INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.
2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.
3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*
4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

***For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.**

UMI University Microfilms International

8614653

Basson, Sara H.

PATTERNS OF PRONUNCIATION ERRORS IN ENGLISH BY NATIVE JAPANESE AND HEBREW SPEAKERS: INTERFERENCE AND SIMPLIFICATION PROCESSES

City University of New York PH.D. 1986

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

Copyright 1986
by
Basson, Sara H.
All Rights Reserved

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __√__.

1. Glossy photographs or pages ______
2. Colored illustrations, paper or print ______
3. Photographs with dark background __√__
4. Illustrations are poor copy ______
5. Pages with black marks, not original copy ______
6. Print shows through as there is text on both sides of page ______
7. Indistinct, broken or small print on several pages __√__
8. Print exceeds margin requirements ______
9. Tightly bound copy with print lost in spine ______
10. Computer printout pages with indistinct print ______
11. Page(s) ______ lacking when material received, and not available from school or author.
12. Page(s) ______ seem to be missing in numbering only as text follows.
13. Two pages numbered ______. Text follows.
14. Curling and wrinkled pages ______
15. Dissertation contains pages with print at a slant, filmed as received ______
16. Other ______

University
Microfilms
International

Patterns of Pronunciation Errors in English by Native Japanese and Hebrew Speakers: Interference and Simplification Processes

by

Sara H. Basson

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1986

© 1986

Sara H. Basson

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

April 27th 1986
Date

Michael Studdert-Kennedy
Chair of Examining Committee

April 25, 1986
Date

[signature]
Executive Officer

Katherine Harris

Arthur Boothroyd

Lawrence Raphael

Supervisory Committee

The City University of New York

Abstract

PATTERNS OF PRONUNCIATION ERRORS IN ENGLISH BY NATIVE JAPANESE AND HEBREW SPEAKERS: INTERFERENCE AND SIMPLIFICATION PROCESSES

by

Sara H. Basson

Advisor: Professor Michael Studdert-Kennedy

Adult second language (L2) learners often exhibit phonetic deviance in English. Pronunciation errors accounting for reduced intelligibility and perceived accent in native Japanese and Hebrew speakers were investigated. Four native Japanese, four native Hebrew, and four native English male speakers recorded syllables and sentences in English. The Japanese and Hebrew speakers also recorded syllables in their native languages. The segments forming the English syllables were classified in terms of their contrastive status, that is, in terms of whether phonemically identical sounds occur in similar phonetic environments in the speaker's first language (L1). The recorded tokens were randomized and presented to 36 experienced listeners to transcribe and rate in terms of degree of accentedness. Transcriptional errors were interpreted as production errors of the non-native speakers. Two models derived to account for L2 learners' errors were evaluated. The first assumes that errors reflect L1 interference. The second model predicts that certain sounds are inherently more difficult than others, based on cross-linguistic data and findings from L1 acquisition.

Of all consonant groups, fricatives, liquids, and clusters were most deviant for Japanese speakers. These findings are consistent with an interference model, though individual consonants within the consonant classes were not all equally deviant. Vowels were inaccurately transcribed more often than consonants for all language groups, and articulatory explanations for vowel errors are discussed. The Japanese speakers performed worse on vowels that do not appear in their native inventory; Hebrew speakers performed poorly on the entire set. Measures of first and second formants reveal that L1 interference largely accounted for vowel deviance in Japanese speakers. The Hebrew speakers centralized the entire back set of vowels irrespective of contrastive status, which may reflect a natural simplification process. The presence of contrastive segments in an utterance did not affect accentedness ratings, though overall intelligibility was impaired. Accentedness ratings for single syllables and sentences correlated highly, suggesting either that segmental information is critical to perceived accent, or that speakers had similarly good or poor control over both parameters.

ACKNOWLEDGMENTS

A project of this magnitude requires an extensive support network of caring people. Their contributions to my efforts pervade this manuscript, though my thanks to only a few of them appear in these acknowledgments.

To my advisor and mentor, Michael Studdert-Kennedy, I offer my deepest appreciation and affection for being an unending source of warmth, wit, imagination, and intelligence. His guidance has inspired me from my first semester in the Ph.D. program through the final weeks of my dissertation, even at a distance of 3000 miles. His meticulous, insightful comments on this thesis as it evolved made the editing process (almost) enjoyable.

I wish to extend my sincere gratitude to the members of my committee. I have thoroughly enjoyed working with Katherine Harris, whose infectious enthusiasm, careful analysis of data, and constructive suggestions have guided me through the doctoral program as well as through the dissertation experience. Arthur Boothroyd benefited me and this work with his technical assistance, statistical support, and a well-defined approach to research. I thank Lawrence Raphael for serving as outside examiner, and for his valuable suggestions on the final draft of this thesis.

Pamela McPartland, the director of the International English Language Institute of Hunter College, graciously allowed me to use the facilities of the Institute, and worked arduously to enlist ESL instructors as listeners for this study. My gratitude to the many instructors who volunteered to listen to a 65-minute tape.

The stimuli creation and acoustic analysis of data were done at Haskins Laboratories, and I am grateful to the supportive staff for all their help. Particular thanks to Rena Krakow and Doug Whalen for assistance in computer manipulation of a huge data base. My deepest gratitude to Judith Rubin-Spitz for her kind help in the mammoth task of statistical analysis.

My thanks to Judith Pierpont for her assistance in data analysis, and to Judith Klavans, for many and varied kinds of assistance from the earliest to latest stages of this dissertation.

My parents have provided innumerable varieties of support - psychological and physical, spiritual and tangible - throughout the Ph.D. experience and for as long as I can recall before that. Their pride and belief in me has served as a constant source of strength. Without their help, the task would have been considerably less pleasant.

I owe a debt to my young son Yarden, for the many, many hours spent in utero in front of computer terminals, and for many hours spent ex utero, without his mother. This will be a new beginning for us.

