THE IMPACT OF PERCEPTUAL DISSIMILARITY ON THE PERCEPTION OF FOREIGN ACCENTED SPEECH

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

in the Graduate School of The Ohio State University

By

Shawn Aaron Weil, M.A.

* * * * *

The Ohio State University

2003

Dissertation Committee:

Professor Mark A. Pitt, Adviser

Professor Keith Johnson

Professor Nadine Sarter

Professor James Todd

Approved by

____

Adviser

Department of Psychology

ABSTRACT

Understanding the effects of Foreign Accented Speech (FAS) is important for both theoretical inquiries on the nature of speech perception and for evaluation of real-world situations in which successful speech communication is critical. Much of the previous research investigating FAS has been limited to measuring gross intelligibility. In all of these studies, FAS was found to be less intelligible than non-accented speech (NAS). The current series of experiments was designed to expand and refine this general finding. The relative intelligibility of different phonetic contrasts was predicted based on descriptions of accented speech, and the implication of reaction time (RT) differences was explored.

A Cross-Modal Matching task was used (Experiment 1) in which listeners are instructed to compare a visual prime word with an aurally presented target. Target words were either accented or non-accented. The prime/target similarity was manipulated to exploit the relationship between the phonetic inventories of the foreign language (L1) and English (L2). Accuracy differences were found to be greatest when the prime and target differed by a single phoneme that was expected to be perceptually confusable given the L1/L2 relationship. No significant RT differences were found.

Experiment 2 used a Word Repetition task to provide a pure test of word intelligibility and perceptual effort. Listeners heard a word and had to repeat it as soon as it could be identified. FAS was less likely to be repeated correctly than NAS, and there was a 30-50 msec delay associated with the presence of the accent.

Previous studies that have investigated the perceptual effort (comprehensibility) of FAS have relied in large part on subjective ratings. Ratings of word comprehensibility were collected in Experiment 3 and correlated with the RT values of the Word Repetition task. There was a strong positive correlation between the two measures, indicating the potential for a shared underlying perceptual cause.

Overall, it was concluded that FAS lowers intelligibility, particularly for linguistic segments which were predicted to be confusable in English production. The RT differences suggest that lexical access is sensitive to variability due to accent and this sensitivity has the potential to negatively impact human performance.

Dedicated to Beren Gayle, the Ernie to my Bert.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank all of the people who have helped me get to this point. My advisor, Mark Pitt, has allowed me the freedom to follow my own path and has always given me sound research advice. I thank my dissertation committee for their insight and participation: Nadine Sarter, James Todd, and Keith Johnson. Tim Anderson and David Woods have provided me with invaluable guidance during the past several years. I cannot begin to thank them all adequately.

I have had some exemplary teachers, and I would be remiss if I did not take this unique opportunity to thank some of them for their encouragement, particularly Vivian Quam, Michael Sopko, Pat Russoniello, James Flatly, Deborah Stuby, Richard Pastore, and Colleen Reardon. I couldn't have done this without the support of my partners-in-crime, Katherine Smith, Lisa Shoaf, Erik Tracy, Melissa Jungers, Noah Mackenzie, and Bret Aarden who all listened to me kvetch for 5 straight years. Boaxia Liu, Maya Bukhman, Anna Fisher, Vince Schmidt, and Raymond Slyh all provided encouragement and material support.

This research was supported and funded in part by Sytronics, Inc. in conjunction with the Air Force Research Laboratory. I am grateful for having the opportunity to work with such a distinguished organization.

Most of all, I'd like to thank my parents, Steve and Shelly Weil, for their encouragement, love, and support. I couldn't ask for better.

VITA

October 21, 1975 …………………….………. Born – Boonton, New Jersey

1998 ……………………………………….. B.A., Binghamton University (SUNY), Vestal, New York

1999 – present ……………………………...Graduate Research Associate, The Ohio State University

2000 ……………………………………….. M.A. Psychology, The Ohio State University

2000 – present ……………………………...Human Performance Specialist, Sytronics, Inc., Dayton, Ohio

PUBLICATIONS

Research Publication

1. E.J. Crawley, B.E. Acker, R.E. Pastore, and S.A.Weil, "Error Detection in Multi-Voice Music: The Role of Musical Structure, Musical Training, and Task Demands." *Journal of Experimental Psychology: Human Perception and Performance, 28,* 367-378, (2002).

FIELDS OF STUDY

Major Field: Psychology

TABLE OF CONTENTS

Appendixes:

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

> Civil war raged in and around Kiev. Between 1918 and 1920 the city changed hands seventeen times, and one had to be very careful in adjusting to the constantly changing rules of conduct and ideologies. I developed a degree of virtuosity in handling the various documents, those suitable for the Bolsheviks, for the White Army, and for some Ukrainian nationalist groups. The one faction with which I was unable to cope was the band of Ukrainian anarchists, known as the Green Army, which fought both the Red Army and the White Army, and whose political slogan was classical in its simplicity, 'Kill the Jews! Save our souls!' The Green Army utilized a rather simple test to ferret out Jews by asking those suspected to say 'kukuruza', which meant corn. Suspects who failed to roll the 'r' the way the Russians and Ukrainians did were put to the sword.
>
> (Slonimsky, 1988)

In the course of normal daily activities, we hear spoken language which is highly variable in its acoustic realization. Any two voices differ in terms of fundamental pitch, voice quality, rate of speech, and a host of other factors. In most cases this does not cause difficulties for listeners – they understand all that is said. When both the talker and the listener are fluent native speakers of the same language, this variability may not even be noticed.

Some types of speech, such as synthetic speech, dysarthic speech, and the speech of the hearing impaired, include component segments which deviate from normative pronunciation to a large enough degree that they are outside the range of acceptable variability, and therefore degrade the speech understanding process (Nusbaum & Pisoni, 1985; McGarr, 1983; Elbert & McReynolds, 1978). In the same vein, foreign accented speech (FAS) is cited as an example of the type of talker variability which leads to an awareness of otherwise automatic perceptual normalization processes (see Nygaard and Pisoni, 1998). FAS adds additional variability to an already variable acoustic signal, which in turn affects the understandability of speech for listeners.

The story at the beginning of this chapter is extreme it its depiction of the consequences of having a foreign accent. In most situations, FAS is not life threatening. Still, situations in which FAS plays a role in communication are increasingly common. It is critical that the air traffic controller successfully communicate with pilots from many countries, all speaking English with different acoustic characteristics. A student in an American university must understand his French teaching assistant if she is going to get a good grade on her anthropology exam. A business executive must be confident that he understands the intricacies of what is being said to him by his Asian colleagues. A scientific understanding of FAS is necessary if we are to improve communication in these types of contexts.

The Origins of Foreign Accented Speech

For the purposes of this dissertation, an accent is defined as the deviation in speech production from a local norm due to the influence of a talker's native language (L1) to that of the spoken, non-native language (L2). As Trubetskoi (1939/1969) comments, second language speech is filtered through the "grid" of L1 phonology. Groups of non-native talkers who share a common L1 will naturally share common characteristics in their L2 production. For example, many Japanese speakers confuse the English /l/ and /ɹ/, which do not occur in Japanese. The misuse of these phonemes is a characteristic of the Japanese accent, and is not due to the idiosyncrasies of a single talker*.

During the 1960s, educators who specialized in English as a Second Language (ESL) instruction attempted to exploit this concept using *contrastive analysis* (Lado, 1957). Given that problems in L2 production are due to the intrusion of L1 characteristics, comparing these two languages at an element by element level should lead to predictions about the difficulty of L2 learning for specific accent backgrounds (Whitman 1970). With these predictions, ESL teachers could target instruction for

*This is a simplified explanation of L2 phonology – for comprehensive descriptions, see James (1996), Archibald (1998), and Major (2001).

particular student populations. It was not only L2 phonology that contrastive analysis was thought to apply to. Many aspects of L1 were thought to transfer to L2 production to some degree: grammar, syntax, vocabulary, etc.

Flege (1995) described in detail how the L1/L2 relationship can lead to speech that is accented as part of his Speech Learning Model (SLM). This model, in many ways a refinement of contrastive analysis concepts, aims to explain second language production as a function of the similarity of the L1 and L2 phonetic inventories, the talker's perception of the salience of second-language phonetic features, and the duration of the talker's experience with the second language. These factors influence the cognitive representation of phonetic categories which are essential to both the perception and production of speech (but see Best [1994] and Kuhl & Iverson [1995]).

Every language has a set of phones that are used to construct syllables, words, and sentences. The physiological and acoustic realizations of the phones depend on the surrounding phonetic context. Flege contends that at this allophonic level, the phonetic inventories of the native and spoken languages influence each other. This is a far more complicated interaction than was assumed by proponents of contrastive analysis, where it was expected that less detailed (i.e. phonemic) comparisons would be necessary to make predictions. The consequence of these relationships, following Flege, is that a phoneme may be noticeably accented in one context, and sound native-like in another.

SLM provides a theoretical framework for understanding – and predicting – characteristics of L2 production. Languages have a variable number of vowels which occupy the same physiologically determined acoustic space. According to SLM, L2 vowels that do not appear in the L1 inventory will be produced correctly if they are perceptually distinct from L1 vowels. If, however, the L2 vowel is perceptually similar to a vowel category in the L1 inventory, a non-native talker may use the closest L1 vowel, even if they are an imperfect match. This substitution is wholly dependent on the vowel inventories of the languages involved. Thus, a native Portuguese talker speaking French may pronounce the French /y/ (a close front rounded vowel) as /i/ (a close front unrounded vowel) because the Portuguese inventory does not have the /y/, and /i/ is the closest L1 vowel. In contrast, English talkers may use the vowel /u/ (a close back rounded vowel) in place of /y/ in French production because the /u/ is closer to the /y/ in English vowel space. The result is L2 vowel production that deviates from the native pronunciation (Rochet, 1995), but is unique to the specific L1/L2 combinations.

Similarly, L1 consonant production influences L2 realization. If L1 and L2 have a similar consonant in their respective phonetic inventories, and the talker cannot discern the difference, the talker will likely use the same phonetic category for both phonemes (i.e. equivalence classification). For instance, the Hebrew pronunciation of the word initial phone [b] is voiced throughout the labial closure (Laufer, 1999). In contrast, the English pronunciation of [b] is not voiced during the closure (Ladefoged, 1999). If the bilingual Hebrew talker does not perceive the difference between these two productions, a single phonetic category will be established, rather than two language specific categories. As a consequence, an Israeli talker may pronounce [b] in the same manner, regardless of the language that he is speaking. As in vowel production, this pronunciation is an imperfect match for L2.

If the L2 inventory includes phones that have no close correspondence in the L1 inventory, SLM predicts that a talker will either learn the new sound or substitute an L1 phone that is relatively distant from the L2 normative production, depending on his L2 experience. For example, [θ] (as in "thin") is not part of the phonetic inventory of Russian. As a consequence, Russian talkers may substitute [t] in its place. This may cause confusion in English language perception when both [θ] and [t] are both contextually appropriate, as in "Get me the thin/tin box." Consonant production in L2 speech deviates from the norm to varying degrees.

Following Flege's (1995) hypotheses, vowels and consonants will be pronounced differently by non-native talkers than the average native talker. However, not all words or segments spoken by a given non-native talker will be accented to the same degree. Instead, each accent has a profile in which some sounds will be further from normative L2 production than others. This profile of accentedness is only valid for a particular L1/L2 combination, and leads to L2 production which is characteristic of that accent. The profile would be different for the bilingual talker of another language with different phonetic characteristics.

Given a detailed description of the L1 and L2 phonetic categories at an allophonic level, perceptual consequences of non-native L2 pronunciation can be anticipated. However, the current project investigates not the production of L2 speech, but the subsequent *perception* of that speech by native speakers of L2. The question is not "what are the characteristics of accents in production," but instead "how do these characteristics affect perception." Can SLM be extended to predict perceptual difficulties in L2 perception? If so, what aspects of perception can it predict?

The Intelligibility of Foreign Accented Speech

A prerequisite for understanding speech in a conversation is recognition of the words that are spoken. Word recognition, in turn, is dependent on successful integration of component sounds. It is a subset of these sounds which are produced non-normally when spoken with an accent. As a consequence, communication between native and non-native talkers is made more difficult because some accented words are misinterpreted or confused for other words.

Speech is said to be intelligible when the words intended by a talker are conveyed successfully to a listener. Studies have sought to measure intelligibility of FAS experimentally using a variety of procedures, such as transcription accuracy (Munro & Derwing, 1995a, 1995b; Weil, 2002), phonetic and word discrimination (Van Wijngaarden, 2001; Rogers, 1997), sentence verification (Munro & Derwing, 1995a), and mispronounciation detection (Schmid & Yeni-Komshian, 1999). The conclusions are consistent – accented speech is less intelligible than non-accented speech (NAS).

Munro and Derwing (1995a) presented subjects with sentences spoken by either fluent accented talkers (Native Language [L1] = Mandarin) or non-accented talkers (L1 = American English). They used a pure transcription task – the participants were instructed to write down the words that they heard. These transcriptions were compared to the intended message, noting discrepancies. They found that the number of errors per talker was approximately three times greater for Mandarin talkers (63.6 errors) than for non-accented talkers (22.0 errors).

The intelligibility of the Mandarin talkers was not uniform among different utterances (Munro and Derwing 1995a). Although the Mandarin talkers had lower overall intelligibility, there were some utterances in which no errors in transcription ever occurred, and some which were *more* intelligible than corresponding non-accented baseline utterances. The reason for these unusual results was not determined, although the authors could not attribute it to sentence length, outlying talkers, or stronger contextual influences. Reevaluating these results with SLM in mind, perhaps the highly intelligible sentences contained fewer phonemes that are predicted to be produced non-normally, just by chance.

Schmid & Yeni-Komshian (1999) used a mispronunciation detection task with both accented (Spanish and Tamil) and non-accented stimuli. Subjects heard sentences in which the predictability of the final word was either high or low, given the surrounding sentential context (see Kalikow et al, 1977). Subjects had to respond if the initial phoneme of the final word was incorrect – if it made the final word a non-word. Accuracy was higher when the talker was a native English speaker (94% in the high predictability condition, 92% in the low predictability) compared to the accented speakers (77% and 67%). FAS reduced the ability to hear the mispronunciations.

Using two different tasks, Schmid & Yeni-Komshian (1999) and Munro and Derwing (1995a) both found gross differences in intelligibility between FAS and NAS.

However, understanding the underlying phonetic causes of the intelligibility gap requires a finer-grade methodology than was used in these studies. The intelligibility of specific linguistic segments – including those predicted to be produced non-normally by SLM – can be assessed by presenting subjects with phoneme and word length stimuli, and comparing the intelligibility of stimuli predicted to be confusable to different degrees. Van Wijngaarden (2001) did just this. After finding gross intelligibility differences between FAS and NAS, he investigated accent perception at the phonetic level to discover which accented segments were confusable for the native listener. Accented and non-accented consonants and vowels were presented in a neutral context to native talkers for discrimination from among a nearly exhaustive response set. For example, when testing the perception of the vowel /øː/, listeners would hear /dʒøːp/ and choose between *jaap*, *jup*, *jeup*, *jip*, and several other possibilities. The analysis indicated that vowels were more likely to be misidentified than were consonants compared to the baseline control group. This finding was especially strong for L2 vowels which are not present in the L1 phonetic inventory. This is consistent with SLM: vowels in L2 that are not present in L1 will likely be produced like perceptually close – but distinct – L1 vowels. This leads to non-normal L2 pronunciation and consequently lower performance in the perceptual identification task.

Predicting segments that will exhibit degraded perception entails descriptions of both the L1 and L2 languages at the contextually dependent allophonic level, and in-depth comparison of these details. While Flege (1995) presented evidence that confirmed his hypotheses for specific phonetic contrasts, this type of analyses may not be practical or feasible for languages in their entirety. The English language contains approximately 40 phonemes, but the number of allophones – acceptable variants of the phoneme based on the surrounding phonetic context – is far larger. A second language would have an equally large set of legal allophones, and the comparison of the first language to the second would be a monumental task. Instead, Rogers (1997) suggested that detailed descriptions of accent characteristics may be a reasonable alternative to the allophone-by-allophone comparison required for SLM to lead to predictions. The causes of accent-related differences in L2 production are only important for L2 perception inasmuch as they describe linguistic segments that are more or less susceptible to degraded perception. If these same segments can be predicted based on detailed descriptions of L2 production, the benefits of the consistency of the L1/L2 relationship can be exploited without the lengthy and perhaps impossible allophonic contrastive analysis. SLM can then be used after the fact, to explain the results of the accent analyses. Description of the accent leads to experimental predictions for data; explanations for the patterns in the data can then be drawn from SLM.

Detailed descriptions of L2 production characteristics have been undertaken in a number of disciplines. A large corpus of descriptions of accented speech is available in an on-line database, constructed by students of phonetics (Weinberger, 2003). Instructors of ESL have produced descriptions of the problems that individuals from specific language background face in L2 production (Chreist, 1964; Nissel, 1959; Kenworthy, 1997). Acting instructors have endeavored to describe accent characteristics indicative of various language backgrounds to achieve more realistic characters, (Herman and Herman

1943, Blumenfeld 2002). It is important to note that none of these sources compared the phonetic inventories of L1 and L2, as is suggested by in the SLM framework However, these descriptions detail which segments sound “more accented” to keen observation by a variety of experienced listeners.

Rogers (1997) performed a detailed phonetic analysis of Mandarin accented English with emphasis on identifying phones that would be confusable for native English listeners. Two native Mandarin talkers were recruited to record lists of real English words. The phonetic composition of the utterances was then analyzed and transcribed. Errors in pronunciation (e.g., non-normative production) were noted and categorized. From this error analysis, Rogers was able to identify phonemes and phoneme combinations that should be misidentified by native English listeners.

A minimal pairs test was developed for Mandarin accented English. English words (Targets) were chosen that contained phonemes which were likely to be mispronounced when spoken with a Mandarin accent. These words were paired with foils which differed from the Target words by only one error-prone phoneme. For example, the phonetic analysis indicated that Mandarin talkers occasionally confused the voiced fricative /z/ with the unvoiced fricitive /s/. For this error, the Target “peas” (/piz/) was paired with the foil “peace” (/pis/). Listeners heard “peas” uttered by a native Mandarin talker, and then saw both the “peas” and “peace” on the visual display. They then had to choose which of the two words they had heard. In essence, they had to discriminate between the phonemes /z/ and /s/ in a similar phonetic context.

