the school and the department they would like to attend.
   □ b) their high school grades and their entrance exam score.
   □ c) their academic strengths and weaknesses.
   □ d) their entrance exam score and the school they would like to attend.

4. When indicating their preferences, students are allowed
   □ a) 8 choices.
   □ b) 18 choices.
   □ c) 80 choices.
   □ d) 180 choices.

5. The number of students admitted to a particular department is
   □ a) determined by each individual department.
   □ b) based on the number of qualified applicants.
   □ c) decided by university-wide policies.
   □ d) a fixed number under 200 for all departments.

6. To fill the available slots in a department,
   □ a) entrance exam scores are sorted and the students with the highest scores choose their departments first.
   □ b) departments select students with the highest grades in high school.
   □ c) departments accept the top-scoring students from those who have indicated a preference for that department.
   □ d) the entrance exam scores, and not the students' preferences, are the most important factor.

7. How many students did the speaker's department accept when she was there?
   □ a) The number varied from year to year.
   □ b) 108.
   □ c) 118.
   □ d) 180.

8. Students may not be admitted to the department that is their first choice if
   □ a) their entrance exam score is not high enough.
   □ b) their high school grades are not high enough.
   □ c) they choose a major that is very popular.
   □ d) they go to a small university.

9. According to the speaker, what people major in for college is
   □ a) their decision, for the most part.
   □ b) fundamentally not their decision.
   □ c) completely randomly selected.
   □ d) basically determined by bureaucrats.

10. In addition to engineering, the speaker's interests included:
    □ a) chemistry
    □ b) biology
    □ c) dentistry
    □ d) genetics

IV. Discrimination Test B
    Indicate whether the speakers have the same accent or a different accent as the speaker you just heard by clicking with the mouse in the appropriate box next to each number to check either 'Same' or 'Different.'

<table>
<thead>
<tr>
<th></th>
<th>Same</th>
<th>Different</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX G:

TRANSCRIPTION SCORING RUBRIC

Errors were tallied by exact word match:

1. inserted word = 1 error (noted by letter—A, B, C)
2. missing word = 1 error (noted by # of word missing)
3. replaced word = 1 error per original word replaced, depending on # of syllables in original ('make the fence' for make defense would be 1 error, not 2)
4. homophones = no error ('flower' for flour, 'then' for than)
5. inverted phrases = 1 error ('which throughout the whole game keeps their structure' for which keeps their structure throughout the whole game)
6. non-contracted forms = no error ('do not' for don't, 'it is' for it's)