I extend my deepest appreciation to Ruth Berkowitz. Her devotion to Yarden has provided me with the clear head necessary to complete this project; her flexible availability has provided me with the time.

An all-consuming endeavor requires an all-patient, all-capable, all-accepting spouse. My husband Amnon has demonstrated these qualities and many more, remaining enthusiastic, optimistic, and easygoing through the most trying stages. He was always there to listen, comfort, advise, and encourage, or provide any emergency assistance called for. I consider myself very lucky to have had him with me throughout graduate school. For us, too, this will be a new beginning.

The best expression of gratitude to those people who have helped me become what I am is to become all I can be. I thank them all for providing me with this opportunity.

Sara H. Basson

April 16, 1986

This work was supported in part by NIH Grant HD-01994 to Haskins Laboratories.

TABLE OF CONTENTS

ACKNOWLEDGMENTS vi

LIST OF TABLES xii

LIST OF FIGURES xiv

Chapter

I. INTRODUCTION 1

II. REVIEW OF THE LITERATURE 8

Models Proposed to Account for Segmental Deviations from Target Norms 8

Contrastive Analysis 8
Alternatives to Contrastive Analysis 9
Domain of Investigation beyond the Segment 12

Specific Features Responsible For Reduced Intelligibility 15

Role of Perception in Production Errors 18

Variability 19

Across L2 Learners 19
Within L2 Learners 20
Across Tasks 22

Prosodic Features in L2 Acquisition 23

II. METHOD 24

Collection of Spoken Materials 24

Talkers 24
Stimulus Materials 26

The Perceptual Experiment 36

Test Preparation 36
Listeners 37

Information Coded for Analysis of Perceptual Results 40

Perceptual Data Analysis 43

Acoustic Analysis 43

IV. RESULTS 46

Perceptual Results 46

The Effect on Intelligibility of Consonant Class and Consonant Position in Syllable 46
The Effect of Vowel Class on Intelligibility 62
The Intelligibility of Consonants vs. Vowels 70
Specific Segmental Confusions 70
The Effect of Contrastiveness on Token Intelligibility 82
The Role of Contrastiveness in Accentedness Ratings 89
The Relation Between Accentedness and Intelligibility 89
Accentedness Ratings for Single Words vs. Sentences 91
Real words vs. Nonsense syllables 91
Summary of Perceptual Results 93

Acoustic Results 95

Overview 95
Reliability and Validity of Acoustic Results 96
Statistical Comparison of F1 and F2 for English and Non-native Speakers: Contrastive vs. Non-contrastive Vowels 99
Vowel-by-Vowel Comparison of Formant Values for Native and Non-native Speakers 99
Relation between Target Intelligibility and Formant Deviation 106
Acoustic Variability of Vowels 107
Five Vowel Systems and Distances Among Vowels 108
Summary of Acoustic Results 110

V. DISCUSSION 112

Models Accounting for L2 Learners' Phonological Distortions 112

L1 Interference and Unintelligible Segments 112
Error Patterns and Alternatives to CAH 115
Predicting the Nature of Errors: CAH and Natural Processes 117

Contrastiveness, Intelligibility, and Accentedness 119

Relation between Contrastiveness and Intelligibility 119
Relation between Accentedness and Intelligibility 119
Features Determining Accentedness 121

Intralingual Vowel Distances and L2 Transfer 124
Intelligibility of Consonants over Vowels 125

Perception and Production of Vowels 126
Vowel Acoustic Deviance
and Loss of Intelligibility 128

Perceptual Consequences of Vowel Acoustic Deviance 132
Stability of L2 Learners' Targets 136
Accentedness Judgments:
Citation Forms vs. Longer Speech Samples 137
L2 Learner Variability 138

VI. SUMMARY 140

VII. APPENDICES 144

Appendix A1. Instructions 144
Appendix A2. Word List 145
Appendix A3. Questions presented
to elicit sentences 147
Appendix A4. Sample of Listener's
Response Sheet 148
Appendix B1. Formant Values of Averaged
English Vowel Data for English Speakers 149
Appendix B2. Formant Values of Averaged
English Vowel Data for Japanese Speakers 150
Appendix B3. Formant Values of Averaged
English Vowel Data for Hebrew Speakers 151
Appendix B4. Formant Values of Averaged Japanese
Vowel Data 152
Appendix B5. Formant Values of Averaged Hebrew
Vowel Data 153

REFERENCES 154

LIST OF TABLES

Table

3.1 Contrastive Analysis of English Phonemic Inventory for Japanese Speakers 28

3.2 Contrastive Analysis of English Phonemic Inventory for Hebrew Speakers 30

3.3 Specific Targets Extracted for Analysis 32

3.4 Classification of Targets in Terms of General Phonetic Class 41

4.1 Percent Error for Consonant Segments Tabulated by Phonetic Class for the Three Language Groups 47

4.2 Percent Error for Consonant Segments Tabulated by Phonetic Class for Japanese and Hebrew Talkers, Corrected against English Scores 49

4.3 Rank Ordering of Consonants, High to Low Intelligibility for Raw and Corrected Scores of Japanese and Hebrew Speakers 50

4.4 Three-Factor Analysis of Variance: Effect of Consonant Class and Position on Intelligibility 51

4.5 Intelligibility as a Function of Consonant Class for Initial and Final Position for Japanese Talkers 54

4.6 Intelligibility as a Function of Consonant Class for Initial and Final Position for Hebrew Talkers 55

4.7 Planned Comparisons Between Consonant Classes in Initial and Final Positions for Japanese Talkers 60

4.8 Planned Comparisons Between Consonant Classes in Initial and Final Positions for Hebrew Talkers 61

4.9 Percent Error for Vowel Classes for the Three Language Groups 63

4.10 Percent Error for Vowel Classes for Japanese and Hebrew Talkers, Corrected against English Scores 64

4.11 Rank Ordering of Vowels, High to Low Intelligibility, for Raw and Corrected Error Scores of Japanese and Hebrew Speakers 65