These words were then recorded by new Mandarin talkers. Non-accented (English) listeners misidentified 17.4% of the accented utterances, choosing the similar foil over the intended utterance. This was significantly higher than the control condition, in which only 4.2% of words spoken by a native English talker were misperceived. Rogers found that she could predict lowered intelligibility by analysis of L2 speech, without having to contrast L1 and L2 at the fine-grain level that SLM requires.

It is clear from these studies that overall intelligibility is lower in FAS than in NAS. In both free transcription (Munro and Derwing, 1995a; Rogers, 1997) and phonetic discrimination (Rogers, 1997; Schmid & Yeni-Komshian, 1999, Van Wijngaarden, 2001), FAS words and sentences are more likely to be misheard, misidentified, and misunderstood. This finding is robust, and should occur in any inquiry into FAS perception. Furthermore, some evidence suggests that some accented segments are more detrimental to perception than others, and this seems to be determined in part by specific accent characteristics. However, in no study have phonetic contrasts been varied experimentally as a function of their predicted effects on intelligibility. One intention of the current study is to provide this manipulation.

The Comprehensibility of Foreign Accented Speech

While FAS does degrade intelligibility, it may also affect other aspects of understandability. Only so much can be learned about FAS perception by measuring the number of word errors that listeners make. In the air traffic control room, operators converse with individuals with diverse linguistic backgrounds in a complex environment.

Even if the controller understands all of the words that are spoken to him, is his comprehension as effortless with accented talkers as with non-accented? What effects does FAS have on his performance in co-occurring tasks? The term *comprehensibility* has been used to indicate the difficulty of perception. It refers to "… how difficult or easy an utterance is to understand" (Derwing and Munro 1997, page 2).

One way to measure comprehensibility is via subjective ratings. In one study (Munro & Derwing 1997), listeners heard accented sentences and had to both transcribe what was heard and judge comprehensibility on a numeric scale. Approximately 89.5% of the words uttered were correctly identified, or had trivial changes in transcription; 60% of sentences were transcribed correctly in their entirety. However, a fair number of sentences which were correctly transcribed received low comprehensibility ratings; comprehensibility ratings were less positively skewed than were intelligibility scores. Munro and Derwing (1997) determined that listener comprehensibility ratings were "harsher" than their transcription scores; a talker could be rated as being very difficult to understand, but still elicit high word transcription accuracy. This suggests that the subjective rating does not necessarily reflect accuracy differences, but may instead be a reflection of perceptual effort. Some of the words in those sentences contained phoneme pronunciation that was far from normative pronunciation. While many of these words were identified correctly, it took more effort to make those identifications. Participants may have noted the extra effort and based their ratings on their subjective impressions of the severity of that effort. However, these results could also be explained as a post-perceptual bias; listeners realize that the talker had an accent, and automatically rate the talker as being difficult to understand. What is needed to dismiss this possibility is an objective, on-line measure of perceptual effort.

Reaction time (RT) has long been used in cognitive science and psycholinguistics to compare the processing time differences resulting from different stimuli. Longer processing times are thought to indicate either extra stages or delayed/non-optimal processing. As Borsky, Tuller, and Shapiro (1998) note, “We assume that processes that theoretically require more ‘processing resources’ or are more difficult to complete will be associated with longer response times.” This seems to correlate well with the definition of comprehensibility put forth by Munro and Derwing (1995a). Perhaps the subjective measures of comprehensibility correspond with RT, an empirically measurable variable.

Consider a model of word perception in which a heard utterance is compared to cognitive representations of words in a mental lexicon. These entries are based on the linguistic experience of the listener. A lexical entry becomes active when there is enough evidence (i.e., similar acoustic or phonetic composition) that the heard utterance and the entry are the same. Activation of the lexical entry intended by the speaker indicates that the word was intelligible; activation of a competitor means the utterance was not intelligible. The amount of evidence needed to reach the activation point, or threshold, is a function of a number of factors, such as the number of similar words in the lexicon, the familiarity of the word to the listener, and the sentential context in which the word is heard. The more evidence that is needed, the more effort required for a word to be activated. All else being equal, an utterance that is highly similar to the lexical entry will require little effort, while those which differ from the entry will require more effort. FAS differs noticeably and predictably from the pronunciation of native speakers, and should require more resources and effort activate the correct lexical entries. It is proposed here that the similarity between accented words and the cognitive representations of the these words in the lexicons of native talkers will be reflected in the RT measurements.

The perception of speech is a dynamic process, one that occurs over time. Intelligibility refers to the number of words identified as intended, which is the end result of the speech perception process. Comprehensibility, in contrast, refers to the difficulty of this perception, which is a characteristic of the process itself. Measures of comprehensibility have the potential to increase our understanding of these perceptual processes with far more sensitivity than simple intelligibility measures. In conjunction with intelligibility, the understanding of these processes is far greater than intelligibility alone.

Munro & Derwing (1995b) investigated the use of RT for the measurement of comprehensibility using a sentence verification task. Sentences, spoken by 10 talkers of Mandarin and 10 native talkers, were presented to native English listeners. Participants had to judge the truth or falsity of sentences such as “Italy is a country in Europe” and “The inside of an egg is blue.” Response times were measured and correlated with subjective ratings for each utterance. In general, the Mandarin talkers elicited slower response times than the native control talkers, despite high accuracy. There was a positive relationship between subjective ratings of comprehensibility and RT; RT was slower for sentences that were rated “Difficult” compared to those rated “Moderate” or “Easy” – a difference of approximately 500 msec.

In the mispronunciation detection task described in the previous section, Shmid and Yeni-Komshian (1999) also collected RT. Subjects heard sentences spoken with or without an accent, and had to indicate if they heard a mispronunciation. Responses to the accented talkers were slower than the non-accented talkers, both in the high predictability (818 msec vs. 749 msec) and low predictability context (874 msec vs. 849 msec). FAS slows mispronunciation detection.

Although both of these studies found the expected relationship between accentedness and RT, interpreting these differences is challenging. The complexity of Munro and Derwing's (1995b) truth assessment task makes it difficult to know which processes are being measured. A listener had to hear the sentence, identify some or all of the words, understand the meaning of the sentence in its entirety, and then evaluate if the sentences were true or false. The RT measurement captures all of this processing, so it is impossible to discriminate between processing delays due to the accent from those due to sentence comprehension or to the truth judgment. Shmid and Yeni-Komshian's (1999) mispronunciation detection task seems less complex than the truth assessment task. However, it is still difficult to know what processes are being measured. It may be reasonably assumed that the differences in RT values are measurements of the differences in processing time required to identify mispronunciations as a function of accent. But it is the time it takes to understand *correctly* produced sentences that reflects real world word processing. Shmid and Yeni-Komshian did not collect responses for sentences in which there were no errors in pronunciation; they only reported RT for *hits*, not for *correct rejections*. It may be that native listeners are not slowed by FAS if there are no contrived errors. Also, perception of the last word of a sentence may be influenced by degraded intelligibility of the context sentence, which was spoken by the same non-native talker.

Are the two measures – ratings and RT – capturing the same processes? Ratings tasks are inherently subjective – RT may be a better metric, if it reflects the same processes as the ratings of Munro & Derwing (1995b). To evaluate this, both ratings and RT must be collected for the same utterances, and these responses compared. Similar patterns may indicate shared origins.

In the preceding discussion, the intelligibility of specific linguistic segments was proposed to be graded due to the factors cited in Flege's (1995) SLM. In addition to a general intelligibility difference between FAS and NAS, segments predicted to be perceptually distant given a L1/L2 paring were hypothesized to have disproportionably lower accuracy. Assuming that the hypotheses of SLM lead to the ability to make predictions of L2 intelligibility, do these same segments take longer to perceive when they are intelligible? Accented speech samples which contain segments predicted to be far from normative pronunciation should differ from the mental representation of those words to a greater degree than other accented words. If this is true, activation RT should be slower for words which contain these specific segments. Comprehensibilty can be predicted using the same methods used to predict intelligibility, and these predictions could be tested using similar designs.

Goals

The current series of experiments was designed with two major goals in mind. The first goal is to assess the impact of FAS on word intelligibility. Previous studies have indicated that FAS lowers intelligibility. Following Flege's SLM (1995), it should be possible to predict the circumstances in which this intelligibility cost is incurred. The second goal is to measure the effects of FAS on word comprehensibility, both in response time and in subjective rating. The utility of RT as a metric for assessing comprehensibility is unclear given the poverty of the existing literature. By using RT measurements in addition to ratings of effort and comparing the results, we can simultaneously measure comprehensibility and assure that the current findings echo what has been referred to as comprehensibility in previous literature. The perceptual cost associated with FAS should not be limited to lowered intelligibility, but also reflected in lower comprehensibility ratings and slower RT.

A discrimination task (Experiment 1: Cross-Modal Matching) was used in which words containing specific phonetic contrasts varying in their predicted confusability were presented to native English speaking subjects. A second, identification task (Experiment 2: Word Repetition) was conducted, in which listeners repeated presented utterances. The Cross-Modal Matching and Word Repetition tasks are ideally suited for the investigation because they can simultaneously measure intelligibility and RT for the same presented word using slightly different methods. This will provide converging evidence that FAS negatively impacts word intelligibility, and that the errors in production are predicted in large part by the arguments of Flege's SLM (1995) and Rogers (1997).

Ratings of perceptual effort (Experiment 3: Comprehensibility Judgment) for these same words were undertaken to provide evidence that the RT data reflects the same processes investigated by Munro and Derwing (1995a) and called comprehensibility.

CHAPTER 2

MATERIALS

The same stimulus materials were used throughout this project. While the details of the designs employed will be described in upcoming chapters, the characteristics of the shared stimulus materials is outlined below. For each accent utilized (Mandarin and Russian), a list of words was compiled (List M and List R). The words included phones that were predicted to be confusing to native English speakers when spoken with these specific accents. Accented and English talkers were recruited to record these words for use as stimuli.

Mandarin Stimuli

Rogers (1997) included a phonetic analysis of the production of Mandarin accented English words in her study of the intelligibility of non-native speech. Based on this analysis, a subset of English phonetic contrasts was predicted to be especially difficult for Mandarin talkers to produce correctly, and for native English listeners to subsequently perceive as intended. These contrasts are listed in Rogers (1997) tables 1 and 2, and included 52 consonant confusions and 20 vowel confusions. Confusable pairs were not limited in terms of phonetic categories – stops, nasals, fricatives, and vowels were all found to be confused. These contrasts varied on several dimensions, including phoneme type and position in word. A minimal pairs test was developed to confirm the difficulty of pronunciation of these specific contrasts. Each pair consisted of two words which differed by only one phoneme predicted to be confusable when spoken with a Mandarin accent. Rogers found that these words were difficult to discriminate

List M was made up of 228 real English words (Appendix A, "Target" words). 180 test words were taken from the Rogers (1997) minimal pairs test. The remaining 48 words were selected by the author for use as stimuli in practice sections of each experiment. The choice of words was not constrained by word familiarity or part of speech. Choice of words was constrained, however, by the design parameters of the Cross-Modal Matching task (Experiment 1), which mandated that all words used include at least one phoneme predicted to lead to non-normal perception. Words were made up of a consonant (C) or consonant cluster followed by a vowel (V) followed by a final consonant or consonant cluster (C). For example, "sap" (CVC), "slap" (CCVC), "saps" (CVCC), and "slaps" (CCVCC) were all acceptable words.

There is great variability in speech. Native speakers of a language who share a common education, linguistic background, and residence may vary significantly in voice quality, speaking rate, fundamental frequency, and a host of other variables. All of these factors influence the understandability of speech, the consequence being that some talkers are more understandable than others. When the individual is not a native speaker of the language, the potential for variability increases greatly. The quality of a talker's L2 instruction, years of immersion in an L2 environment, and perhaps natural ability for phonetic mimicry may all affect the resulting L2 speech production, in addition to factors that influence the production of all individuals who share a common L1.

This degree of variability poses a problem for a study of the type described in this project. It is not possible to find a single speaker of a foreign language who has *the* Mandarin or Russian accent, any more than it is possible to find a single native English talker who's dialect reflects the range of native English speech production. For this reason, multiple talkers within each accent were used in an attempt to approximate the variability of accented speech and thus assure maximum generalization to the accent while limiting adaptation to particular talkers (Weil, 2002b). Variability was further reduced by choosing talkers with many demographic commonalities, such as sex, place of primary residence during L1 and L2 instruction, educational background, age, and years of L2 instruction.

To record List M, ten male talkers were recruited: 5 native Mandarin talkers and 5 native English talkers. Table 2.1 presents the biographical information about the talkers used in List M. To reduce variability, effort was made to find a relatively homogeneous Mandarin accented population in Columbus, OH. The Mandarin talkers were all male, and were raised in the greater Beijing area and surrounding Chinese provinces. All of the talkers were Ohio State University graduate students and reported a usable knowledge of English. The average age of the Mandarin talkers was 25.8 years, with an average of 2.5 years living in the United States and an average Age of First English Instruction (AFI) of 12 years old.

The English talkers used in List M were all male Ohio State University undergraduates. All were born in proximity of Columbus, OH. The parents of all

English talkers were themselves born and raised in the United States. None of the English talkers reported fluency in languages other than English. The average age of the English talkers used in List M was 18.6 years.

Talker	Talker L1	Place Raised	Age	AFI	Years in US
Man1	Mandarin	Xintai, Shandong, P.R.China	27	12	1
Man2	Mandarin	Xi'an, Shann Xi, P.R.China	23	11	1
Man3	Mandarin	Qiqihaer, Heilongjiang, P.R.China	29	11	3.5
Man4	Mandarin	QingDao, Shandong, P.R.China	29	14	1.5
Man5	Mandarin	Xi'an, Shann Xi, P.R.China	21	-- *	7
Eng1	Ohio English	Columbus, OH, USA	22	n/a	n/a
Eng2	Ohio English	Springfield, OH, USA	18	n/a	n/a
Eng3	Ohio English	Grove City, OH, USA	17	n/a	n/a
Eng4	Ohio English	Columbus, OH, USA	18	n/a	n/a
Eng5	Ohio English	Pickerington, OH, USA	18	n/a	n/a

Table 2.1: Biographical information about talkers in Mandarin variations. (* did not reveal AFI)

The average length of the 180 test words was calculated from the onset of the initial phoneme to the end of phonation. The length was 481 msec (SD = 90) for English talkers and 556 msec for the Mandarin talkers (SD = 123). An independent samples t-test indicated that these samples were significantly different, $t(178) = -4.64$, $p<.001$.

Russian Stimuli

For the List R, ten male talkers were recruited: 5 Russian talkers and 5 native English talkers. Table 2.2 presents the biographical information about the talkers used in

List R. Certain biographical information limited the choice of talkers used to constrain variability in the accent characteristics. Only Russian talkers from urban areas in the northern part of Russia were recruited, insuring that accent characteristics were consistent[†]. All of the Russian talkers used were male Ohio State University graduate students. The average age of the Russian talkers was 25 years, with an average of 4.2 years living in the United States and an average AFI of 14.2 years old.

Talker	Talker L1	Place Raised	Age	AFI	Years in US
Rus1	Russian	Nizhny Novgorov, Russia	24	11	2
Rus2	Russian	Yakaterinburg, Russia	20	16	4
Rus3	Russian	Moscow, Russia	27	10	5
Rus4	Russian	Adygeya, Maikop, Russia	27	12	5
Rus5	Russian	Tomsk, Russia	27	22	5
Eng1	Ohio English	Columbus, OH, USA	21	n/a	n/a
Eng2	Ohio English	Bayview, OH, USA	20	n/a	n/a
Eng3	Ohio English	Sagauwe, OH, USA	22	n/a	n/a
Eng4	Ohio English	Grove City, OH, USA	20	n/a	n/a
Eng5	Ohio English	Westerville, OH, USA	21	n/a	n/a

Table 2.2: Biographical information about talkers in Russian variations.

The English talkers were all male Ohio State University undergraduates. None of the talkers used in List R had been recorded for List M. All were born in proximity of Columbus, OH. The parents of all English talkers were themselves born and raised in the

[†] Criteria for talker recruitment was based on the recommendations of Daniel Collins chair of the Department of Slavic and East European Languages and Literatures at Ohio State University (personal communication, April 8, 2003).

United States. None of the English talkers reported fluency in languages other than English. The average age of the English talkers was 20.8 years.

Instead of choosing L2 contrasts based on expert phonetic analysis, the words used in List R were based on two less analytical sources. Descriptions of Russian accent characteristics are found in books designed for acting instruction (Blumenfeld, 2002; Herman & Herman, 1943), and include information about the deviations in English pronunciation that are distinguishing for specific accents. The word choice criteria for List R differed from List M in part to assess the potential of using non-technical descriptions of accent characteristics for predicting perceptual difficulties. For example, Blumenfeld (2002) writes "There is no voiced TH [/ð/] in Russian and no voiceless th [/θ/]. [When playing a Russian character, you should] substitute 'v' and 'f' for voiced and voiceless, or 'd' and 't'."

These descriptions were evaluated by a native Russian talker. In an interview, this informant described her experiences learning English speech production, and provided specific examples of English production that were difficult for her and other native Russian talkers to learn and pronounce, such as word final consonant voicing. Her intuitions supported the characterizations described by the actor's training manuals as well as the reflections of other Russian learners of English. In his biography *Perfect Pitch,* Russian émigré musicologist Nicholas Slonimsky wrote, "The worst obstacle in my pilgrim's progress were the English phonetics – I could not tell the difference between 'food' and 'foot', 'hat' and 'head'. The length of each vowel did not register in my ears, and the final 'd' and 't' sounded identical" (pp. 87-88).

Based on both of these sources, English phonetic contrasts were chosen that were predicted to be difficult to discern when spoken by native Russian talkers. These contrasts are listed in table 2.3. In total, there were 8 vowel confusions and 8 consonant confusions were identified. Based on these contrasts, a list of 180 test words and 48 practice words was constructed and recorded. The English stimuli were significantly longer than the Russian words (586 msec vs 537 msec), $t(178) = 3.03$, $p<.005$. The list of the words used in List R is included in Appendix B (Target Words).

In addition to the practice stimuli recorded for Lists M and R, a third set of practice stimuli was recorded for use in Experiments 1 and 2. To remove possible effects of talker adaptation (Weil, 2000), a native English talker (the author) recorded 24 words for use in the practice sessions that preceded each experiment. These sessions were used primarily to introduce the task demands to the participants. The author is a male graduate student, enrolled at The Ohio State University. He is the child of native English talkers, was raised in New Jersey, and was 27 years old at the time of recording.

Recording Procedures

Talkers were individually recorded in a sound attenuated booth while the experimenter sat in an adjacent room with the recording equipment. Talkers were first given several paragraphs to read, to calibrate recording equipment and to allow the talkers to acquaint themselves with the microphone recording procedures. English talkers recorded three paragraphs written to reflect the phonetic variety in English: "The North Wind and the Sun" (IPA, 1999), "Rainbow" (Fairbanks, 1940), and "Please Ask Stella"

Vowels	Consonants
/e/ with /ɛ/	/dʒ/ with /tʃ/
/æ/ with /ɑ/	/s/ with /ts/ (WF)
/i/ with /ɪ/	/w/ with /v/
/ɛ/ with /æ/	/j/ with /dʒ/ (WI)
/ɪ/ with /i/	Inserted /j/ (WI)
/ɑ/ with /ɔ/	/p/ with /b/ (WF)
/u/ with /ʊ/	/t/ with /d/ (WF)
/ʌ/ with /ɔ/	/k/ with /g/ (WF)

Table 2.3: English phonetic contrasts expected to be confusable for Russian talkers.

(Weinberger, 2003). Accented talkers recorded these passages as well as Mandarin or Russian translations of "The North Wind and the Sun," as translated by native talkers of those languages.

Written words and paragraphs were presented to the talkers to read. Talkers were encouraged to ask the experimenter if any of the words were unfamiliar to them. Talkers were asked to read each of the words out loud three times. The utterances were monitored for potential misreading, and any words which included obvious dysfluencies were rerecorded.

The microphone used was a head-mounted Crown CM-311A Differoid Condenser. Stimuli were recorded onto Digital Audio Tape using a Tascam DA-30 MKII DAT recorder and a Yamaha MV802 mixer. The recordings were then transferred to PC via a Zefiro Acoustics ZA2 digital sound card. Recordings were mono, with a sampling rate of 22050 and 16 bit resolution. Each recorded word was amplitude normalized.

CHAPTER 3

EXPERIMENT 1: CROSS-MODAL MATCHING

The initial experiment in this project examines FAS perception in a discrimination task. The ideas proposed in Flege's (1995) SLM suggest that, for a given L1 and L2, the accentedness of particular L2 phones will be graded. As a result, FAS will be less intelligible than NAS, and this disparity should be greatest for segments which are perceptually distant from the cognitive representation of normal pronunciation. Experiment 1 is designed to addresses the consequences of accent on the subsequent perception of L2 speech, both in terms of intelligibility and comprehensibility.

In her minimal pairs task, Rogers (1997) found that phonemes that were predicted to be confusable in Mandarin accented speech were more difficult for native listeners to discriminate. For example, Rogers found that Mandarin talkers occasionally confused /p/ with /b/ but never /p/ with /k/ in their English production. Consequently, a native Mandarin talker's pronunciations of "tap" and "tab" may be confusable for native English listeners, and Rogers found this to be true. There were significant numbers of accented phonemes which were ambiguous for native English listeners.

One problem with the minimal pairs design is that subjects were not presented with a control group of accented words that contained alternate phonemes expected to be perceptually distinguishable. This is vital because it eliminates the possibility that all accented phonetic pairs are difficult to discriminate regardless of the L1/L2 relationship, or that confusability is a function of global phonetic similarity, independent of accent characteristics. For example, Rogers found that Mandarin talkers never produced /p/ in a way that would make it confusable with /k/; therefore "tap" and "tack" should be easily discriminated. Rogers did not present this contrast. If those words are found to be confusable, then the predictions and conclusions of Rogers (1997) would have to be reassessed. The current study includes both the predictably confusable and non-confusable contrasts, specific to two language (Mandarin and Russian) to address this concern.

A Cross-Modal Matching task was employed in which a participant is presented visually with a word (Visual Prime), and then aurally presented with a comparison word (Auditory Target). The participant must then judge if the Prime and Target represented the same or different words, and respond via a button press (buttons labeled "Same" and "Different"). Both the accuracy and speed of the response were measured. This paradigm was chosen because it allows simultaneous measurement of accuracy (to assess intelligibility) and reaction time (as a gauge of comprehensibility), as a function of predicted word discrimination difficulty. Two variations of this design were conducted; one with English and Mandarin talkers, and a second with English and Russian talkers.

Targets and Primes

The Targets used were those recorded words described in Chapter 2. They consisted of real English words uttered by both accented and non-accented talkers. Each Target included a phoneme predicted to be confused with another phoneme by native English listeners, given the L1/L2 relationship and the descriptions of accented English speech production (Rogers, 1997; Blumenfeld, 2002; Herman & Herman, 1943). These confusable phonemes could occur word initially, word medially, or word finally, and could be either consonants or vowels. The proportion of initial, medial, and final phonemes is unequal because the make up of each stimulus list was determined by the number and type of predicted phonetic confusions. For example, Blumenfeld (2002) and Herman & Herman (1943) mentioned few word initial phonemes that would be confusing when spoken with a Russian accent. As a consequence, there were only 3 word initial phoneme confusions presented to subjects in the Russian variation. Table 3.1 summarizes locations and phonation types of these confusable phonemes.

Phonation Type:	**Consonant**		**Vowel**	
Position in Word:	**Initial**	**Final**	**Medial**	**Total**
Mandarin	42	69	69	180
Russian	3	87	90	180

Table 3.1: Within word placement of confusable phonemes in the stimuli used in the Cross-Modal Matching task.

There are four Visual Prime conditions: Same, Confusable, Non-Confusable, and Dissimilar (table 3.2). The Same primes were identical to the intended recorded word; the subject saw and heard the same word. The other three Prime conditions were mismatches in which subjects saw one word and heard another. Confusable primes were identical to the Targets, except that the phoneme predicted to be confusable was replaced by the phoneme that it was predicted to be confused with. The Non-Confusable primes also differed from the Target by only one phoneme; the phoneme predicted to be confusable was replaced by a phoneme that was predicted to be perceptually distinct. The Dissimilar primes had no phones in common with the auditory target. Note that these words could contain more than one phone predicted to be confusing given a particular accent. However, only one phone per word was contrasted in each Confusable and Non-Confusable trial.

Prime Type	**Visual Prime**		**Auditory Target**	**# Trials**
Same	"tap"	/tæp/	/tæp/	90
Confusable	"tab"	/tæb/	/tæp/	30
Non-Confusable	"tack"	/tæk/	/tæp/	30
Dissimilar	"fell"	/fɛl/	/tæp/	30

Table 3.2: An example of Prime and Target types in the Cross-Modal Matching task.

Following the predictions of Rogers (1997) and the implications of the SLM, a distinct pattern of discrimination results should emerge. Accented words should be less discriminable than non-accented words. In the Same condition, this is a pure test of word intelligibility, and FAS should be less intelligible than NAS. Accuracy differences in response to accented and non-accented words should occur on both the Confusable and Non-Confusable conditions. However, if the phonemes of greatest confusability can be predicted by SLM or error analysis, the magnitude of these differences should be significantly larger in the Confusable condition than in the Non-Confusable and Dissimilar conditions.

It is unclear what the RT data will look like. While RT differences based on FAS have been found in previous studies, the exact causes of these differences are difficult to determine. If RT is a measure of perceptual effort, it is likely that RT will vary as a function of the gross phonetic similarity of the Primes and Targets. The greater the difference, the faster the RT should be, all other factors being equal. In the Dissimilar condition, this difference is great because there is little in common between Prime and Target. As a consequence, response time should be relatively fast. In the Confusable and Non-Confusable conditions, the decision process may be more laborious because of the overall similarity of the Prime and Target; they share all but one phoneme in common. Response time in the Same condition will likely be relatively fast because the match between Prime and Target is perfect.

Of interest are the effects of the finer phonetic variation due to accent. Differences in RT should be found between accented and unaccented stimuli because the accented utterance is farther from the listener's perceptual representation of the intended word than are non-accented equivalents. As in the intelligibility results, these accent differences should be greater in the Confusable condition than in the Non-Confusable, but only if the perceptual similarity of the spoken utterance and the listener's representation of that utterance determines decision speed. There should be no difference in the Dissimilar condition because the accented and non-accented utterances are both perceptually distinct from the Dissimilar Primes. In the Same condition, the accented stimuli will likely lead to slower RT than the non-accented because the Target utterance will be perceptually farther from the listener's representation of the Prime.

Experimental Methods and Procedures

A total of 50 listeners participated in the Mandarin variation, and 53 in the Russian. All received credit in an introductory psychology course. To qualify for participation, both the participants and their parents had to be native English speakers, born and raised in the United States. Subjects reported fluency in only one language. One participant in the Mandarin variation and one participant in the Russian indicated non-normal speech or hearing; their data were subsequently eliminated from analyses.

Between 1 and 4 participants were tested during a given session. Each was seated in an individual sound attenuated booth, and instructed that they would be seeing words on a computer monitor and then hearing words over headphones. Their task was to decide if the words they saw and heard were the same or different, and reply via a button press, with buttons labeled "Same" and "Different." Participants were told to use their two index fingers to make their responses. Both their response and the response time were collected. Words were presented on color monitors with a 7" x 13" display. Sony MDR-V900 headphones were used to present the auditory stimuli.

For each of the practice and test trials, participants saw the visual prime in 1.5 cm tall black letters in the center of a white background. The regular Times New Roman font was used. The visual prime remained on the display for 750 msec, after which the screen was cleared. Following a 500 msec pause, the auditory target was presented via the headphones. Participants were encouraged to respond to the auditory stimuli as soon as they knew if it was the same or different than the visual prime – speed was emphasized. Response time was measured using a Acromag precision timer and recorded onto a computer data file

Each session began with 24 practice trials, during which the experimenter monitored the performance of each participant. Further instruction was given to those participants who were not performing the task as expected. The practice trials were spoken by a native English talker whose voice was not used in the subsequent test session. The test session followed, which began with 12 practice trials and 90 test trials. There was a short break, followed by 12 additional practice trials and 90 test trials. The total time for the experiment was approximately 50 minutes. Following the experiment, a short questionnaire was distributed.

For each variation, six counterbalanced stimuli lists were created to assure equal presentation of all talkers and prime conditions to each subject. In each of these lists, the 180 Target words were presented to the participants. Fifty percent of words were uttered by accented talkers, and 50% by native English talkers. All subjects heard the same talkers speak the same words, and they heard each word only once. However, the Visual

Prime was varied among lists; 50% of the participants heard a given word paired with the identical Prime (Same), and 50% of participants heard the same word paired with a distinct Prime (Confusable [16.67%], Non-Confusable [16.67%], or Dissimilar [16.67%]). Within each session, the order of trials varied randomly.

Results and Discussion

Intelligibility.

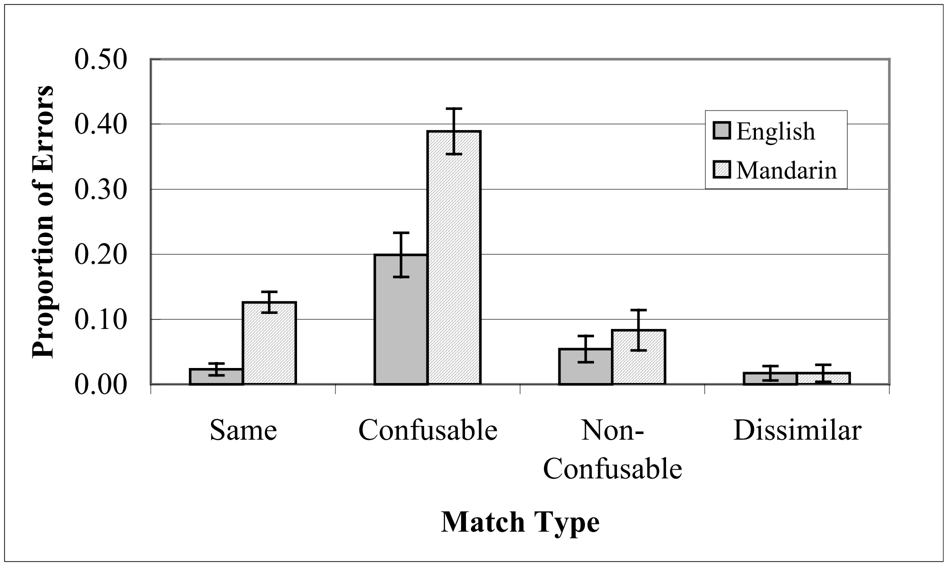
Similar accuracy patterns were found for the Mandarin and Russian variations, and they will be discussed together. For each variation, a 2x4 repeated measures Analysis of Variance (ANOVA) was conducted[‡], with Accent and Prime as the two within-subjects factors and proportion of errors as the dependent measure. For the Mandarin variation (figure 3.1a), there were significant main effects of both accent $F(1, 47) = 124.43$, $p<.001$, and prime type, $F(3,141) = 193.36$, $p<.001$, as well as a significant interaction of accent and prime, $F(3, 141) = 43.56$, $p<.001$. This pattern of results was repeated in the Russian variation (figure 3.1b): there were significant main effects of both accent, $F(1, 51) = 51.43$, $p<.001$, and prime type, $F(3,153) = 203.95$, $p<.001$, as well as a significant interaction of accent and prime, $F(3, 153) = 16.9$, $p<.001$. Interpretation of the main effect of accent is straightforward. The accented utterances led to a higher proportion of errors than did the non-accented utterances (Mandarin 10% more; Russian: 5% more). The main effect of Prime condition can be accounted for as well. For all

[‡] Preliminary analysis suggested that performance did not vary as a function of Balancing List; this factor will not be considered in the subsequent analysis and discussion.

language backgrounds, the Confusable prime led to the highest proportion of errors followed by the Non-Confusable, Same, and Dissimilar primes. The significant interaction was due to the magnitude of the differences among the Prime conditions – the contour of the pattern of results was the same regardless of language background of the talker, but the degree of the differences varied as a function of Prime condition. Tukey's post-hoc tests reveal significant differences between the accented and non-accented stimuli in the Same, Confusable, and Non-Confusable prime conditions (Mandarin HSD = .0338; Russian HSD = .0126; $p<.01$), but not in the Dissimilar condition.

Of great interest is the comparison between the responses to the Confusable and Non-Confusable conditions in each variation. The major difference between these conditions was that the phonetic contrast in the Confusable condition was predicted to be problematic for intelligibility, while the contrast in the Non-Confusable was not. Between subjects, the same auditory Target was paired with either Confusable or Non-Confusable primes. Thus it is possible to directly compare responses between the conditions, on a Target by Target basis. The difference in accuracy results is an indication of the relative discriminability of the phonetic contrasts. A difference of zero would indicate that the Confusable and Non-Confusable alternatives were equally discriminable from the target. This difference in accuracy is presented in figure 3.2, for accented and non-accented stimuli in both variations. While the difference was greater than zero for both the native and non-native talkers, paired sampled *t*-tests confirm that this difference was significantly larger for the Mandarin and Russian accented words than for non-accented controls (Mandarin: ~15% larger difference, $t(47) = -7.641$, $p<.001$;

a



b

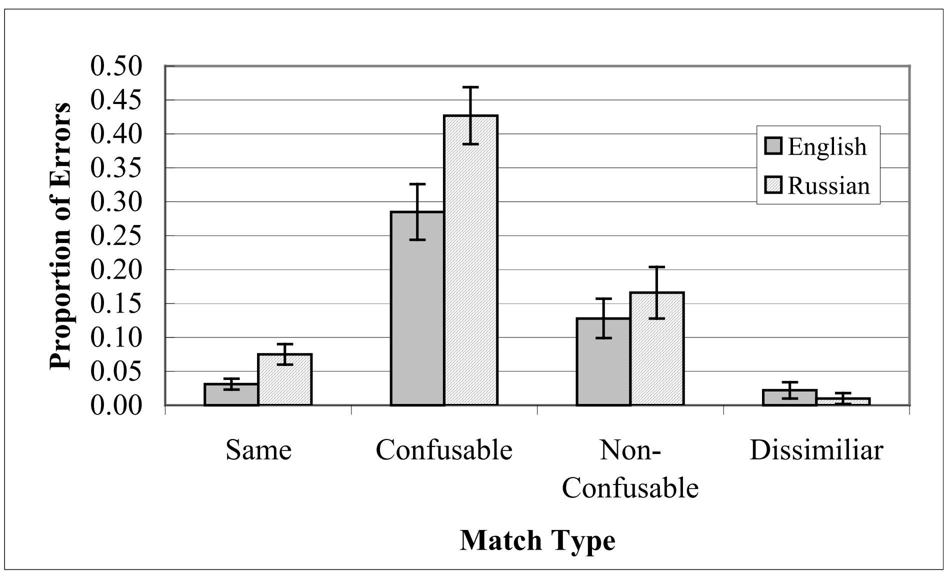


Figure 3.1: Accuracy results for the Cross-Modal Matching task. Error bars indicate 95% confidence intervals. (a: Mandarin Variation; b: Russian Variation).

Russian: ~10% larger difference, $t(51) = -3.528$, $p<.001$). This supports the conclusion that those contrasts which were predicted to lead to lower discriminability were indeed less discriminable.

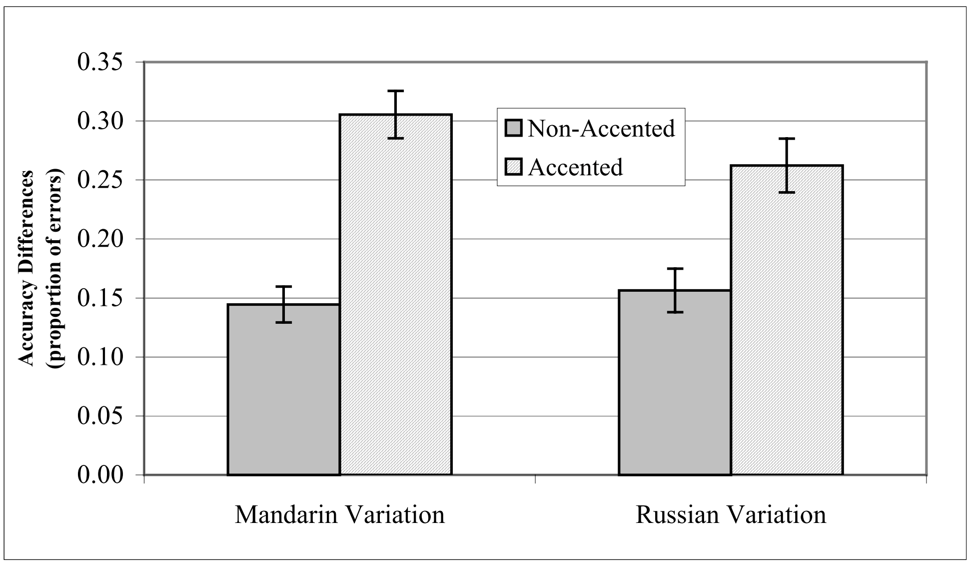


Figure 3.2: Accuracy Difference Scores: (Non-Confusable Accuracy) – (Confusable Accuracy). Error bars indicate standard error.