4.12 Planned Comparisons Between Vowel Classes for Japanese Talkers 68

4.13 Planned Comparisons Between Vowel Classes for Hebrew Talkers 69

4.14 Percent Error: Vowels vs. Consonants, by talker 71

4.15 Percent Transcription Errors for Vowels: English Talkers 73

4.16 Percent Transcription Errors for Vowels: Japanese Talkers 74

4.17 Percent Transcription Errors for Vowels: Hebrew Talkers 75

4.18 Percent Transcription Errors for Consonants: English Talkers 76

4.19 Percent Transcription Errors for Consonants: Japanese Talkers 78

4.19 Percent Transcription Errors for Consonants: Hebrew Talkers 80

4.21 Percent Target Segments Mistranscribed by Contrastiveness Category 83

4.22 Intelligibility of Phonetically Contrastive and Non-contrastive Segments for Japanese Talkers 85

4.23 Intelligibility of Phonetically Contrastive and Non-contrastive Segments for Hebrew Talkers 86

4.24 Mean Formant Values for the Four English Speakers Compared with those for the 33 Speakers of Peterson and Barney (1952) 97

4.25 Percentage Differences for English Speakers' Vowels (Percentaged against Peterson and Barney's Values) 98

4.26 Range of F1 and F2 Across All Vowels in Hz for the Three Language Groups 109

LIST OF FIGURES

Figure

4.1 Percentage of Error as a Function of Consonant Class for Initial and Final Position for Japanese Speakers 56

4.2 Percentage of Error as a Function of Consonant Class for Initial and Final Position for Hebrew Talkers 57

4.3 Percent Intelligibility of Vowel Classes for Japanese and Hebrew Talkers 67

4.4 English and Native Vowels Produced by Japanese and English Talkers 101

4.5 English and Native Vowels Produced by Hebrew and English Talkers 102

Chapter One

INTRODUCTION

Adults are generally described in the literature on second language (L2) acquisition as poor language learners. There is good evidence, however, that they often learn L2 syntax even more readily than children (Krashen, Long, and Scarcella, 1979). The main problem facing an adult second language learner is simply to sound like a native speaker (Tahta, Wood, and Loewenthal, 1981; Oyama, 1976). The poor success of foreign sounding adult speakers in remediating their accents has been ascribed to a biological critical period for language learning (Seliger, 1978), to some yet unknown physiological constraints (Neufeld, 1980), and to affective variables such as "ego preservation" (Tarone, 1978). Sociolinguists point out the ego-preserving importance of retaining identity with the native language and culture (Guiora, Beit-Hallahmi, Brennan, Dull, and Scovel, 1972) and therefore criticize attempts to eradicate foreign sounding speech. However, the advantages of adopting native sounding patterns in the new language are also obvious. Foreign sound patterns not only limit a speaker's participation in the L2 environment, but, if very marked, may actually interfere with intelligibility.

The most prevalent account of persisting non-native speech and language patterns is an interference model that derives from contrastive analyses of the speaker's native language (L1) and the newly acquired language. The Contrastive Analysis

Hypothesis (CAH) proposes that regions where L1 and L2 differ most will be maximally difficult for the learner, while areas of similarity will pose less of a problem (Lado, 1957). This assumes that the learner produces targets appropriate to either L1 or L2, and that no other factors contribute to error.

An alternative approach in second language acquisition theory posits that the regions of difficulty cannot be determined a priori simply on the basis of contrasts between L1 and L2. Some errors may originate from dissimilarity of the new form with forms in L1, but additional sources of error exist, such as overgeneralization of target language material. It is necessary to analyze errors after they have been produced to determine their source. The Interlanguage Hypothesis (ILH) therefore postulates that learners internalize a set of rules that is not simply a carryover from L1, including features such as overgeneralization and undergeneralization (Selinker, 1972). The ILH also claims that the learner's competence changes as the learner moves through various language learning stages, termed "stages of interlanguage" (Richards, 1971).

No large scale studies have investigated which model accounts for the data more thoroughly. If interference accounts for the L2 learner's difficulty, the researcher should be able to predict the loci of phonetic problems as well as the direction in which the accented production will shift. If phonetic problems do not appear regularly in the predicted loci, if the shifts are not in the anticipated direction, or if there is marked variability across speakers, the interference model will be inadequate to

account for the findings. In most studies designed to test the relative roles of L1 interference and of developmental or universal processes on L2 phonological acquisition, a particular segment or feature is examined (Flege 1980; Wode 1980; Eckman 1977). None of these studies accumulated a large number of possible sound sequences in L2 with the goal of determining which sound sequences were most disruptive to the L2 learner's intelligibility, and which model was best supported when several types of error were considered.

The interference model of second language learning predicts the learner's errors based on the source language. The alternative intralingual model cannot yet provide a set of specific predictions, since several factors in addition to L1 contrast are presumed to contribute to the L2 learner's errors. While it is useful to point out the presence of additional processes besides L1 interference, it would be more interesting to specify those processes exactly and to predict which of them will be most prevalent at various stages of language learning (Selinker, 1972). For example, Ingram (1976) has proposed a set of universal simplification processes in L1 acquisition that are "repressed" at later stages of language development. Phonological disorders are then described as the presence of unusual simplification processes or the persistence of normal processes beyond their prescribed time. A similar measure may be appropriate for L2 learners (Hecht and Mulford, 1982). Certain substitutions or distortions might be considered less deviant than others with greater likelihood of changing over time. For example, poor differentiation between vowels /i/ and /ɪ/ might occur at early stages of L2 acquisition for many English learners un-

familiar with the distinction from L1, but differentiation of the two vowels may improve after sufficient L2 exposure. However, the aspiration feature of English voiceless plosives may be more difficult to modify.