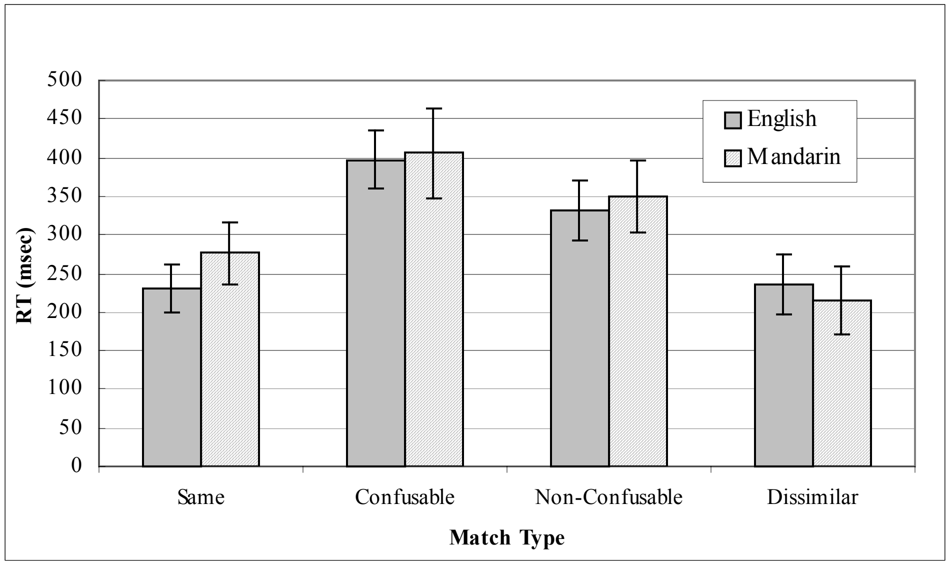
Comprehensibility

Additional ANOVAs were conducted with accent and prime as within-subjects factors and RT as the dependent measure (figure 3.3a, 3.3b). RT was calculated from the offset of the word to the response. Only the RTs for correct responses were considered in this analysis. Because slightly different patterns of results were found between the Mandarin and Russian variations, they will be discussed separately.

In the Mandarin variation, the main effect of prime was significant, $F(3, 141) = 103.74$, $p<.001$. Replicating the profile of results found in the accuracy data, the Confusable condition led to the slowest responses, followed by the Non-Confusable, Same, and Dissimilar conditions. The time needed to make a correct response increased as the gross phonetic similarity increased. The main effect of Accent was not significant, $F(1, 47) = 2.1$, ns. Responses to accented words were not slower or faster than responses to non-accented words. The interaction between Prime and Accent factors was significant, $F(3, 141) = 4.71$, $p<.05$. Tukey's post-hoc tests reveal a significant difference between the RT for accented and non-accented words in the Same condition (45 msec), but not in the other conditions (Tukey's HSD = 30 msec).

In the Russian variation, the pattern of results was slightly different. There was a significant main effect of Prime, $F(3, 153) = 142.38$, $p<.001$. Like the Mandarin variation, RT increased as gross phonetic similarity increased. However, unlike the Mandarin, there was also a significant main effect of Accent, $F(1, 51) = 71.26$, $p<.001$. Responses to accented words was slower than to non-accented words, as predicted. There was also a significant interaction ($F[3, 153] = 2.7$, $p<.05$). The difference in response times between Russian and English words was significant in all Prime conditions (Tukey's HSD = 20.5 msec), but was largest in the Same condition.

a



b

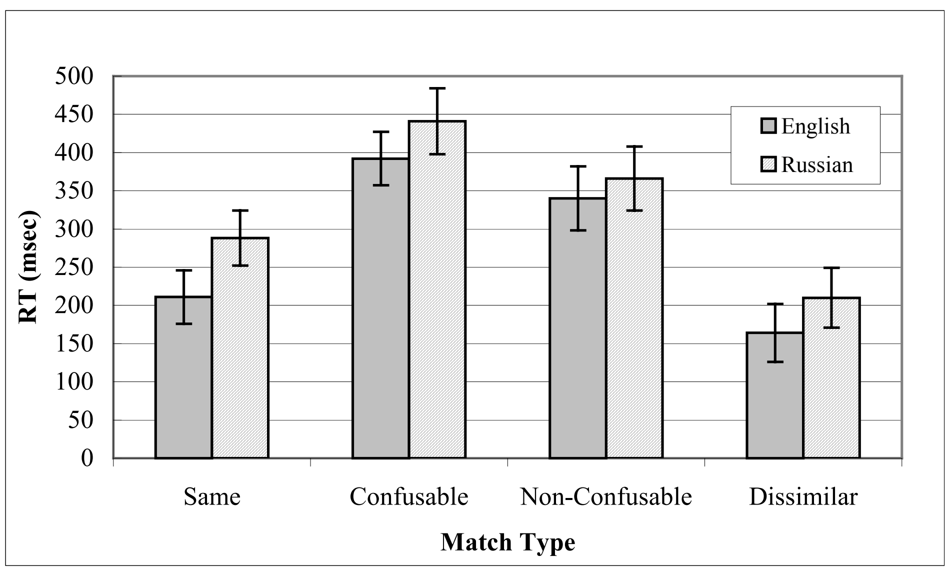


Figure 3.3: RT results for the Cross-Modal Matching task. Error bars indicate 95% confidence intervals. (a: Mandarin Variation; b: Russian Variation).

Of interest is the difference between the Confusable and Non-Confusable conditions. If the phonetic pairs that are predicted to be confusable are the locus of comprehensibility differences, RT in the Confusable condition should be noticeably slower than RT in the Non-Confusable for the accented talkers, but no different for the non-accented talkers. Even if there is a difference between the Confusable and Non-Confusable responses, the degree of difference should be greater for accented talkers. These differences were calculated for each of the linguistic backgrounds used (figure 3.4). These differences were not reliable in either variation (Mandarin: $t(47) = .290$, $p = .773$; Russian: $t(53) = -.955$, $p = .344$). The difference between the Confusable and Non-Confusable was no larger for accented stimuli than for non-accented, although the trend in the Russian variation was in the predicted direction.

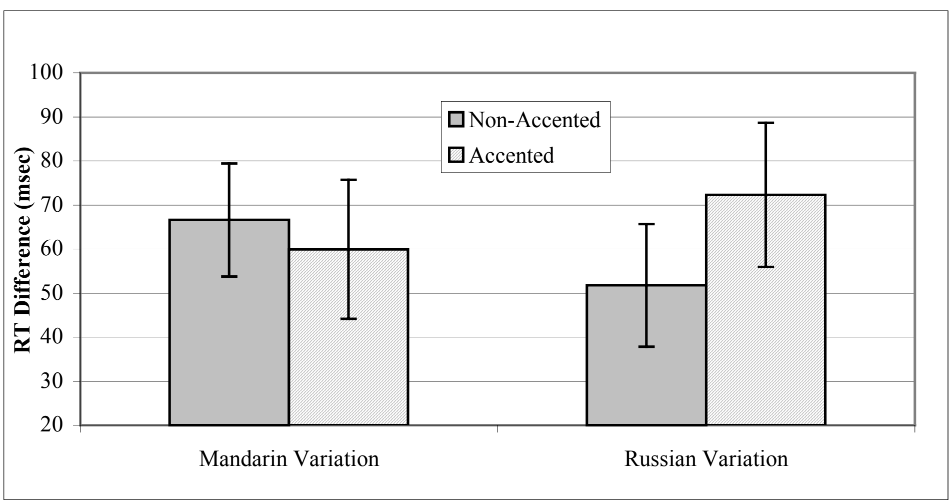


Figure 3.4: RT difference scores: (Non-Confusable Accuracy) – (Confusable Accuracy). Error bars indicate standard error.

It was expected that both Prime condition and talker accent would impact RT. As expected, there was a significant main effect of Prime condition, which indicates that RT is sensitive to the gross phonetic similarity between the Prime and Target in the same way that the accuracy data is; as the gross phonetic difference between the Prime and Target increases, the RT decreases. However, RT does not seem to reflect the finer grade accent differences that were found in the accuracy results. In the Mandarin variation, there was no difference in RT in response to the accented and non-accented words. In the Russian variation, a significant main effect of Accent was found; accented words took longer to respond to than non-accented words. But the pattern of interaction results do not follow the expected patterns. RT to the Russian stimuli should be disproportionately slow in the Confusable condition, when compared with the other Prime types However, the Confusable and Non-Confusable phonetic changes were equally disruptive to RT. There should not have been accent differences in the Dissimilar condition, yet this was found in the Russian variation. The causes of the main effect of accent in the Russian variation may be due to non-phonetic factors.

Conclusions

The results of Experiment 1 indicate that FAS lowers accuracy in a phonetic discrimination task. More discrimination errors occurred when the stimuli was accented than when it was not. This replicates the results of previous literature, which found lower intelligibility for accented phonemes, words, and sentences when compared to non-accented controls. The predicted perceptual dissimilarity of various phonetic contrasts was established by consulting existing characterizations of accents (i.e., Rogers 1997;

Blumenfeld 2002), and the impact of these specific contrasts on discriminability assessed. The disparity in accuracy was greater for phonemes predicted to be confusable, given a specific L1/L2 pairing. These results are consistent with the implications of the SLM (Flege, 1995), which posited graded accentedness as a function of the L1 and L2 phonetic inventories.

In both the Mandarin and Russian variations of the Cross-Modal Matching task, RT slowed as a function of prime condition. The RT was greater in the Confusable condition compared to the Non-Confusable, and indication that RT is sensitive to gradations of gross phonetic similarity. However, the impact of FAS on comprehensibility is still unclear. This increase in RT among the Prime conditions was no greater for accented stimuli than for non-accented. The lack of a consistent main effect of accent on RT contradicts the prediction that FAS lowers RT and thus negatively impacts comprehensibility. Still, the findings of Munro and Derwing (1995b) and Schmid and Yeni-Komshian (1999) provide clear indications that FAS does lower RT in some conditions. Why, then, were these results not found in Experiment 1?

One explanation for the lack of the predicted pattern of results is that accented speech does not lead to lowered comprehensibility compared to non-accented controls. While RT does change as a function of global phonetic similarity (i.e., Prime type), the more subtle variation due to accent does not effect discrimination time. This would imply that the speech perception process occurs at the same speed, regardless of phonetic variability. Differences in intelligibility may arise because there is insufficient time or information to make the correct responses.

Alternatively, the problem may lie in the Cross-Modal Matching task itself. Variability due to accent may influence response time in speech communication, but this task may not have been sensitive to those changes. To complete the discrimination task successfully, the listener did not necessarily need to know what word he was hearing; he only needed to know if what he heard was different from what he saw. The phonetic differences between a native and non-native utterances may not have been great enough to exhibit differences in this task. What is needed to overcome this confound is a task that requires the listener to access the cognitive representation of the intended word before making the response. A different paradigm was used in Experiment 2, spoken word repetition, to test this possibility.

CHAPTER 4

EXPERIMENT 2 – WORD REPETITION

A word repetition task was used in Experiment 2 in order to measure word comprehensibility with different task requirements than were needed to complete the Cross-Modal Matching task used in Experiment 1. In this word repetition task, listeners hear utterances over headphones and are instructed to repeat what they hear out loud, into a microphone. Using recording equipment and voice activated timers, both the accuracy of their responses (intelligibility) and the time needed to make the response (comprehensibility) are measured. Repetitions errors are then analyzed to assess the predictions of SLM. Like the Cross-Modal Matching task used in Experiment 1, the word repetition task can measure intelligibility and comprehensibility simultaneously, and the same recorded utterances can be used between experiments, allowing cross-experiment comparisons. However, the Word Repetition task has a simpler decision criteria and open response set, which will make the RT measure more interpretable.

The goals of Experiment 2 were the same as those of Experiment 1: To evaluate how FAS impacts intelligibility and comprehensibility for non-accented listeners. Accented words should be repeated with lower accuracy than non-accented control words. The types of errors made by subjects should contain the types of phonetic confusions predicted by SLM – those used in the Confusable condition of Experiment 1. If FAS lowers comprehensibility, the speed of the word repetition should be lower for accented words than for non-accented, even when the listener interprets the word as intended.

The freedom of responses in Experiment 2 necessitates that intelligibility predictions be evaluated differently than in Experiment 1. In the Cross-Modal Matching task, the relative intelligibility of different phonetic segments was determined by comparing discriminability among the four Match conditions. Accuracy changed as a function of Match type, and indicated higher confusability for predicted phonetic contrasts. In the Word Repetition task, there is no such comparison because the responses are not primed. Instead, listeners can respond with any utterance. To evaluate phonetic intelligibility in Experiment 2, an analysis of the types and frequency of response errors is required.

Experimental Methods and Procedures

Both Mandarin and Russian variations of a word repetition task were conducted. The words recorded for use in Experiment 1 were used again in Experiment 2, and the procedures were in large part identical to those of Experiment 1. Subjects were run in groups of 1 or 2 people in individual sound attenuated booths. In each trial, a black "+" appeared on the video display for 750 msec, to indicate the beginning of a new trial. There was then a 500 msec pause before the utterance was presented via headphones to the participants at a comfortable listening level. Participants were instructed that all of the utterances they would be hearing were real English words, spoken by a variety of individuals. Responses were to be made as soon as they could identify the word spoken. Speed and accuracy were equally emphasized in the instructions. Subjects then reported what they had heard by repeating the word into a Electro-Voice PL91 Dynamic Centroid microphone. The responses were recorded using a Tascam 102mkII analog audio tape recorder and normal bias cassettes for analyses, and the response time recorded to a computer data file. A 500 msec pause followed each trial. There was one break at the mid-point of the experiment. A post-experiment questionnaire was completed by each subject.

Sixteen subjects participated in the Mandarin variation, and sixteen in the Russian. The two variations consisted of different subjects, and none of the subjects had participated in Experiment 1. All participants were Ohio State University undergraduates who received course credit for their involvement. None reported fluency in any language but English, nor any speech or hearing difficulties. The participants and their parents were all native English talkers, raised in the United States.

Results and Discussion

For each trial, two responses were collected: repeated utterance and RT. Correctness of the repeated utterance was determined by comparing the intended word with the recorded utterance of the participant. Two fluent English talkers listened and transcribed the subject responses. The transcribers agreed on the response for 95% of the trials in the Mandarin variation and 93% in Russian. Trials for which there was not agreement were not considered in subsequent analyses.

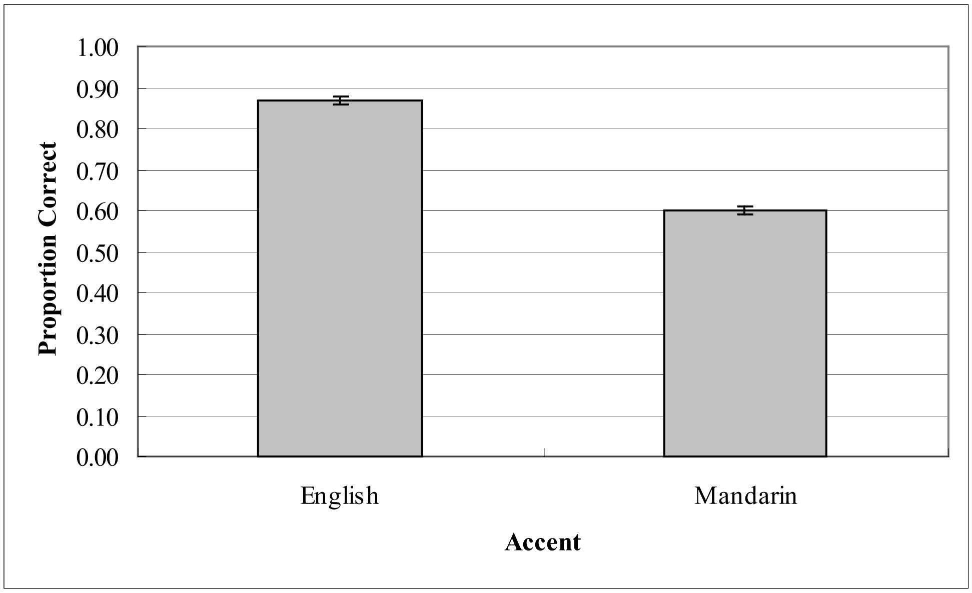
Intelligibility.

Paired-samples *t*-tests were conducted for each variation to determine if there were differences in word repetition accuracy. In the Mandarin variation, this mean difference was significant, $t(15) = 30.77$, $p < .001$. Participants correctly repeated 87% of non-accented words compared to 60% of the Mandarin accented words (figure 4.1a). In the Russian variation, the means were also significantly different, $t(15) = 18.481$, $p < .001$. Participants were able to correctly repeat 79% of non-accented utterances compared to 52% of Russian accented utterances (figure 4.1b).

It is clear given this consistent pattern of results that accented speech is less intelligible than NAS when single words are presented to native speakers. Word identification is 25%-30% lower for accented words compared to non-accented words. This replicates both the results of Experiment 1, and those of many of the studies in the literature (Rogers, 1997; Munro and Derwing, 1995a; Schmid and Yeni-Komshian, 1999). The presence of an accent increases the chances that a word will be misidentified.

To assess the causes of the intelligibility differences, a broad phonetic analyses of incorrect responses was conducted on participants' recorded responses. Each participant's utterance was compared to the intended word, and differences at the phonetic level were categorized. For instance, if the intended word was "lab" (/læb/) and the participant repeated "lap," one error would be recorded: /b/ to /p/. If the participant repeated "lip," an additional error would be recorded: /æ/ to /ɪ/. These errors were tallied for all incorrect trials for all participants, and categorized by phoneme confusion. The types of errors made were compared with the types of errors predicted by Rogers' (1997)

a



b

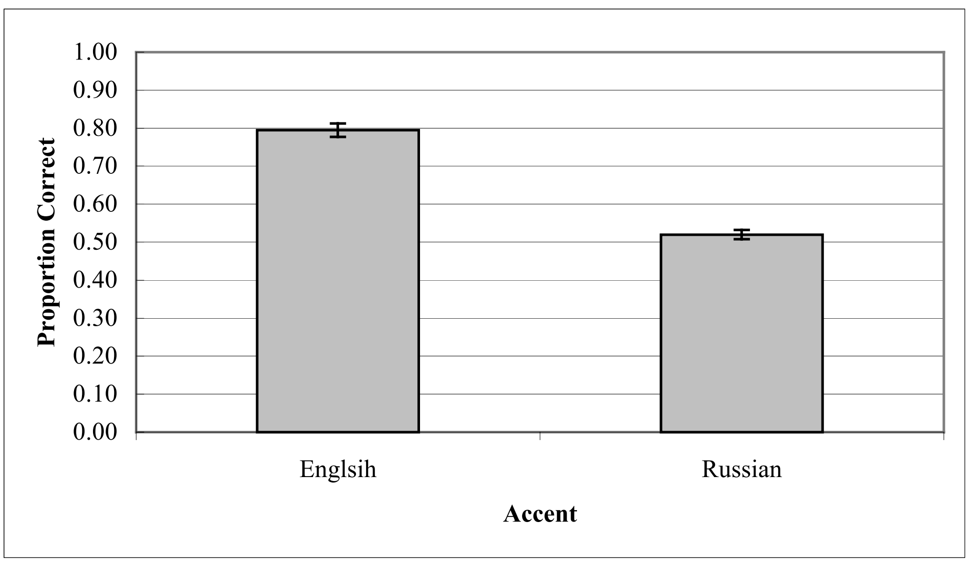


Figure 4.1: Accuracy results in word repetition task. Error bars indicate 95% confidence intervals. (a: Mandarin Variation; b: Russian Variation).

analysis of Mandarin accented speech, and the characterizations of Hermann & Herman (1943) and Blumenfeld (2002) for the Russian accent. More detailed descriptions of these predictions are given in Chapter 2.

Because both Flege (1995) and Rogers (1997) indicated that accent errors would be positionally dependent, errors were calculated separately for word initial, word medial, and word final phonemes. In the Mandarin variation, 132 types of errors were made at least one time, with 562 total errors. Of these, 14% of the errors occurred in response to non-accented talkers, 86% in response to Mandarin talkers. The errors predicted by Rogers (1997) and described in Chapter 2 accounted for 36.3% of the total repetition errors made. The breakdown by phonation type and position is given in table 4.1. There were 150 types of errors in the Russian variation, 656 total errors. Twenty-four percent were in response to non-accented stimuli, and 76% for Russian accented stimuli. Predicted Russian contrasts accounted for 30.2% of the total errors.

Phonation Type:	**Consonant**		**Vowel**
Position in Word:	**Initial**	**Final**	**Medial**
Mandarin Accented Words	31%	56%	17%
Russian Accented Words	0%	30%	44%

Table 4.1: Percent of repetition errors accounted for by the predicted confusable phones.

The predicted phonetic confusions accounted for a sizable proportion of errors. Remarkably, the Rogers' predictions accounted for 56% of the final consonant errors in response to Mandarin speech, and the Russian vowel characteristics described by Blumenfeld (2002) accounted for 44% of errors.

Note that very few predictions were made in the Russian initial position, so the proportion of errors accounted for is very low. Only 1 of the 122 errors made on word initial phonemes was predicted. The majority (99.2%) of word initial errors made in response to Russian talkers was not predicted by the characterizations of accents used to create the stimuli.

Participants reported hearing a variety of words for a given stimuli. For example, the word "raid" was spoken by a Russian talker in Experiment 2. It was predicted that the /d/ might be confused with a /t/ given the characteristics noted for the Russian accent; that "raid" would be heard as "rate." It turns out that "raid" was repeated as "pride," "prayed," "cried," "fight," "right," and the non-words "pred" and "krid"[§] – but never "rate." The expected error did occur – two participants reported hearing /t/ rather than /d/. But many other errors were reported as well. These responses, all incorrect, contained a mix of the intended phonemes and errors, although each phone was interpreted correctly by at least one listener.

Rogers argues that this type of error assessment is not suitable to evaluate phonetic confusions because listeners have a strong bias to respond with a real word, rather than reporting the actual sounds that they heard. For example, if an accented talker's pronunciation of "phone" (/fon/) was pronounced "foat" (/fot/), non-accented listeners might interpret the utterance as "foot" (/fUt/) because "foat" is not a real word. The experimenter would then assume that two errors in perception occurred (/n/ → /t/, /o/

[§] These non-word responses were made despite instructions that all stimuli were real words.

→ /U/), when there was actually only one. This may have been the case in Experiment 2; a bias to respond with a real word – as stipulated in the instructions – may have artificially inflated the number and types of errors reported.

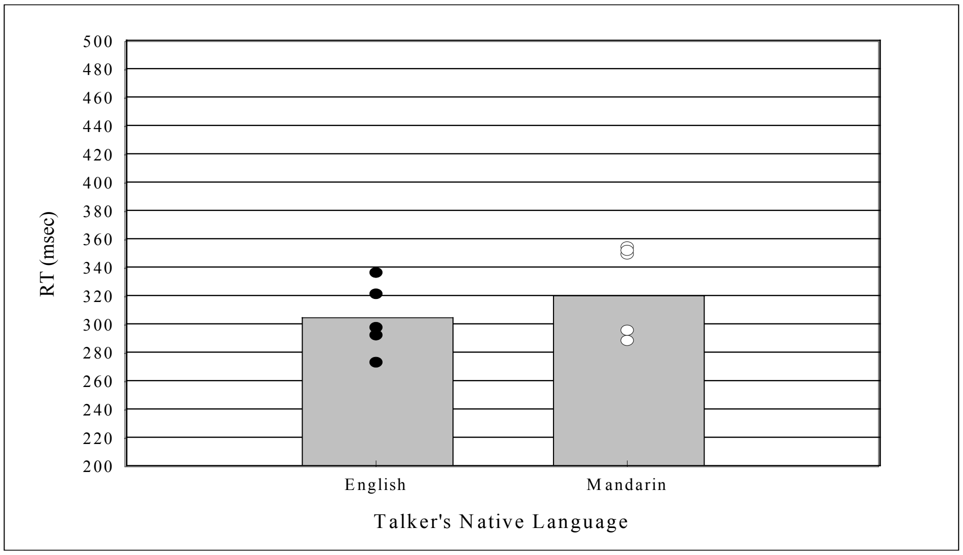
This bias makes it difficult to assert that the predicted phonetic confusions were more or less prevalent than other errors in Experiment 2. It is unclear which errors were due to bottom-up perceptual confusions, and which were due to top-down compensatory mechanisms. The results of Experiment 1 unambiguously show that the predicted confusions are more problematic in FAS perception than the alternatives. However, the responses made in Experiment 2 may more accurately reflect the breadth of errors that are found outside the controlled laboratory setting.

Comprehensibility.

Reaction time differences in Experiment 2 were compared by paired-samples *t*-tests. Only responses judged to be correct were used in RT analyses. RT was measured from the end of the word to the beginning of the spoken response. In the Mandarin variation, this difference was of marginal significance, $t(15) = -1.95$, $p = .07$. Correct responses to Mandarin accented words averaged 320 msec compared to 304 msec for the NAS (figure 4.2a). The Russian variation showed similar trends in RT (figure 4.2b). Repetition of Russian accented words was 321 msec after the end of the word, compared to 262 msec for the non-accented trials ($t[15] = -9.77$, $p < .001$).

The reason that the RT difference for the Mandarin was not significant can be seen when the Mandarin data are broken down by talker (figure 4.2a, points). Two distinct groups of Mandarin talkers emerge: one group of three talkers who elicited RTs

a



b

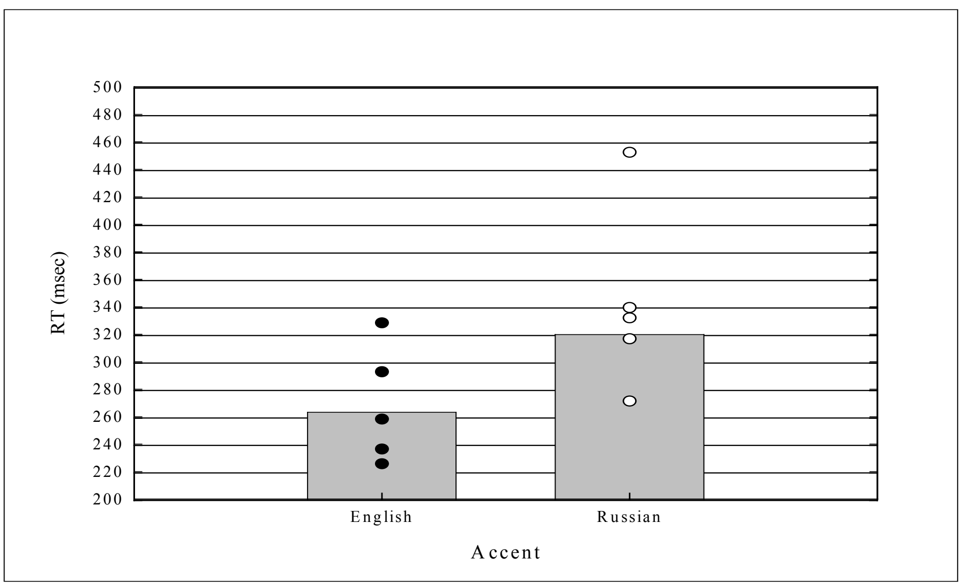


Figure 4.2: RT in the Word Repetition task. Circles indicate performance in response to individual talkers. (a: Mandarin Variation; b: Russian Variation).

which were slower than the English talkers, and one group of two talkers (Man3 and Man5) who elicited responses which were noticeably faster. These same talkers had the highest intelligibility scores as well as the fastest responses, and had lived in the United States for the longest period of time (table 2.1). Given this information, a second t-test was conducted with these two talkers were eliminated from the analyses, along with the two fastest English talkers (Eng2 and Eng4). With these talkers removed, there was a significant difference in RT as a function of accent, $t(15) = -9.77$, $p < .001$.

These results suggest that FAS decreases comprehensibility. The presence of an accent adds 30-60 msec to the processing time required to identify and repeat a word. This result was consistent between the variations; both Russian and Mandarin talkers were slower than their respective English controls.

Conclusions

The presence of a foreign accent lowers intelligibility. When asked to identify isolated words spoken with or without an accent, listeners are more likely to incorrectly identify the accented word. This replicates the results of Experiment 1, where discrimination was found to be lower for accented utterances. Examining the responses of the native English subjects revealed a diversity of perceptual errors. Although the phonetic confusions predicted by the descriptions of Mandarin accents (Rogers, 1997) and Russian accents (Herman & Herman, 1943; Blumenfeld, 2002) appeared in great numbers, many errors that were not predicted to occur in fact occurred.

When talkers correctly identify accented words, they are relatively slow to do so. Responses to accented words were ~50ms slower than to non-accented words, for both

Russian and Mandarin accented words. This differs from the results of Experiment 1, where consistent main effects of accent were not found. Because the same stimuli were used in both tasks, perhaps the demands of each task led to the differences in response.

All of the words used as stimuli in this series of experiments included at least one phoneme that was expected to be potentially confusable with another English phoneme, given the linguistic background of the talker. As a consequence, it was not possible to compare responses to words that contained confusable phonemes to those that did not. This would have been beneficial to determine the relative contribution of confusable phonemes to the overall comprehensibility. This evaluation is possible with a modified stimulus list. Words spoken by non-native talkers which do not contain confusable phones would be an appropriate contrast to those words which contain 1 or more confusable phonetic segments. Comparing response times to these words with those that contain one or more confusable phones may establish the causes of comprehensibility differences.

The RT measure was used to determine if there is a processing cost involved in successfully identifying accented speech. The SLM posits that certain linguistic segments will deviate from the native pronunciation norm. That a word is intelligible does not necessarily imply that it was pronounced normally – only that the intended message was recovered. The longer RT values found in Experiment 2 imply that the cognitive processes involved in the processing of accented speech are more labored. The word repetition design was used to gauge the effectiveness of RT as a measure of comprehensibility or effort. Previous studies that have tried to measure perceptual effort in FAS have focused on larger linguistic units (i.e., sentences) or have used more subjective measures. Experiment 3 addresses the relationship between the RT values found in Experiment 2 with subjective ratings of comprehensibility.

CHAPTER 5

EXPERIMENTS 3 - COMPREHENSIBILITY JUDGMENT

The results of Experiment 2 support the conclusion that FAS negatively impacts comprehensibility. Repetitions of accented words took longer than non-accented words. This conclusion is in-line with predictions drawn from Flege's (1995) SLM, which imply that segments contained within accented speech will differ perceptually from normal pronunciation. Because these spoken segments are imperfect matches with the native speaker's cognitive representations, it is more difficult to successfully identify what was spoken.

Does this RT cost mirror our impressions of accented speech? Previous studies that have investigated comprehensibility have relied on subjective ratings tasks. Munro and Derwing (1994a) presented accented phrases and sentences to native English listeners. Participants were asked to transcribe the utterances and to rate the effort needed to understand the phrase (Comprehenbility). They used a 9 point scale, 1 being "extremely easy to understand" and 9 being "impossible to understand." Participants made 63.6 transcription errors per accented talker, but more than 80% of the sentences were transcribed with high accuracy (above 85% correct). Comprehensibility ratings were skewed towards the "easy to understand" range (figure 5.1), but there were far more ratings of "moderately difficult to understand" than would have been expected given the high transcription accuracy.

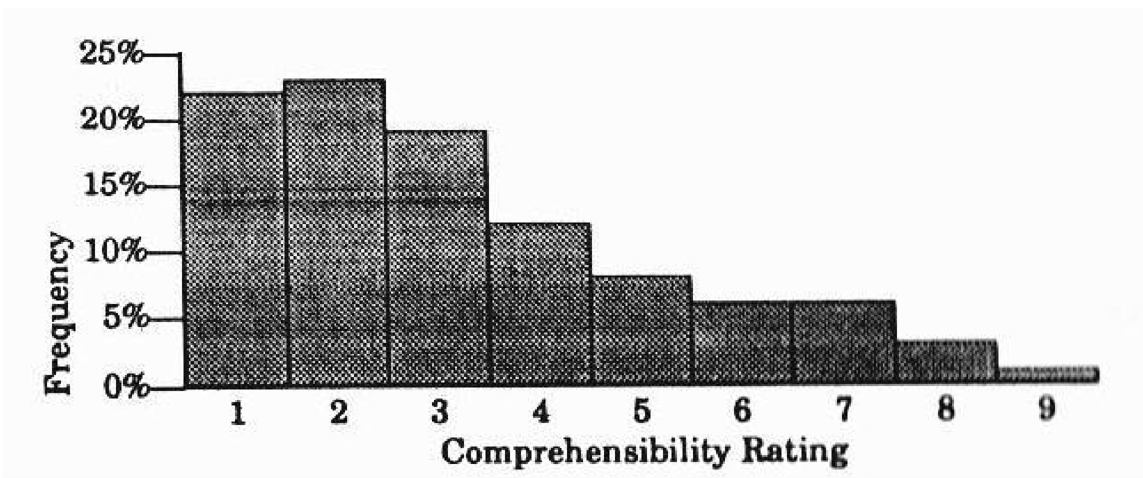


Figure 5.1: Comprehensibility Ratings (from Munro and Derwing, 1995a).

Do these ratings reflect a perceptual cost attributable to accent? The discrepancy between intelligibility and judgments of comprehensibility is intriguing; if a listener is able to understand most or all of what he hears, why would he then say that the utterance was difficult to understand? One explanation is that certain words are difficult to correctly identify for a given talker or a given accent because of the way in which accent expresses itself in L2 production. It takes longer because the component phones are produced non-normally, and the lexical access time is relatively laborious.

In Experiment 2, response time was interpreted as a measure of cognitive effort. Similarly, the rating scales of Munro and Derwing (1995a) purport to measure effort. If they are both measuring the same property of perception, the two measures should correspond. Words which take longer to identify should also have lower comprehensibility ratings, and the reverse should be true as well. To test this, Experiment 3 was conducted. Echoing the design of Munro and Derwing (1995a), the effort needed for intelligibility was evaluated on a rating scale for Mandarin and Russian accented words and non-accented controls. Significant differences in rating values should be found between accented and non-accented words. This would replicate the pattern of judgments found by Munro and Derwing (1995a). Mandarin and Russian accented words should have high ratings for both effort and accentedness. Non-accented words should have low ratings for effort and accentedness. If reaction time and subjective judgment are tapping into the same processes, a significant correlation should be found between performance in the word repetition task and ratings of effort in Experiment 3.

Experimental Methods and Procedures

Fourteen Ohio State University undergraduates participated in the Mandarin variation, and 14 in the Russian. All received course credit for their participation. None reported any speech or hearing deficiency. The methods used were similar to those used in Experiments 1 and 2, with the following changes.

Participants were run in groups of 1 to 4 individuals, seated in separate sound attenuated booths. A visual signal would appear on the monitor to signal the beginning of a new trial. This was followed by the auditory presentation of the word. Subjects were asked to judge the difficulty of perception on a scale from 1 to 6, 1 being "Very Easy to Understand" and 6 being "Very Difficult to Understand," by pressing a marked button[**]. After they made their decision, the intended word was presented to the visual display. Participants were instructed to indicate if the word on the screen matched the word they believed they heard by pressing buttons labeled Same and Different. This second response was collected to assure that the ratings were being made for the intended words, and not for an alternate word with a similar phonetic make-up.

Results and Discussion

In the Mandarin variation, subjects reported correctly understanding 83% of the utterances – 91% of the non-accented words, and 71% of the Mandarin accented words. This is comparable to the accuracy ratings in Experiment 2, where 87% of the non-accented words were repeated correctly, compared to 60% of the accented words. Only words that were reported as identified correctly were used in subsequent analyses. Paired-samples *t*-tests were conducted for each variation, comparing the mean rating for the accented and non-accented groups. There was a significant difference between the ratings of the two groups, $t(13) = -10.116$, $p<.001$. The mean rating for the Mandarin talkers was 2.26 compared to 1.53 for the English talkers. There was a significant

[**] Please bear in mind that lower comprehensibility is indicated by higher ratings on the subjective scale (i.e., very difficult to understand); high comprehensibility is indicated by low ratings (i.e., very easy to understand).

correlation in rating level for each accent, $r = .892$; subjects who rated the non-accented talkers high tended to rate the accented talkers high as well.