Contrasts between two languages resulting in accented speech are commonly investigated at the phonemic rather than the phonetic level (Dickerson, 1975). Segments are the unit of analysis, and those that are not part of the systematic phonology of the speaker's L1 are assumed to be difficult. Yet the phonetic level may prove to be more interesting and fruitful for at least two reasons. L2 learners who have acquired the phonological system of the new language may still exhibit speech errors at both inter- and intra-segmental levels: for example, errors in the ratio of the duration of one segment to that of the next, as well as timing errors within the various articulatory components of the individual segments themselves (Flege and Port, 1981; Mack, 1982).

Few studies have actually set out to determine the phonetic problems in second language acquisition that most significantly interfere with communication. Studies of the phonetic patterns of second language learners typically choose a single segment or feature, compare and contrast its use in a learner's L1 with its use by native L2 speakers, and then measure its acoustic structure as the learner produces it in L2 (Caramazza, Yeni-Komshian, Zurif, and Carbone, 1973). Shifts of such a segment or feature with continued exposure to the target language are also reported in the literature (Flege and Port, 1981). Several studies have compared L2 learners' pro-

ductions of voice onset time (VOT) differences between voiced and voiceless consonants in L1 and L2 (e.g., Williams, 1979). However, there is no evidence that the features examined most extensively, such as VOT, are in fact important contributors to breakdown of a speaker's intelligibility.

Non-native sounding accents surface as production errors in the learner, but they may in fact reveal underlying perceptual problems. Clearly, learners will not produce systematic distinctions they cannot hear. Several experiments using categorical perception paradigms have assessed discrimination abilities of L2 learners (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura, 1975; MacKain, Best, and Strange, 1981; Gass 1983). Results suggest different strategies at different levels of L2 familiarity, with L1 influence decreasing as a function of time. Attempts to explain production errors of L2 speakers must not ignore the possibility of perceptual interference.

Finally, non-native sounding accents may be so slight as to be just detectable, or they may interfere substantially with the speaker's intelligibility. We do not know how important each segmental substitution or distortion is in terms of disrupting the entire system. Nor do we know whether the factors that mark a speaker as a second language learner are the same as those that, in extreme forms, interfere with the speaker's intelligibility. Perhaps "unintelligible" speakers of the target language exhibit additional distortions not found in their more intelligible, but still clearly accented, counterparts. Most remediation techniques assume that all L2 learners of the

same language background exhibit similar problems along a continuum of proficiency. This view must be partially accurate, since listeners are often able to identify the L1 backgrounds of L2 speakers. Therefore, there must be some shared characteristics across all speakers from the same language background. But similarity is not identity.

The general purpose of the present study was to trace the origin of L2 learners' difficulties with the sound pattern of English by examining a representative sample of English syllables, spoken by native Japanese and Hebrew speakers. Two main hypotheses were considered:

(1) The learner's difficulties are primarily the effect of interference from L1;

(2) The learner's difficulties are primarily the effect of natural phonetic simplification processes (e.g., vowel centralization).

The quality of the speakers' productions was assessed perceptually by experienced listeners who were asked both to transcribe each utterance (thus providing a measure of its intelligibility) and to rate its degree of accentedness. A subset of the utterances (those found to be least intelligible) was analyzed acoustically and compared with the target values of a control group of English speakers.

Specific questions addressed include the following:

(1) Which model better accounts for pronunciation errors of L2 learners: L1 interference or natural simplification processes?

(1.1) Are there systematic differences between Japanese and Hebrew speakers reflecting differences in their L1 phonologies?

(2) Which classes of sound most significantly interfere with phonetic intelligibility and/or account for perception of accentedness in the L2 learner?

(2.1) What is the acoustic basis of poor intelligibility?

(2.2) Does intelligibility correlate significantly with the extent of acoustic deviance from the target?

(2.3) Does the L2 learner have stable but incorrect targets, or appropriate targets with high variability in production?

(2.4) Are the L2 learner's phonetic difficulties rooted in problems of perceiving L2 targets, producing them, or both?

(3) What is the relation between accentedness and intelligibility?

(4) Do judgments of accentedness for isolated, citation form syllables and for longer speech samples correlate significantly?

Chapter Two

REVIEW OF THE LITERATURE

2.1 Models Proposed to Account for Segmental Deviations from Target Norms

2.1.1 Contrastive Analysis

The contrastive analysis hypothesis was described earlier as one model for L2 acquisition patterns. The assumption underlying the CAH is that speakers transfer native language phonemes when attempting target language forms. Lado (1957:12), an early proponent of interference theory, states:

> In learning the sound system of a foreign language, one finds sounds that are physically similar to those of the native language, that structure similarly to them, and that are similarly distributed. "Learning" of such phonemes occurs by simple transfer without difficulty. On the other hand, one also finds sounds that are not part of the sound system of the native language, that structure differently, or that are differently distributed. Learning of these occurs more slowly, and difficulty with them is more persistent. In fact, learning of the latter actually means learning the sounds of the language. We therefore seek to find those problems, and we will find them by the structural comparisons of the two sound systems.

In its strongest interpretation, the CAH can be used to predict difficult targets for L2 learners (i.e., targets not present in L1) as well as what types of substitutions the speakers will produce (i.e., phonetically similar sounds from L1). The strong version of the contrastive hypothesis was tested by Johanssen (1973), who had speakers of various language backgrounds produce speech samples in Swedish. She found that learners did indeed pronounce Swedish sounds most deviantly either when there were no sound equivalents between their L1 and Swedish (e.g., front rounded vowels for English speakers) or when the Swedish sounds were similar but not identical to L1 forms (e.g.,front rounded vowels for Danish speakers). Additional general processes in attempts to handle unfamiliar sounds seemed active for all speakers, regardless of L1 background. Thus, high vowels in Swedish with no L1 counterpart were frequently mispronounced as lower vowels, while front or back vowels without L1 counterparts gravitated towards the central region: the pattern seemed to be to move targets from extremely high, low, front, or back position towards the tongue's rest position.