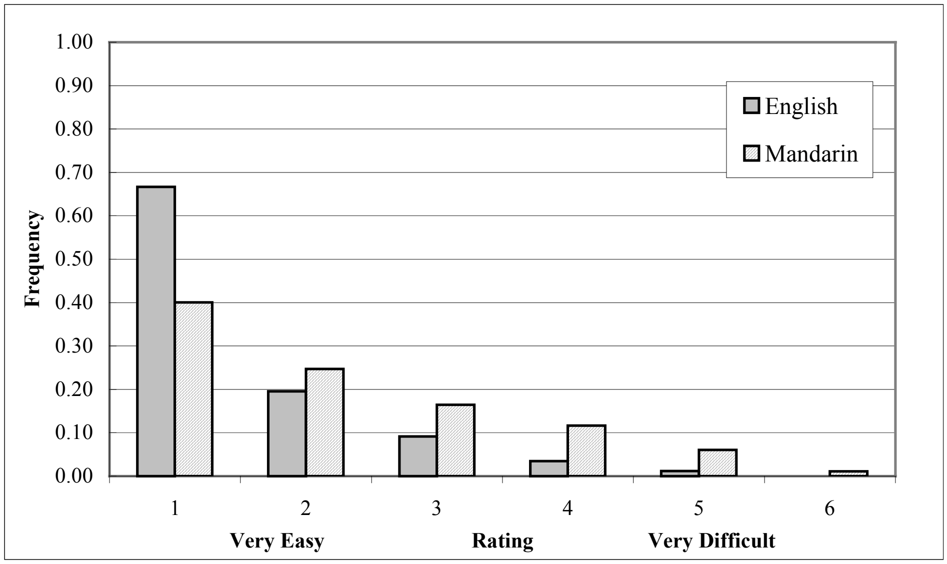
The Russian variation yielded similar results. Subjects reported understanding 80% of the words presented; 91% of the non-accented words, and 69% of the Russian words (compared to 79% for the non-accented and 52% for the accented in Experiment 2) The Russian were rated as significantly more difficult to understand (mean rating: 2.4) than the non-accented talkers (mean rating: 1.75), $t(13) = -9.449$, $p < .001$.

Figures 5.2a and 5.2b show the distribution of ratings for words that subjects reported they heard as intended. The majority of responses – for both accented and non-accented words – was a “1” or “Very Easy” response. For non-accented, the number of responses descended quickly as the rating increased. For accented words, this curve was shallower, with fewer ratings of “1.” These results are very similar the pattern of ratings found by Munro and Derwing (1995a; figure 5.1): a skewed distribution of ratings towards the “very easy” edge of the scale. This similarity implies that participants were using the same criteria to make their ratings in both the current and previous studies, even though the type of stimulus presented in each was different. This validates the use of the rating task for single word presentation.

Cross-Experiment Comparison

The primary motivation for conducting this rating experiment was to correlate performance between subjective ratings and reaction time. In the current investigation, the comparison was possible because the same words spoken by the same talkers were presented in each of the experiments.

a



b

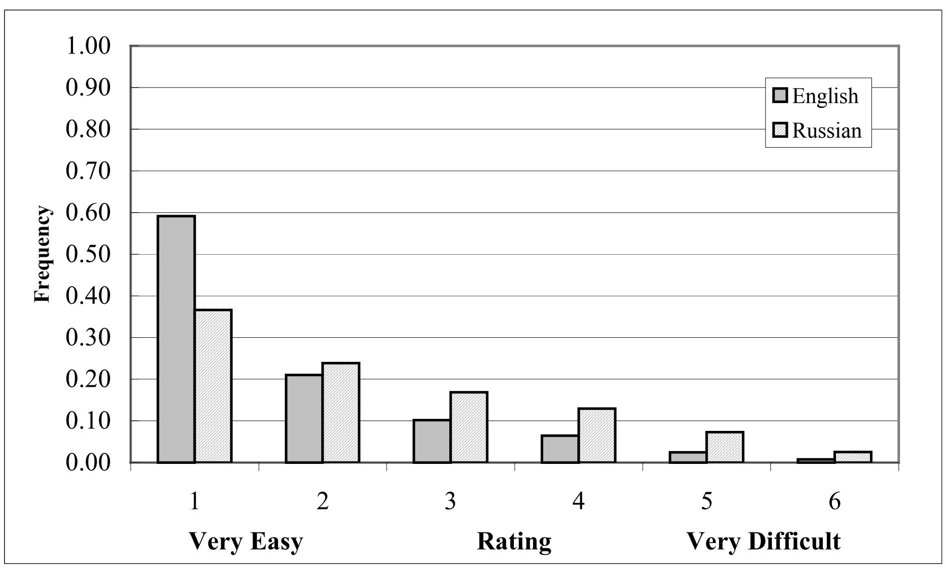


Figure 5.2: Distribution of Comprehensibility Ratings reported correct. (a: Mandarin Variation; b: Russian Variation).

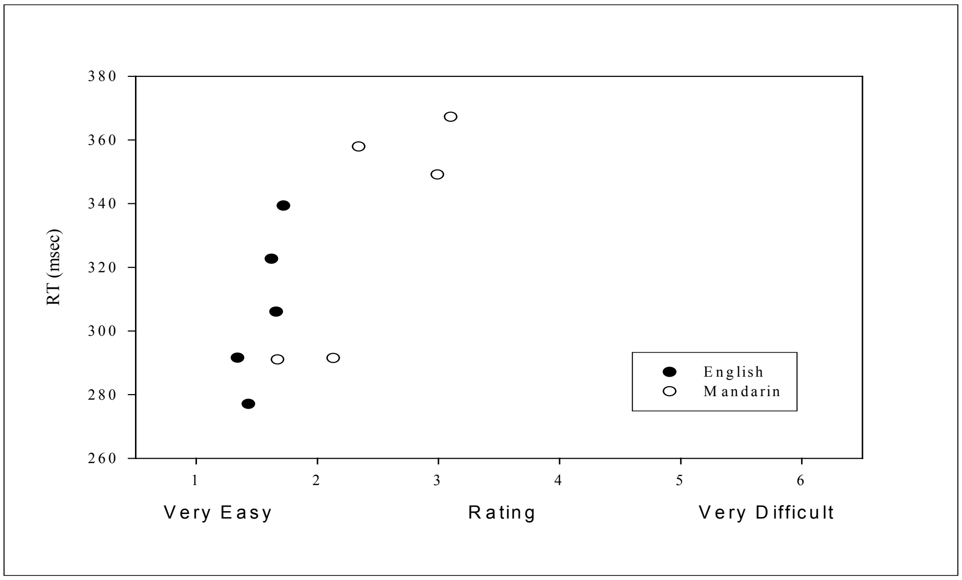
The comprehensibility ratings (Experiment 3) and the repetition RT (Experiment 2) were collapsed over talkers (180 stimuli/10 talkers = 18 trials), and these scores correlated. These data are presented in figure 5.3a. These correlations were significant (r=.774 for the Mandarin list, r= .90 for the Russian List). The speech of talkers who received low comprehensibility ratings tended to take longer to perceive correctly.

It was expected that there would be a certain degree of variability in responses to each talker's stimuli. Although a particular talker may be easy to understand in general, some of his utterances may nonetheless be difficult to understand. Examining the correlation between RT and rating at a word by word level gives a more complete picture of the relationship between these two measures of comprehensibility. Repetition RT and comprehensibility ratings were significantly correlated, r=.525. These results are presented in figure 5.4a. A similar pattern of results was found for the Russian list, r=.50. These correlations are presented in figure 5.4b. In general, words that elicited longer RT values had comprehensibility ratings which were farther towards the "Difficult to Understand" end of the scale.

Conclusions

These results support the hypothesis that subjective measures of perceptual effort and objective measures of on-line speech perception such as RT measure the same underlying processes. The results found by Munro and Derwing (1995a) were replicated using single words, indicating that the same criteria are used when listeners are presented a with single word instead of a larger utterance. Words that were rated as having low comprehensibility tended to require more processing time. Although causality can not be

a



b

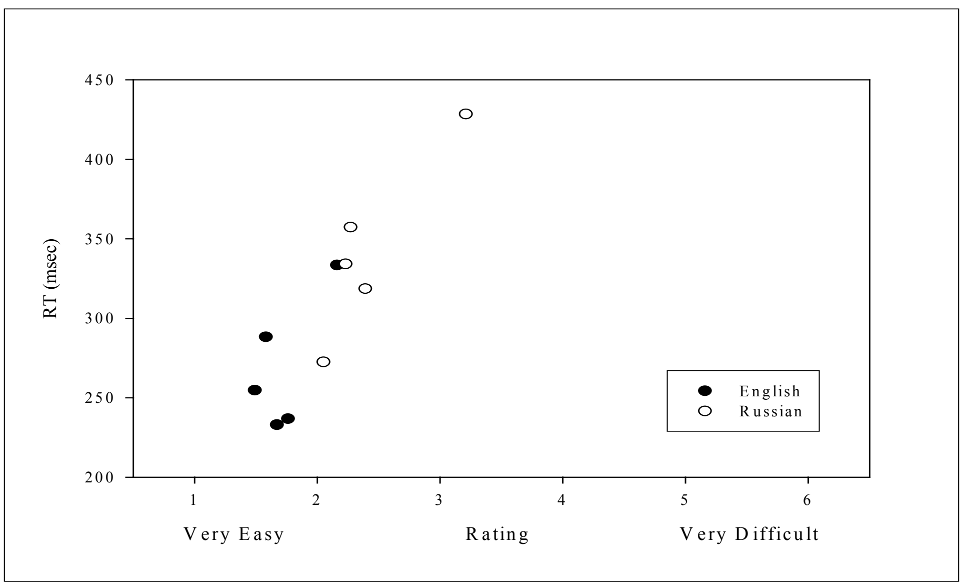
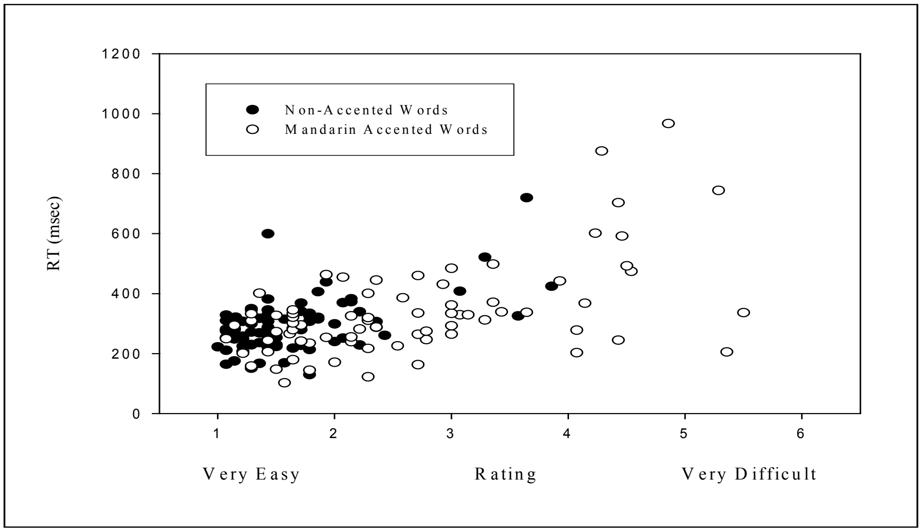


Figure 5.3: Correlation between comprehensibility ratings and repetition RT by talker. (a: Mandarin Variation; b: Russian Variation).

a



b

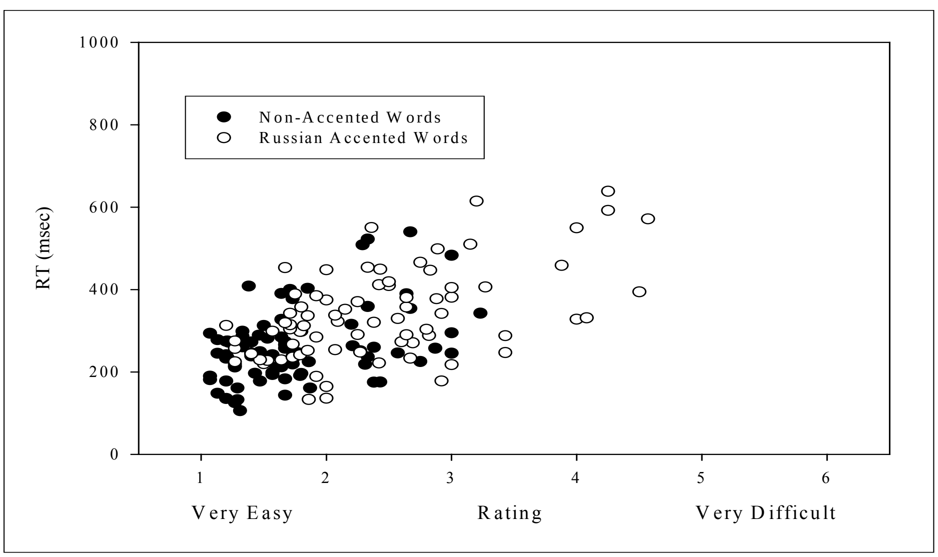


Figure 5.4: Correlation between comprehensibility ratings and repetition. (a: Mandarin Variation; b: Russian Variation).

attributed based on these results, it is a straightforward conclusion the accent-based RT differences reflect the increased processing demands of non-normal pronunciation, and that subjective judgments of effort are based, in part, on the magnitude of this additional cost.

CHAPTER 6

GENERAL DISCUSSION

Studying foreign accented speech perception provides the rare opportunity to simultaneously address both theoretical questions about speech perception and practical questions about the impact of accent on human performance. This series of experiments represents an attempt to do just this. At the outset of this paper, two research goals were outlined. Both concerned the impact of foreign accent on the process and success of speech perception. First, how intelligible is FAS to native listeners, and can the phonetic causes of intelligibility deficiencies be determined a priori? Second, does the presence of an accent require more effort for effective communication than NAS? Three experiments were conducted to explore the intelligibility and comprehensibility of FAS.

The Impact of FAS on Speech Perception

Speech is said to be intelligible if words are interpreted as intended by a speaker. While native speakers of a given language have a remarkable tolerance for variability in the speech signal, the characteristics of foreign accented speech push the boundaries of the ability to compensate for differences in speech production within one's native language. The results of Experiments 1 and 2 demonstrate that accented speech has consistently lower intelligibility than non-accented speech. Listeners are more likely to confuse similar sounding words when uttered with an accent. Participants in Experiment 1 were presented with a discrimination task; they saw a visually presented word and had to determine if it was the same as or different from a word presented aurally. In essence, listeners were required to call to mind the visually presented word and compare their memory of the sound of that word with that of the spoken target presented. When these words differed by only a single phoneme, accented words were more difficult to discriminate than non-accented words (73% correct versus 83% correct, averaged over the Russian and Mandarin variations). This accuracy gap was largest when the visually and aurally presented words differed by phonemes that were predicted to be difficult to discriminate based on the characteristics of the specific accent presented (59% correct in the Confusable condition versus 78% correct in the Non-Confusable).

Similarly, listeners are more likely to misunderstand a word when it is spoken by a non-native talker than when it is spoken by a native talker. In Experiment 2, participants were presented with single words and instructed to repeat the words they heard. Repetitions of accented words were more likely to contain errors than were similar non-accented words (38% error rate for FAS, 11% error rate for NAS). Many types of repetition errors were made, including those confusions predicted based on the characteristics of the accents.

Taken together, these results indicate that accented words are more difficult to both identify and discriminate than non-accented words. They are less intelligible overall, and are particularly difficult to identify when they include specific phones that are produced non-normally due to the accent characteristics. These phones can be determined by comparing the allophonic inventories of the two languages (e.g., Flege, 1995) and noting differences, or by using descriptions of the phonetic characteristics of the accents themselves (e.g., Rogers, 1997).

Intelligibility is a prerequisite for successful communication. However, identifying and understanding a word is the end product of the speech perception process, it is not the process itself. Two words can both be intelligible, but one may require far more time, cognitive resources, or attentional focus for the same end result. The perceptual difficulty or effort involved in successfully understanding accented speech was tested using two measures, reaction time and subjective ratings. In Experiment 2, the repetition time was calculated and compared for FAS and NAS. Responses to both Mandarin and Russian words were slower than to non-accented control words. When people hear an accented word, it takes them more time to access the cognitive representation of that word. Ratings of the effort needed to understand words correctly were solicited in Experiment 3. Although most accented words were found to be intelligible, these words received lower average comprehensibility ratings than did non-accented equivalents. When the ratings of Experiment 3 were compared with the repetition RT of the same words presented in Experiment 2, significant correlations were found. By and large, words that elicited slower responses tended to have ratings that indicated lower comprehensibility. While it is not clear if participants based their ratings on the increased time needed for lexical access, or if both the ratings and RT data reflect another underlying process, it is clear is that there are processing costs that can be attributed to the degree of variability inherent in FAS, independent of word intelligibility.

Attributing comprehensibility differences to specific accent characteristics proved to be more difficult. The Cross-Modal Matching task used in Experiment 1 was designed to induce perceptual errors and slowed responses by forcing listeners to compare accented utterances with words that had similar phonetic compositions. The response times between the accented and non-accented stimuli differed as a function of the prime/target similarity, but not as a function of the linguistic background of the talker. Responses to Mandarin and Russian stimuli were no faster or slower than to English stimuli. The task demands of the Cross-Modal Matching paradigm may have obscured measurable RT differences.

These results obtained in these three experiments agree with and extend the previous literature on the perception of FAS. Munro and Derwing (1995a), Shmid and Yeni-Komshian (1999), and Rogers (1997) all found that accented speech is less intelligible than NAS, using a variety of designs. The Cross-Modal Matching task was designed to evaluate FAS intelligibility, so the clear main effect of accent bolsters the robustness of the intelligibility gap seen previously. The word repetition task is similar to the transcription task used by Munro and Derwing (1995a) and others in that the impressions of the talkers are recorded. The current results replicates the main effect of previous studies: lower accuracy when a word is accented.

The propositions of Flege's (1995) SLM were used as the theoretical foundation for many of the arguments presented in this project. He described the demographic conditions that would cause phones to be produced non-normally based on the influence of L1 onto L2. The result is a mismatch between the accented talker's realization of a phone and the native listener's cognitive representation. This implies that a comparison of the allophonic inventories of both languages would yield predictions of the phones that lead to substandard intelligibility. The process of comparison is beyond the scope of the present investigation.

One of the primary hypotheses of Rogers (1997) was that phonetic analysis of accented speech can be used to make predictions about the intelligibility of FAS with far less difficulty than the direct phonetic comparison implied by SLM. Through careful scrutiny of accented speech samples, potential phonetic ambiguities can be noted. From these, predictions can be formed as to the likely perceptual pitfalls for listeners. She found, in a minimal pairs test, that native English listeners often confuse accented words with foils that differed by a single predicted phoneme. The Cross-Modal Matching task used in Experiment 1 was designed in part to extend these results by comparing the accented utterance not only with the confusable alternative, but with foils that were not expected to be difficult to discern. Rogers' conclusions were confirmed by this manipulation. Not only were English-speaking participants susceptible to higher error rates for accented talkers, these error rates were highest for those phonetic contrasts that were predicted to be confusable for a given accent. These results were replicated using the Russian accent which had a different set of confusable phones.

The RT and rating data collected in Experiments 2 and 3 extend the previous literature on FAS comprehensibility. Munro and Derwing (1995a) found that ratings of the perceptual effort needed for a sentence to be intelligible were "harsher" than intelligibility measures indicated. Many sentences that were transcribed perfectly had ratings that indicated that the sentences were moderately difficult to understand. They did not consider that the intelligibility of individual words within a sentence will vary, altering the overall impression of intelligibility of the sentence. If a single word in a sentence is not intelligible, the ratings of Munro and Derwing (1995a) will reflect the effort needed to recover that word, even if all of the other words in the sentence were easily interpretable. Perhaps this caused a bias towards ratings of perceptual difficulty, and thus overstated the magnitude of the comprehensibility gap between accented and NAS. This question was addressed in Experiment 3. A rating task similar to that used in Munro and Derwing (1995a) was employed, but with single words as stimuli. A pattern of results similar to that of Munro and Derwing (1995a) was found. Words that were intelligible were nonetheless given ratings that indicate low to moderate comprehensibility.

Implications of Comprehensibility

What are the implications of lower comprehensibility? The results of Experiments 2 and 3 clearly indicate that what people describe as greater perceptual effort correlates significantly with increased reaction time. What has not been investigated are the other potential consequences of this discrepancy in perceptual effort. How else does comprehensibility impact speech perception and task performance in general?

Consider two words, one spoken with a foreign accent, and one by a native English talker. In normal conditions, native listeners correctly identify both words 98% of the time. However, although the accent has not lowered intelligibility, it takes about 40 msec longer for native listeners to identify the accented word. Subjectively, listeners feel the accented word was more difficult to understand. In essence, the accented word is less comprehensible than the non-accented word.

What happens to the intelligibility of these words when the signal is distorted by the addition of noise? Given that the baseline (noiseless) intelligibility of both words is the same, will the noise affect both words equally or will the word with lower comprehensibility have disproportionately lower intelligibility in noise? Several studies have investigated this question, and have found that accented speech is affected by disruption in noise to a greater degree than NAS. Rogers et al (1999) found that transcription accuracy in noise was disproportionately lower for low proficiency accented talkers than for high proficiency accented talkers. That is, the slope of an intelligibility curve is a function of both the amount of noise added to a signal and the severity of the accent. Scott (1999) confirmed these results using a mispronunciation detection task in noise. Noise disrupts accented speech more than it does NAS.

Van Wijngaarden (2001) used a variation of a transcription task (Speech Reception Threshold [SRT], Plomp & Mimpen, 1979) to quantify the effects of noise on FAS perception. The SRT measures a talker's susceptibility to disruption due to noise, which is used as an indirect measure of intelligibility. Native and non-native Dutch talkers were recorded, and their utterances were presented to native and non-native Dutch speaking listeners. In noiseless conditions, both the accented and non-accented talkers had near ceiling performance. In noise, this pattern changed. Van Wijngaarden found that performance was best when both the talker and listener were native speakers of Dutch. If either the talker or listener were not native speakers, the SRT was approximately 3 dB higher; non-native speech perception was far more susceptible to degradation due to noise masking then native speech, indicating lower intelligibility for accented talkers. From these SRT values, psychometric curves can be plotted which show the impact of noise on intelligibility (figure 6.1).

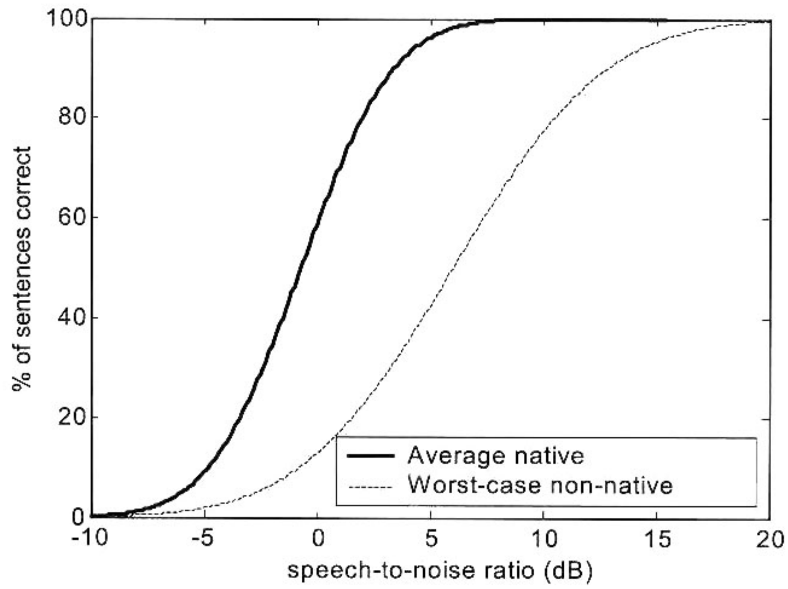


Figure 6.1: Speech intelligibility curve (from Van Wijngaarden, Steeneken, & Houtgast, 2002)

As SNR increases, intelligibility grows as well. However, when the speaker has an accent, this improvement is far more gradual. Although near ceiling intelligibility is reached with high SNR for both FAS and NAS, the difference between the accented and non-accented curves may reflect comprehensibility differences. Recall that similar words with equivalent intelligibility may require different lengths of time to process, all other factors being equal. However, if only accuracy is being recorded, these differences may be overlooked. When noise is added to the signal, the non-accented utterances become somewhat less intelligible, while the accented speech becomes far less intelligible. Van Wijngaarden (2001) did not use the term comprehensibility in describing the phenomena he was studying, but a strong argument could be made that the patterns of results he found were due to differences in perceptual effort.

The relative perceptual similarity of the spoken word and the cognitive representation of that word determines comprehensibility. In clear conditions, the two words from our example are both perceptually similar enough to activate the correct lexical entry, but it takes longer for this to occur for the accented word. When noise is added to the signal, the perceptual difficulty of the lexical activation process increases. The result of this is that the non-accented word may still be intelligible, but response time should be slower. The accented word is less likely to be intelligible because the perceptual distance between the cognitive representation of the word and the degraded, accented spoken word is too large for lexical access to occur. These predictions could be tested experimentally by repeating the Word Repetition task with different signal to noise ratios, and comparing both response time and accuracy.

These cognitive representations are not fixed. Schwab, Nussbaum, & Pisoni (1985) found that with training, the intelligibility of synthetic speech improved. Weil (2000) replicated these results using foreign accented speech stimuli, and found exposure to a single talker improved intelligibility for both that talker and generalized to other talkers with similar linguistic backgrounds. Clarke (2000) did not find generalization within among talkers who share accent characteristics, but Bradlow & Bent (2003) found that the intelligibility of novel accented talkers can be improved by exposure to multiple talkers with similar L2 characteristics. If exposure to FAS improves intelligibility, this may imply that the cognitive representation changes as new, valid exemplars of linguistic segments are heard and identified.

Context and Prosody: Other Aspects of FAS

For the sake of simplicity, the current study has considered only a few aspects of L2 communication. Only the effects of accent based phonetic variation in single words was considered. This was an essential first step in the investigation of comprehensibility; understanding the effects of accent in smaller linguistic units is crucial for interpreting the effects of accent in more complicated situations. Of course, spoken communication is not limited to the exchange of monosyllabic utterances. Prosodic variability and context both influence how FAS is perceived – for good and for bad. This does not negate the importance of the results obtained here. Instead, it adds perspective to the meaning of these results, and exposes directions for further inquiry.

One of the purposes of presenting only single words to listeners was to assure that the RT values reflected processes related to the perception of that specific word. Comprehensibility is lower for accented words than for non-accented in isolation, but it is unclear how the magnitude of this difference would change in connected speech. One possibility is that the negative effects of FAS on processing time would be compounded by the negative effects of FAS on the preceding context. Accent related disruption in this case would be cumulative, making intelligibility of each word in a sentence more difficult than the last.

Connected speech is not equivalent to a series of spoken, isolated words. When words are spoken in normal speech, coarticulatory effects modify the allophones used, and a larger number of phonetic combinations are expressed at word boundaries. For example, in English the consonant combination /zm/ is not found word initially in any real word. However, in the phrase "His mom went to the store," the phonetic combination occurs in the words "his mom." The principals of Flege's (1995) SLM indicate that accent may be expressed differently in different phonetic contexts. To illustrate, an accent may have the characteristic that the nasal consonant /m/ is expressed as /n/ only if it is preceded by a voiced fricative (i.e., /z/ or /v/), otherwise, it is pronounced as in English. It follows that, words in connected speech contain more opportunities for the negative effects of accent on speech perception than words in isolation. In our example, this characteristic of the accent would never be expressed when presenting single words, but would emerge in connected speech because of the overlap among words; our accented talker would pronounce "mom" correctly in isolation, but incorrectly in the context of that sentence.

Also emergent in multi-syllabic words and connected speech are prosodic variations. Prosody refers to the variations of pitch, speaking rate, timing, and loudness that occur in connected speech (Crystal, 2003). Just as the phonetic inventory of L1 influences L2 production, the prosodic characteristics of a talker's native language may have influence in second language production. The consequence of this is a further degradation of L2 intelligibility, perhaps because prosodic cues help with the word segmentation process. The impact of these differences can not be understated. Gass and Veronis (1984) found that the prosodic aspects of FAS are the most salient aspect of production as rated by trained listeners. Studying prosodic elements of accent is difficult to accomplish because it is difficult to manipulate prosodic cues experimentally. In one successful study, Tajima, Port, and Darby (1997) presented subjects with accented and non-accented sentences, and found that accented sentences were less intelligible. Then they resynthesized the stimuli, modifying the accented sentences so the prosody better resembled the native English prosody, and also doing the reverse. Intelligibility increased for the modified Mandarin stimuli, and decreased for the modified English. They demonstrated that intelligibility increases as the prosody becomes more native-like.

Outside of the laboratory, words are not often presented in isolation. In most situations, listeners have some context in which to base their expectations for what is being said. The simplest one of these is the lexical bias discussed in the analysis of errors in Experiment 2. The types and number of errors subjects made in word repetition may have been altered as the best matching word was accessed in the individual's mental lexicon. It is likely that some of these words contained component phonemes that would not have been identifiable in isolation. However, in the process of lexical decision, this variability is compensated for. That is, given the word "bud" with an ambiguous p/b, the subject may have heard "pud," but responded with "bud" because the experiment called for real word responses and "bud" was the closest match. The result of this is that some perceptual errors were not expressed in the intelligibility results.

These same processes may have created the impression of perceptual errors when none actually occurred. If the speaker said "bud" with the same accent, and listener heard "pud" and responded with the word "putt" (another word with phones in common), it would be assumed that two mistakes were made: /p/ for /b/ and /t/ for /d/. The lexical bias in this case inflates the number of supposed perceptual errors because /t/ was not truly confused with /d/. While this makes interpreting the intelligibility data more difficult, two important aspects of this process must be noted. First, this bias may reveal itself in comprehensibility measures. Words in which a clear lexical entry is not activated should take longer to repeat. Second, lexical bias is one compensatory mechanism for overcoming the variability in second language production. The phenomena should be studied in its own right, not dismissed as a confound to the speech perception process.

As the context in which words are presented becomes more complex, it becomes increasingly difficult to assess the frequency of perceptual errors. Both semantic and syntactic context influence the perception of single words. In Experiment 1, perceptual errors were drawn out from the listeners by priming them with the visual display. In sentential contexts, the priming occurs as a natural consequence of the relationship among the words uttered. If the word "nurse" was pronounced "nuz" /n^z/, it may not be interpretable in isolation. However, in the context of the sentence "the doctor and the nurse went to the hospital," this same word may be identified correctly (see Kalikow et al, 1977).

This does not imply that the phonetic components of these words were perceived differently, only that the criteria for activation in the mental lexicon was altered by the environment. The semantic and syntactic context of the sentence impacts intelligibility in ways that oppose the potential cumulative effects of noise discussed at the beginning of this section. In that case, words at the end of the sentence may be less intelligible than words at the beginning. To evaluate these complex relationships, controlled experiments would have to be designed in which both word predictability and accent severity are varied.

Foreign Accented Speech Outside the Laboratory

Rogers (1997) noted that impact of context on perception makes interpretation of transcription errors a poor test of perceptual acuity. There are just too many factors influencing the outcome. Her minimal pairs test and the Cross-Modal Matching task used in Experiment 1 were attempts to avoid these problems, and they were in large part successful. However, Rogers' criticism highlights an important point: the characteristics required for successful laboratory studies do not necessarily reflect real-world conditions. Scientists may study FAS to inform models of L2 acquisition, phonetic category formation, or theories of lexical access, which all require considerable control of the speech context. Outside of the laboratory, a particular industry may need to understand the impact of FAS on aspects of human performance in a number of complex environmental conditions, indifferent to the underlying perceptual causes. These reasons have some overlap, but the types of data that are informative may differ. As a consequence, the approach taken to a specific problem should vary with the characteristics of the problem.

Physicians and hospital workers must keep meticulous records of the status and treatment progression of patients. Because there is a high patient turnaround, a single doctor will see many patients, and a single patient may see a number of physicians, nurses, and other health care professionals over the course of treatment. To assure that all parties involved know that has been done in the past, records of each procedure, test, diagnosis, or other event is kept in a personal record. However, because of the number of patients a physician must attend to, there is insufficient time for a detailed written report. Instead, the report may be dictated for transcription by an outside party.

It is in this situation that research onto the perception of FAS may be important for patient safety. Physicians in the United States come from many language backgrounds, and will speak English with different degrees of proficiency. The lexicon of the medical field contains many terms that are unfamiliar to the layperson, and which differ only slightly in terms of the phonetic make-up, but drastically in terms of medical course of action. The phonetic similarity of the Latin and Greek roots used for many of these terms may be the linguistic culprit (i.e., hypo- vs hyper-, endo- vs ecto-), but the end result is an increase in potential confusion and catastrophic error. If the individual transcribing the utterance is unfamiliar with these terms, the number of transcription errors may increase disproportionately. This is an instance where the minimal pairs approach may be of some benefit. If two terms that differ by a single phoneme are both semantically valid in a given context, but will lead to disparate treatments, it is imperative to know if some accents are more prone to L2 speech that leads to this ambiguity.

In air traffic control, it is essential that communication be clear between the pilots and the ground for the safety of the thousands of passengers traveling each day. In a busy international airport, pilots will represent many nationalities and language backgrounds. However, in the US and often abroad, all communication is conducted in English. It is inevitable that many pilots will have foreign accents. Controllers on the ground do their work in high demand environments. In addition to communicating with several aircraft simultaneously, they are also performing other tasks. How does the FAS of the pilot impact the performance of the controller? This question is impossible to address without examining the context in which the communication occurs in its entirety. You have pilots of many language backgrounds speaking in a variety of noisy conditions (i.e., cockpit noise), speaking over band-limited radio to listeners who are simultaneously performing other tasks. In this context, both intelligibility and comprehensibility differences may be reflected in poor operator performance. However, the study of intelligibility in isolation may not generalize to actual Air Traffic Control conditions. Although the results of an experiment that presents accented and non-accented versions of aviation terms to controllers in isolation may provide some insight into the types of problems to look for, these results may be tempered when all of the factors are compounded. An alternate approach is to examine performance in the field in a variety of settings that differ in the amount and types of accents that are encountered, and compare differences among them.

If the full extent of the impact of foreign accented speech onto perception is to be understood, a two pronged approach is best. Controlled laboratory studies, such as those presented in this project, begin from the elemental approach. As more and more aspects are studied in isolation, the bigger picture should start to emerge. Contextual, prosodic, and experiential factors should be studied to evaluate the ways in which each impacts processing. More contextually dependent inquiries in specific contexts require a less reductionist approach, where the impact of accent is evaluated in relation to other co-occurring events. When accent is studied in several of these environments, repeating patterns of behavior will emerge, indicating the overall impact of accent. When both the laboratory and situational studies are combined, the result will be a richer understanding of the effects of foreign accented speech on perception and subsequent performance. This will be to the benefit of many domains.

APPENDIX A

LIST M - MANDARIN STIMULI

Listed here are the words used in the Mandarin variations of Experiments 1, 2, and 3. In each experiment, listeners heard only those words in the Target column.