2.1.2 Alternatives to Contrastive Analysis

Predictive statements derived from contrastive analyses clearly cannot account for all the sounds a speaker finds easy or difficult. For example, although English has no /ʃ/ clusters besides /ʃr/, native English speakers have no problems in producing or discriminating /ʃm/ or /ʃt/. On the other hand, word-initial /ts/ or

/pt/ will often be replaced by single consonants, which are presumably simpler (Kohler, 1971) (e.g., as in tsetse and pterygoid).

Since several types of error are difficult to predict from principles of transfer alone, Eckman proposed that contrastive analyses be supplemented with considerations of universal principles (Eckman, 1977). He suggested that forms in the learner's L2 will only pose a problem if they both differ from forms in the learner's L1 and are linguistically "marked," compared to the nearest L1 equivalents (Cairns, 1969). For example, voiced final stops occur in English, yet, based on frequency of occurrence in different languages and on data from child language acquisition (Locke, 1983), they are more marked than voiceless final stops. In German, only voiceless stops occur in final position, and German speakers learning English tend to devoice stop consonants in final position. Thus a German speaker's substitution of /kæp/ for "cab" in English is predictable, due to L1 interference coupled with the fact that voiced final stops are marked. This is referred to as the "Markedness Differential Hypothesis" (MDH).

The MDH is a special case of a general approach to study of L2 acquisition that appeals to so-called "natural processes." "Naturalness," presumed to reflect anatomical and physiological properties of the human speech mechanism, is inferred from frequency counts in L1 acquisition and from adult usage across typologically distinct languages: frequent phones, features and processes (e.g., deletions, substitutions) are taken to be, in some sense, more natural. Locke's (1983) comprehensive

compilation of phonological processes across languages reflects marked similarities among the consonantal sounds that infants babble and use while acquiring L1 phonology, and the sounds exploited for phonological distinctions across several languages. The example of devoicing word-final obstruents was noted earlier. Other examples include stopping (the substitution of a homorganic stop for a fricative), word-final consonant deletion, and cluster simplification. Locke concludes that the sounds children produce most accurately also occur most frequently in the world's languages. If the sounds children produce earliest and best reflect some sort of natural phonetic process, we might expect L2 learners to perform better when they must learn sounds in L2 that exploit natural abilities (irrespective of whether they exist in L1), than when they must learn sounds that violate these natural processes. However, given Locke's claim that processes preferred in L1 acquisition are exploited in many languages, it may be difficult to tease apart sources of phonological errors in L2 acquisition. Are speakers reverting to L1 acquisition patterns or are they using universal simplification processes?

An example of research exploring the "naturalness" hypothesis comes from Davidian and Flege (1984). They evaluated the difficulty of /g/ vs. /b/ and /d/ for Polish and Spanish speakers in word-final position. In word-final position, Polish allows no voiced stops, while Spanish rarely allows stops at all. The palatal stop is the least frequent of the three voiced stops in the world's languages. Accordingly, if L1 interference alone determines performance, no differences among /b/,/d/, and /g/ would be predicted for either language group, while a theory based on natural proc-

esses would predict that both groups would experience more difficulty with /g/ than with /b/ or /d/. The outcome favored L1 interference.

2.2 Domain of Investigation beyond the Segment

One way of studying the relative roles of L1 interference, language universals, and developmental processes in interlanguage (IL) phonology has been through language differences in syllabification (Anderson, 1982; Broselow, 1983 ; Broselow, 1984; Tarone, 1980). Broselow (1984) points out the importance of considering issues in IL phonology at the level of the syllable, claiming that syllable structure restrictions are particularly susceptible to transfer. Broselow investigated consonant cluster production of Iraqi and Egyptian Arabic speakers. Iraqi Arabic, unlike Egyptian Arabic, allows consonant clusters (e.g., /θnɪn/), but has an optional rule allowing insertion of epenthetic /i/ before the cluster (e.g.,/iθnin/). Therefore, transfer principles accurately predict Iraqi production of /ʔiflɔr/ = "floor," /ʔiθri/ = "three" in English, as well as the fact that fewer cluster related mistakes were detected among Iraqi speakers than among Egyptian speakers. However, interference could not initially predict the Egyptian speakers' choice to use an epenthetic vowel to interrupt the cluster of /θr/ in "three," producing /θiri/ , rather than other means available, such as deletion of one of the consonants. A closer look at the phonological rules of Egyptian dialect, however, provides adequate explanation within the CA model. In Egyptian Arabic, underlying C-C-C are separated as follows:

$$0 \rightarrow i / CC __ C$$

Thus, katab + t + l + u $\rightarrow$ /katabtilu/ . Iraqi dialect, however, resolves the same underlying cluster in the following fashion:

$$0 \rightarrow i / C __ CC$$

Thus, katab + t + l + a $\rightarrow$ /katabitla/.

To explain these differences, it is necessary to refer to a syllabic rather than a segmental analysis. Iraqi, which allows CV,CVC, CCVCC clusters, would divide the example above as follows: (ka)(tab)t(la). Egyptian dialect would similarly derive (ka)(tab)t(lu), since it allows CV,CVC, CVCC. In Iraqi phonology, the separation rule inserts a vowel to the left of a non-syllabified consonant, yielding a closed syllable: C $\rightarrow$ /i/C. Egyptian phonology inserts the vowel to the right of a nonsyllabic consonant, resulting in an open syllable, C $\rightarrow$ C/i/. This analysis predicts the differences found between the two dialects in handling English consonant clusters, capturing the consistency between Iraqi and Egyptian rules for epenthesis.

An additional study by Broselow (1984) analyzed American English and Egyptian Arabic word juncture across syllable boundaries, a phenomenon generally permissible in this Arabic dialect, but not in American English. A strict contrastive analysis would expect the Arabic speaker to produce "this ink" in English as /ðɪ sɪŋk/, since cross word syllabic linkages are the rule in Arabic. In fact, Arabic speakers produce /ðɪs ʔɪŋk/. Broselow attributes the insertion of a glottal stop to universal principles, as most languages of the world permit insertion of /ʔ/ before

initial vowels, thereby creating an unmarked CV sequence opening the syllable. Thus she also advocates incorporating a notion of relative markedness of the target form into the CAH.

The most frequent syllable structure in the world's languages is the CV syllable, with a single consonant onset and a vowel peak. Anderson (1982) investigated IL syllabification of native Chinese and Arabic speakers learning English. Arabic permits 2-consonant clusters initially, while neither Chinese dialect investigated allows clusters at all. Based on the CAH, the Arabic group was expected to perform better than the Chinese groups on clusters of two consonants, and this was confirmed. Arabic phonology also has a rule for dividing consonant clusters with epenthetic vowels (see above) from which it was predicted that Arabic speakers would frequently break up word-initial consonant clusters. In fact, deletion was the most common strategy for both groups, and this corresponds to an L1 developmental strategy. Thus Anderson proposes that both L1 transfer and developmental processes need to be considered in describing L2 phonological acquisition.

Several studies investigating syllable structure in IL phonology ignore differences that may result from the phonetic context of the target segment or cluster. The production of L2 consonant clusters by L2 learners needs to be analyzed not only with reference to whether those clusters exist in L1, but also where in the syllable those clusters may appear in L1. Greenberg (1983) found that Turkish and Greek speakers produced errors in clusters that existed in their respective native languages,

but were not permissible in certain positions within the syllable. Thus, a contrastive analysis which merely compares inventories across languages may not adequately predict loci of difficulty. Additional support for this approach comes from the literature on misarticulating children, where context of the target sound is relevant. Hodson and Paden (1981) noted that both normally developing and speech impaired children frequently misarticulate /l/ postvocalically. Their normal group, however, generally produced prevocalic /l/ accurately, while their speech impaired group produced /l/ inaccurately in both contexts.

2.3 Specific Features Responsible For Reduced Intelligibility

It appears intuitively plausible that certain deviations in L2 phonetic output are more disruptive to the speaker's overall intelligibility than others. Yet little IL research considers which features require more precision or contain more linguistic information than others. Altenberg and Vago (1983) analyzed the English speech production of two native Hungarian speakers. Speaker A used more sound substitutions, more extensive L1 phonological transfer, and more articulatory/acoustic simplifications than speaker B. Naive listeners overwhelmingly identified speaker A as sounding more foreign than speaker B. The authors concluded that the differences noted between speakers A and B contributed to native speakers' perception of subject A's speech as less native-like than speaker B's. This analysis did not determine which articulatory features are most critical, since speaker A demonstrated all possible general characteristics that might correlate with foreign sounding speech.

In a study by Brennan, Ryan, and Dawson (1975), eighteen phonological variables typical of Spanish-accented English were isolated, such as substitutions of /s/ for /z/, /b/ for /v/. Native English listeners, evaluating tapes produced by native Spanish speakers talking English, rated the speakers on a scale of one to seven. The number of mispronunciations gleaned from transcriptions and the ratings by the native English listeners correlated highly. However, further breakdown into the critical features most strongly determining accentedness was not attempted.

The studies cited above provide an inventory of deviations from the target norm for speakers of a given L1, but a simple inventory of deviations from the target norm does not suffice to predict the degree of a speaker's accent. Not all deviations have equivalent effects on the learner's speech, some errors sounding more foreign than others. For example, Briere (1966) pointed out that the substitution of an alveolar for a dental target was not perceived by Vietnamese judges listening to accented Vietnamese, while substitution of a non-aspirated for aspirated sound was immediately detected.

Attempts to determine acoustic factors critical to intelligibility appear in the literature on evaluation of deaf speech. Monsen (1978) obtained intelligibility scores on 37 hearing impaired subjects. He then analyzed their speech production in terms of acoustic variables such as distinctions between voiced and voiceless stops, variation in F1, F2 and F0. The three best predictors of overall intelligibility were VOT distinctions between /t/ and /d/, F2 differences between /i/ and /ɔ/, and per-

ceptual adequacy of liquids and nasals. Such a ranking allows the researcher to devise not only an inventory of deviant sounds, but also a ranking for each feature in terms of its importance for overall intelligibility.

Singh and Black (1966) investigated speaker-listener relations in the perception of consonants when speakers and listeners were from the same or different language groups. It was found that sounds were best identified when speaker and listener shared the same L1. The unaspirated [p] of Hindi speakers was perceived primarily as /b/ by native English listeners, but as the intended /p/ by Hindi listeners. However, /s/, /ʃ/, /h/, /m/, /n/, and /l/ were perceived accurately by listeners of different native backgrounds. This suggests that certain sounds may be inherently more confusable perceptually, irrespective of L1 background. Conversely, certain sounds may be inherently easy to perceive, despite considerable deviation from the target norm. The learner may have little difficulty with L2 sounds that allow for a considerable range before production is considered deviant, foreign, or unintelligible. The L2 learner may have more difficulty learning a new target similar to an L1 sound that demands a high degree of precision to be perceived accurately.

Nonetheless, listeners may be sensitive to seemingly trivial, phonologically inert phonetic distinctions in a foreign language. For example, Flege and Hammond (1982) set themselves the modest goal of determining whether language learners can even detect certain phonetic differences. They took shift in a learner's productive

output to reflect some perceptual awareness. In their study, native English speakers mimicked a Spanish accent in English. In addition to the expected substitutions, the authors found subjects shifting VOT production in a direction appropriate for Spanish. Therefore, non-distinctive phonetic differences between languages are perceptible to listeners not proficient in that language.