In Experiment 1, participants were also presented with four types of Visual Primes. Same Primes were identical to the words used in the Target column. Confusable and Non-Confusable primes differed from the Target by a single phoneme. Dissimilar primes had no phonemes in common with the Target. In column 'C/V', the phonation type of the change is indicated by either C (consonant) or V (vowel). The position in the word is listed in the last column; consonant changes could be either word initial (I) or word final (F), but vowel changes were always word medial (M). Target and Confusable words were adapted from Rogers (1997)

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
batch	patch	match	sir	C	I
beach	peach	teach	twin	C	I

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
beak	peak	geek	tap	C	I
buys	pies	dies	whip	C	I
cave	gave	pave	bud	C	I
clean	queen	preen	buys	C	I
clear	queer	jeer	fees	C	I
clock	crock	block	put	C	I
coal	goal	bowl	dog	C	I
crack	quack	plaque	mob	C	I
crick	quick	brick	bed	C	I
die	tie	thy	shot	C	I
dig	big	pig	coal	C	I
dime	time	chime	hone	C	I
draw	jaw	claw	glean	C	I
drug	jug	chug	verve	C	I
game	came	dame	sick	C	I
gave	cave	pave	not	C	I
glean	green	clean	duke	C	I
gloom	groom	bloom	breeze	C	I
goal	coal	bowl	rang	C	I
good	could	should	wren	C	I
green	glean	clean	but	C	I
groom	gloom	plume	cave	C	I
herds	hers	hearse	phone	C	I
patch	batch	catch	tune	C	I
peach	beach	teach	lap	C	I
peak	beak	meek	groom	C	I
pies	buys	guys	car	C	I
please	fleas	sleaze	tomb	C	I
they	say	hay	look	C	I
thick	sick	chick	got	C	I
thing	sing	ding	good	C	I
though	so	foe	vein	C	I
tie	die	buy	cards	C	I
time	dime	chime	fail	C	I

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
train	chain	Jane	sag	C	I
trip	chip	drip	clock	C	I
tuck	duck	puck	gown	C	I
vein	wane	sane	wipe	C	I
vet	wet	pet	so	C	I
vine	wine	fine	mouse	C	I
bathe	base	bait	loose	C	F
beads	beats	beach	cut	C	F
beep	beef	beet	soon	C	F
breathe	breeze	breech	tie	C	F
bud	but	bug	feels	C	F
came	cane	cape	dug	C	F
cane	came	cake	pot	C	F
cards	cars	carbs	home	C	F
carve	car	carp	pass	C	F
cords	cores	corks	run	C	F
cores	course	cord	path	C	F
curve	curb	curse	mess	C	F
doom	dune	dude	pond	C	F
duck	dug	dub	loan	C	F
dune	doom	dude	falls	C	F
falls	false	fault	beep	C	F
feed	feet	fees	sense	C	F
feet	feed	fees	sends	C	F
fields	feels	fears	cane	C	F
god	got	gone	ropes	C	F
gong	gone	god	pain	C	F
got	god	gong	cords	C	F
gum	gun	gush	pat	C	F
holds	holes	hulks	gloom	C	F
home	hone	hope	breathe	C	F
hone	home	hole	dude	C	F
lab	lap	lad	gone	C	F
lad	lab	lag	thing	C	F

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
lap	lab	lack	cars	C	F
lib	lip	lit	team	C	F
live	lib	lid	town	C	F
lose	loose	loot	roam	C	F
mob	mop	mock	slide	C	F
mop	mob	mock	jug	C	F
neat	need	kneel	curve	C	F
need	neat	niece	drug	C	F
path	pass	pat	sin	C	F
peas	peace	peach	boat	C	F
ram	ran	rang	sit	C	F
ran	rang	rag	pies	C	F
rang	ran	rack	clone	C	F
robe	rope	rogue	cooed	C	F
robes	ropes	roads	thin	C	F
rode	robe	rose	beak	C	F
rope	robe	rose	say	C	F
rove	robe	rote	dime	C	F
rum	run	rub	holds	C	F
run	rum	rut	big	C	F
rung	run	rub	silt	C	F
sack	sag	sad	need	C	F
sag	sack	sat	ten	C	F
seem	seen	seed	ball	C	F
seen	seem	seat	fool	C	F
sends	sense	sex	late	C	F
serve	sir	surf	hop	C	F
sin	sing	sick	ride	C	F
sing	sin	sip	beads	C	F
some	sun	sub	verb	C	F
sun	some	sud	mop	C	F
tab	tap	tag	peas	C	F
tap	tab	tad	lose	C	F
team	teen	teal	could	C	F

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
teen	team	tease	mice	C	F
thin	thing	thick	carve	C	F
tomb	tune	tooth	pan	C	F
verve	verb	verse	game	C	F
whip	whiff	wit	hers	C	F
win	wing	whip	beach	C	F
wing	win	wick	duck	C	F
wipe	wife	white	killed	C	F
bad	bed	bide	fall	V	M
bag	beg	big	goal	V	M
bed	bad	bod	ram	V	M
beg	bag	big	sewn	V	M
bull	ball	bell	foam	V	M
but	boat	bought	thick	V	M
chows	chose	cheese	said	V	M
clown	clone	clean	find	V	M
cooed	code	kid	vet	V	M
cop	coop	keep	beef	V	M
cut	coat	kite	seek	V	M
dug	dog	dig	herds	V	M
fail	fell	feel	chows	V	M
fall	fool	fill	curb	V	M
find	fond	fund	lib	V	M
fond	phoned	fend	patch	V	M
fool	foal	fail	train	V	M
fun	phone	fawn	sad	V	M
gild	good	gold	pull	V	M
gone	gown	goon	please	V	M
gown	gone	gain	tuck	V	M
gun	gone	gain	bathe	V	M
hop	hope	hip	rode	V	M
killed	could	cooled	teen	V	M
late	let	light	robe	V	M
lick	leak	lack	wane	V	M

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
lip	leap	loop	clear	V	M
look	Luke	luck	wet	V	M
mess	mass	miss	rove	V	M
mice	mess	mass	neat	V	M
moss	mouse	mean	note	V	M
mouse	moss	man	lab	V	M
night	not	note	course	V	M
not	note	net	robes	V	M
pain	pen	pun	fun	V	M
pan	pen	pine	moss	V	M
pat	pet	pot	clean	V	M
pen	pan	pawn	gild	V	M
pet	pat	pit	clown	V	M
pine	pen	pawn	crock	V	M
pond	pound	pinned	night	V	M
pool	pole	peel	seen	V	M
pot	pout	pat	beg	V	M
pound	pond	pinned	holes	V	M
pout	pot	put	vine	V	M
pull	pool	peel	live	V	M
put	pot	pet	chip	V	M
rain	wren	ran	beats	V	M
ride	rod	rid	though	V	M
rod	rude	read	seem	V	M
room	roam	ream	patch	V	M
sad	said	sod	pet	V	M
said	sad	sod	coat	V	M
shot	shout	sheet	fleas	V	M
shout	shot	shoot	wait	V	M
silt	soot	salt	god	V	M
sit	seat	sat	batch	V	M
slide	sled	slowed	bull	V	M
soon	sewn	sin	quack	V	M
soot	soak	salt	green	V	M

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
suck	sock	sack	gun	V	M
tan	ten	toon	peace	V	M
ten	tan	tune	quick	V	M
town	tone	tune	chose	V	M
tune	tone	town	feet	V	M
wait	wet	wheat	pine	V	M
white	wet	wheat	draw	V	M
wren	ran	rune	cores	V	M

APPENDIX B

LIST R - RUSSIAN STIMULI

Listed here are the words used in the Russian variations of Experiments 1, 2, and 3. In each experiment, listeners heard only those words in the Target column.

In Experiment 1, participants were also presented with four types of Visual Primes. Same Primes were identical to the words used in the Target column. Confusable and Non-Confusable primes differed from the Target by a single phoneme. Dissimilar primes had no phonemes in common with the Target. In column 'C/V', the phonation type of the change is indicated by either C (consonant) or V (vowel). The position in the word is listed in the last column; consonant changes could be either word initial (I) or word final (F), but vowel changes were always word medial (M).

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
fuel	fool	jewel	laid	C	I
jeer	cheer	veer	hit	C	I

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
pure	pour	boar	mud	C	I
yet	jet	let	cheer	C	I
back	bag	bat	lid	C	F
bag	back	bad	let	C	F
bat	bad	back	rid	C	F
bid	bit	bib	gloat	C	F
bleat	bleed	bleak	logs	C	F
brad	brat	brag	tote	C	F
brat	brad	brag	green	C	F
bride	bright	bribe	wheel	C	F
cite	side	psych	bad	C	F
clock	clog	clot	frizz	C	F
clog	clock	clod	sip	C	F
cop	cob	cot	ten	C	F
cot	cod	cop	full	C	F
cub	cup	cud	plot	C	F
cued	cute	cube	lost	C	F
cup	cub	cut	flak	C	F
debt	dead	deck	gall	C	F
fix	figs	fits	laud	C	F
flag	flak	flab	cute	C	F
flags	flax	flaps	cup	C	F
flak	flag	flat	grade	C	F
glowed	gloat	globe	brat	C	F
got	god	gob	slid	C	F
grate	grade	grape	pod	C	F
hag	hack	had	claws	C	F
height	hide	hype	clog	C	F
hid	hit	hick	suit	C	F
hog	hock	hop	well	C	F
kid	kit	kick	mess	C	F
kit	kid	kick	hall	C	F
knot	nod	knock	skit	C	F
late	laid	lake	brad	C	F

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
led	let	leg	nip	C	F
let	led	leg	grid	C	F
light	lied	like	pour	C	F
lit	lid	lip	cod	C	F
logs	lox	lots	sick	C	F
loop	lube	loot	bled	C	F
lox	logs	lots	wean	C	F
mat	mad	map	sped	C	F
mate	maid	make	song	C	F
muck	mug	mutt	tagged	C	F
mud	mutt	mug	leap	C	F
mug	muck	mud	sill	C	F
mutt	mud	muck	lied	C	F
nib	nip	nick	tax	C	F
nod	knot	knob	talk	C	F
plod	plot	plop	slept	C	F
plot	plod	plop	stack	C	F
pod	pot	pock	bet	C	F
pot	pod	pop	muck	C	F
rack	rag	rat	nod	C	F
rate	raid	rake	hock	C	F
rib	rip	rid	lend	C	F
right	ride	ripe	kin	C	F
rip	rib	writ	bag	C	F
rod	rot	rob	men	C	F
sat	sad	sap	wed	C	F
serge	search	serve	pot	C	F
skid	skit	skip	least	C	F
slap	slab	slat	cob	C	F
slit	slid	slip	cub	C	F
squad	squat	squab	peak	C	F
squat	squad	squawk	dell	C	F
stack	stag	stat	lox	C	F
stag	stack	stab	led	C	F

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
sued	suit	soup	tell	C	F
suit	sued	soup	hell	C	F
swap	swab	swat	hack	C	F
tack	tag	tap	jet	C	F
tact	tagged	tapped	lawn	C	F
tag	tack	tad	hall	C	F
tags	tax	tabs	phlegm	C	F
taps	tabs	tags	keen	C	F
tied	tight	type	gel	C	F
tight	tied	tyke	figs	C	F
toad	tote	tope	seek	C	F
tote	toad	toke	lag	C	F
tried	trite	tribe	mall	C	F
trite	tried	tribe	rag	C	F
tweed	tweet	tweak	ride	C	F
tweet	tweed	tweak	search	C	F
wed	wet	web	meat	C	F
wet	wed	web	sad	C	F
wick	wig	wit	sheep	C	F
wrote	road	rope	seal	C	F
bail	bell	bill	mutt	V	M
bait	bet	beat	fool	V	M
blade	bled	bleed	pull	V	M
cat	cot	cut	sell	V	M
close	claws	clues	tried	V	M
coal	call	cool	guess	V	M
crick	creak	crook	slab	V	M
dale	dell	dill	naught	V	M
flame	phlegm	flume	kid	V	M
flap	flop	flip	squad	V	M
foal	fall	fool	side	V	M
fool	full	foal	vat	V	M
freeze	frizz	phrase	plod	V	M
gas	guess	geese	spot	V	M

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
goal	gall	ghoul	tweed	V	M
greed	grid	grayed	sin	V	M
green	grin	grain	bell	V	M
grin	green	grain	flop	V	M
grit	greet	great	hog	V	M
hail	hell	heal	creak	V	M
hawk	hock	hook	squat	V	M
heal	hill	hail	sand	V	M
hill	heal	hail	lick	V	M
hole	hall	hull	trite	V	M
hug	hog	hag	call	V	M
hull	hall	hell	pot	V	M
jail	gel	jewel	raid	V	M
keen	kin	cane	moss	V	M
kin	keen	cane	mess	V	M
lace	less	lease	greet	V	M
land	lend	leaned	bit	V	M
lass	less	loss	road	V	M
leak	lick	lake	wet	V	M
leg	lag	lug	walk	V	M
less	lass	lease	scene	V	M
lip	leap	loop	toad	V	M
list	least	lest	track	V	M
load	laud	lewd	tech	V	M
loan	lawn	loon	read	V	M
lust	lost	lest	mad	V	M
mace	mess	miss	bleed	V	M
main	men	mean	lube	V	M
mass	mess	miss	slapped	V	M
mess	mass	miss	pall	V	M
mitt	meat	met	sad	V	M
mole	mall	mill	sued	V	M
muss	moss	mess	knot	V	M
nut	naught	net	grin	V	M

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
pat	pot	putt	quell	V	M
pet	pat	pit	win	V	M
pick	peak	peck	tack	V	M
pole	pall	pool	wig	V	M
pool	pull	pole	hock	V	M
prone	prawn	prune	rot	V	M
quail	quell	quill	trek	V	M
raid	red	rid	rip	V	M
reed	rid	red	pat	V	M
rid	read	raid	mug	V	M
rung	wrong	ring	stall	V	M
said	sad	seed	less	V	M
sail	sell	seal	kit	V	M
scene	sin	sane	flag	V	M
seal	sill	sail	tabs	V	M
seek	sick	sake	heal	V	M
seep	sip	sup	hide	V	M
send	sand	sinned	tight	V	M
sheep	ship	shape	dead	V	M
ship	sheep	shape	lass	V	M
sick	seek	sake	wrong	V	M
sill	seal	sail	bright	V	M
sin	scene	sane	tweet	V	M
skim	scheme	scum	tag	V	M
slapped	slept	slipped	cot	V	M
slept	slapped	slipped	maid	V	M
spade	sped	speed	flax	V	M
spat	spot	spit	god	V	M
stole	stall	steal	red	V	M
sung	song	sing	back	V	M
tail	tell	teal	mass	V	M
take	tech	tick	prawn	V	M
tan	ten	ton	less	V	M
track	trek	trick	swab	V	M

Target	Confusable	Non-Confusable	Dissimilar	C/V	Word Position
trek	track	truck	ship	V	M
tuck	talk	tech	fall	V	M
vet	vat	vote	clock	V	M
wail	well	will	scheme	V	M
wean	win	wane	hill	V	M
will	wheel	whale	rib	V	M
win	wean	when	tied	V	M
woke	walk	wick	stag	V	M

WORKS CITED

Archibald, J. (1997). *Second Language Phonology.* Philadelphia, John Benjamins Publishing Company.

Blumenfeld, R. (2002). *Accents: A Manual for Actors.* New York: Limelight Editions.

Borsky, S., Tuller, B., & Shapiro, L. (1998). "How to milk a coat:" The effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America, 103,* 2670-2676.

Bradlow, A.R. & Bent, T. (2003). Listener adaptation to foreign-accented English. In the proceedings of the *15th International Congress of Phonetic Sciences (15th ICPhS).*

Chreist, F.M. (1964). *Foreign Accent*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Clarke, C.M. (2000). Perceptual adjustment to foreign-accented English. *Journal of the Acoustical Society of America, 107,* 2856.

Crystal, D. (2003). *A Dictionary of Linguistics & Phonetics, Fifth Edition.* Malden, MA: Blackwell Publishing.

Derwing, T.M., & Munro, M.J. (1997). Accent, Intelligibility, and Comprehensibility; Evidence from Four L1s. *SSLA, 19,* 1-16.

Herman, L., & Herman, M. S. (1943). *Foreign Dialects; A Manual for Actors, Directors, and Writers.* New York, NY: Theatre Arts Books.

Elbert, M. & McReynolds, L.V. (1978). An experimental analysis of misarticulating children's generalization. *Journal of Speech and Hearing Disorders, 21,* 136-150.

Fairbanks, G. (1940). *Voice and Articulation Drillbook.* New York: Harper.

Fayer, J.M. & Krasinski, E. (1987). Native and Nonnative Judgments of Intelligibility and Irritation. *Language Learning, 37,* 313-326.

Flege (1995). Second Language Speech Learning: Theory, findings, and problems. In Strange, W. (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* chap. 8, pp. 233-277. York Press, Timonium, MD.

Flege, J.E., Munro, M.H., & MacKay, I.R.A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97,* 3125-3134.

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

International Phonetic Association (1999). *Handbook of the International Phonetic Association.* New York: Cambridge University Press.

James, A.R. (1996). Second Language Phonology. In Jordens, P. & Lalleman, J. (Eds.), *Investigating Second Language Acquisition,* pp. 293-320. Mouton de Gruyter, New York.

Kalikow, D.N., Stevens, K.N., & Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentences materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337-1351.

Kenworthy, J. (1987). *Teaching English Pronunciation.* New York: Longman.

Kilborn, K., Moss, H. (1996). Word Monitoring. *Language and Cognitive Processes, 11,* 689-694.

Ladefoged, P. (1999). American English. *Handbook of the International Phonetic Association,* Cambridge, England: Cambridge University Press.

Lado, R. (1957). *Linguistics across cultures; applied linguistics for language teachers.* Ann Arbor, Michigan: University of Michigan Press.

Laufer, A. (1999). Hebrew. *Handbook of the International Phonetic Association,* Cambridge, England: Cambridge University Press.

Major, R.C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology.* Mahwah, NJ: Lawrence Erlbaum Associates.

McGarr, N. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research, 26,* 451-458.

Miller, G.A., Heise, G.A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41,* 329-335.

Munro, M.J., & Derwing, T.M. (1995a). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning, 45,* 73-97.

Munro, M.J., & Derwing, T.M. (1995b). Processing Time, Accent, and Comprehensibility in Perception of Native and Foreign-Accented Speech. *Language and Speech, 38,* 289-306.

Nissel, J.J. (1959). Analysis of the Predictable Pattern of Error in English among Native Speakers of Japanese. Unpublished Masters Thesis, Georgetown University, Washington, DC.

Nusbaum, H.C., & Pisoni, D.B. (1985). Constraints on the perception of synthetic speech generated by rule. *Behavior Research Models, Instruments, & Computers, 17,* 235-242.

Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics, 60,* 355-376.

Plomp, R., & Mimpen, A.M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology, 18,* 43-52.

Schmid, P.M., & Yeni-Komshian, G.H. (1999). The Effects of Speaker Accent and Target Predictability on Perception of Mispronunciations. *Journal of Speech, Language, and Hearing Reasearch, 42,* 56-64.

Rogers, C.L. (1997). Intelligibility of Chinese-accented English (Doctoral dissertation, Indiana University, 1997). *Dissertation Abstracts International, 58,* 9.

Rogers, C.L., Dalby, J.M., and Nishi, K. (2001). Effects of noise and proficiency level on intelligibility of Chinese-accented English. Poster presented at the 141st meeting of the Acoustical Society of America, Chicago IL.

Schwab, E.C., Nusbaum, H.C., & Pisoni, D.B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27,* 395-408.

Slonimsky, N. (1988). *Perfect Pitch*. New York: Oxford University Press.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics, 25,* 1-24.

Trubetskoi, N.S. (1969). *Principles of Phonology*. (C. A. M. Baltaxe, Trans.). Berkeley, CA: University of California Press. (Original work published 1939).

Van Wijngaarden, S.J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication, 35,* 103-113.

Van Wijngaarden, S.J., Steeneken, H.J.M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America, 111,* 1906-1916.

Weil, S.A. (2000). Foreign Accented Speech; Adaptation and Generalization. Unpublished Masters Thesis, Ohio State University; Columbus, OH.

Weil, S.A. (2002a). Comparing Intelligibility of Several Non-Native Accent Classes in Noise. In J.H.L. Hansen & B. Pellom (Eds.), *International Conference on Spoken Language Processing (ICSLP 2002)* (pp. 253-256). Adelaide, Australia: Causal Productions Pty Ltd.

Weil, S.A. (2002b). Unpublished data. Ohio State University; Columbus, OH.

Weinberger, S.H. (Ed.). (2003). Speech Accent Archive. http://classweb.gmu.edu/accent/.

Whitman, R. (1970). Contrastive Analysis: Problems and Procedures. *Language Learning, 20,* 29-41.