2.4 Role of Perception in Production Errors

An obvious problem arises in measuring perceptual ability through productive skills: it is never clear at which level the breakdown has occurred; whether the input was interpreted deviantly or whether the listener/speaker is unable to perform the productive task effectively. In an attempt to assess L2 learners' perceptual abilities, Scholes (1968) presented native speakers of various languages with a tape recording of synthetic vowel sounds. Subjects identified the vowels first as though they were native language vowels, and a second time as though they were English. English classifications were predictable from native language responses, rather then from the inventory of English vowel phonemes. Interpretation of the findings must be cautious, however, as the study does not report proficiency levels or familiarity of the non-native speakers with English. If subjects were relatively proficient in English, the results suggest that non-native English speakers both produce and perceive English in terms of their native language phonetic systems. However, task-specific factors mitigate this interpretation. Stevens, Liberman, Studdert-Kennedy, and Ohman (1969) found that American English subjects were able to discriminate

vowels that were non-phonemic in English as well as Swedish listeners for whom the vowels were familiar. English speaking subjects performed worse than Swedish subjects, however, on absolute identification. The experiment by Scholes required absolute identification only. The non-native speakers were unable to label the stimuli correctly according to English categories, but it remains unclear what this reflects about their perception of distinctions in English.

2.5 Variability

2.5.1 Across L2 Learners

Researchers listing the typical errors of a given native language group learning a second language have never evaluated whether certain deviations from the target are evidenced only in the speech of highly unintelligible non-native speakers. This issue was recently investigated with reference to articulatorily impaired children. Hodson and Paden (1981) questioned whether the phonological processes used by speech impaired children were qualitatively or quantitatively different from the processes employed by normally developing children. They found several processes peculiar to the phonologically impaired children. Thus a partial dissociation was noted, with unintelligible children employing strategies such as final consonant deletion, substitution of glottal stops for consonants, and deletion of weak syllables, while normal children demonstrated strategies like depalatalization (e.g., /ʃ/ → /s/) and assimilation. Similar dichotomies might be found for intelligible vs. unintelligible

non-native English speakers. Processes employed by the latter group may be unlike any processes detected in less foreign- sounding speakers.

A transcriptional study by Dickerson (1975) described IL phonological variants as similar across speakers of the same L1. However, Flege and Davidian (1983) found wide variability even within language groups with respect to devoicing final stops. The discrepancy may stem from the different methods used to collect data. Several studies depend on transcriptional notation to classify variants, while Flege and Davidian analyzed their data spectrographically. Phonetic transcription can smear slight differences by clumping phonetically different variants into a single category. In order to determine clearly the similarities and differences among variants, a finer acoustic analysis must be performed.

2.5.2 Within L2 Learners

Intrasubject variability among hearing impaired speakers has been found to be greater than intrasubject variability among normal speakers (Rubin, 1983). Similar results may be found in IL phonology when the issue of variability is carefully investigated. In fact, variability seems to be built into IL syntax, morphology, and phonology. Several models strive to account for it (Tarone, 1982).

In a study of variability in IL phonology of Japanese speakers learning English, Dickerson and Dickerson (1977) asked the following questions: Does the L2

learner's pronunciation reflect a random approximation of English sounds, or is it systematic? Does the pronunciation change over time in some patterned, predictable way? Do learners from the same L1 exhibit similar patterns of English phonology?

In response to the first question, it was found that learners produced several variants for a given target, but that the variants were systematic with respect to linguistic context. Vowel height appeared to affect the selection of elements and the relative frequency of the variants chosen; for example, a non-flapped /r/ occurred more frequently before low than high vowels. Non-linguistic context also played a role in determining variants chosen, as the correct form was elicited most in word list reading and least in free speech. In response to the second and third questions listed above, all ten subjects used the same set of variants in the varying linguistic and non-linguistic contexts, supporting the view that listeners of the same background demonstrate similar phonological patterns. Finally, it was found that learners change in similar directions over time.

Since IL phonology is in a constant state of flux, it is of interest to determine whether the strategies employed at one stage differ from those at another. Studies investigating IL syntax suggest that L1 transfer may prevail at elementary stages of L2 acquisition, while overgeneralization prevails at more intermediate stages (Taylor, 1975). An example of a syntactic overgeneralization error would be "He may goes," where the L2 speaker has learned to use third person singular affixes, but overgeneralizes their usage to inappropriate contexts. Lobo and Yoshida (1982)

asked whether similar results could be found for perceptual acquisition of L2 phonology. Japanese students learning English were presented with actual and nonsense minimal pairs, with instructions to circle the word they heard (e.g., /so/ - /θo/). The subjects produced overgeneralization errors when both test words were familiar (e.g., hearing /s/ as /θ/) and more transfer errors when neither word was familiar (e.g., hearing /θ/ as /s/). Thus, when presented with meaningful pairs, subjects chose the less familiar phone even when it was incorrect. When listeners had poor control over the content of the target items, they were unlikely to hear the less familiar phone. The authors concluded that more overgeneralization and fewer transfer errors occur when learners are more familiar with L2 phonology.

2.5.3 Across Tasks

Speech behavior of native speakers differs depending on the register of the task. Locke (1983) claimed that "laboratory speech," such as reading isolated words, conceals speakers' natural tendencies to the point where dialectal differences are nearly erased. The advantage of "laboratory speech" is that instrumental analysis is more precise, and comparisons across subjects are more uniform. Similar variability across tasks probably occurs in non-native speech. Thus, conclusions about the phonetic output of L2 learners elicited in one setting may not generalize to other settings. Tarone (1982) also points out that much variability can be explained in terms of the demands of the task, but claims that vernacular style exhibits the most stable forms and tasks demanding a more metalinguistic focus elicit more variability.

A model based on a Chomskyan competence paradigm cannot effectively account for the source of variability, since the goal of a competence model is to ignore vagaries of performance. Tarone constructs a "capability continum" based on Labov (1969) to account for the range of styles available to L2 learners. Varying styles of the speaker's IL system are displayed depending on the demands of the task. Elicitation tasks will reflect a different style than unattended tasks, but both types of data reflect viable aspects of the IL systems. Thus the language produced in either setting is valuable towards understanding the components that contribute to the IL continuum. The researcher must recognize the potential style shifts depending on the task, however, and analyze data elicited through different tasks separately.

2.6 Prosodic Features in L2 Acquisition

L2 learners' errors at the segmental level have been discussed in the preceding sections. Non-native speakers also exhibit difficulty controlling the prosodic features of L2; in fact, errors in intonation, stress, and juncture are often cited as the primary components of foreign accents (Bush, 1967; Stevick, 1976). However, little research has been done on L2 learners' prosody. It would be interesting to determine whether control over segmental and prosodic features in L2 develops separately or in parallel.

Chapter Three

METHOD

3.1 Collection of Spoken Materials

3.1.1 Talkers

Four male talkers from each of three native language groups were used. There were four native Japanese, four native Hebrew, and four native English talkers (controls). The Japanese and Hebrew talkers had spoken only Japanese or Hebrew before coming to the United States, although they had studied English in school in their native countries. The native English talkers spoke a standard variety of New York English. Male talkers only were chosen for two reasons. First, the formant structure of a male utterance can be measured more reliably than that of a female utterance with currently available instrumentation. Second, with a sexually homogeneous group, listeners could less easily label talkers and identify the individual's voice for each token on tapes that included tokens from all talkers in randomized order.

The non-native English talkers were selected from language institutes in New York City, and had all been placed in intermediate or advanced English as a Second

Language (ESL) classes, based on placement exams at the beginning of the semester, or had been promoted to these levels after completing introductory level classes. A certain level of English proficiency was important to ensure that the students would be able to read the stimulus materials (even though reading ability was not being assessed). However, high levels in ESL classes reflect a student's syntactic abilities and vocabulary skills more than his pronunciation skills.

The Japanese and Hebrew talkers had been in the U.S. for at least three months, but no more than one year. Thus, the talkers' immersion times in English were fairly equivalent. These non-native talkers ranged in age from 19 to 28. If, as the Critical Period Hypothesis claims, acquisition of native-sounding accents in a new language becomes difficult or impossible after the speaker reaches puberty (Lenneberg, 1967; Oyama, 1976), adults of different ages should be on the same language learning plateau. Alternatively, second language learning flexibility may continue to deteriorate as the speaker gets older (Seliger, 1978). The age group chosen was clearly past the critical period for language acquisition, yet close enough in age to assume roughly equivalent language acquisition abilities, if chronological age is a determining factor.

3.1.2 Stimulus Materials

Contrastive analyses of the phonemic inventories of English, Hebrew, and Japanese were derived. The phonemic inventory of English was then subdivided into three classes:

(1) sounds for which there is a phonemically identical sound in Hebrew and/or Japanese, with phonological constraints similar to those in English;

(2) sounds for which there is a phonemically identical sound in Hebrew and/or Japanese, with phonological constraints unlike those in English;

(3) sounds that do not occur at all in Hebrew and/or Japanese.

These three categories will be referred to as "non-contrastive sounds," "phonologically contrastive sounds," and "phonetically contrastive sounds," respectively. Tables 3.1 and 3.2 list the members of these categories for Japanese and Hebrew, respectively.

Certain items in the "non-contrastive" category (marked with an asterisk in Tables 3.1 and 3.2) would be considered "contrastive" according to standard Japanese and Hebrew phonological description (Chayen,1973; Martin,1952). However, these segments have recently entered Hebrew and/or Japanese through loan words. Native informants of Japanese and Hebrew were asked to decide on the acceptability of the sounds in various environments in modern Japanese (Tokyo dialect) and modern standard Hebrew. When the sounds were acceptable in a particular context for several words, they were classified as "non-contrastive." Since this study as-

sesses the phonetic skills of talkers, this approach seemed more appropriate than one based on prescriptive classification systems.

Forty-eight English segments or segment clusters were included as targets, each segment occurring at least twice (Table 3.3). A total of 132 tokens was devised, such that a given token often carried more than one target. All consonants were presented in both initial and final position, unless this violated English phonotactic structure. Vowels appeared only in medial position.[1] The test items were primarily single syllables, but four two-syllable words were included in order to assess pronunciation of /l/ and /r/ in medial contexts. The single syllable format was chosen to eliminate the redundancy available to a listener in multi-syllabic real words.

Standard speech intelligibility tests such as the closed response set for consonant differentiation derived by House, Williams, Hecker, and Kryter (1965) were considered inappropriate for the specific goals of this study. One objective of the present study was to investigate how L2 learners handle sounds which do not appear in their native phonetic inventories. By devising word lists with this purpose in mind, a sufficient number of contrastive sounds were obtained for further analysis.

---

[1] In some instances, non-contrastive vowels occurred before or after contrastive consonants. This could increase the difficulty of an otherwise easy segment. Ingram (1976) notes that children mispronounce segments they have already mastered when presented in environments with poorly controlled sounds.

Table 3.1

Contrastive Analysis of English Phonemic Inventory for Japanese Speakers

---

Position of Segment:
I=Initial
M=Medial
F=Final

	Vowels	Consonants
1. NON-CONTRASTIVE:[2]		
	/i/	/p/
	/e/	/b/
	/ɑ/	/k/
	/o/	/g/*
	/u/	/tɑ/, /ti/*, /te/, /to/
		/dɑ/, /di/*, /de/, /do/
		/sɑ/, /su/, /se/, /so/
		/zɑ/, /zu/, /ze/, /zo/
		/hɑ/, /hi/, /he/, /ho/
		/ʃ/*
		/m/
		/nɑ/, /nu/, /ne/, /no/
		/ŋ/
		/lɑ/, /le/, /lo/
		/wɑ/
		/jɑ/, /ju/, /jo/
		/tʃ/*
		/dʒo/*, /dʒe/*, /dʒɑ/*
2. PHONOLOGICALLY CONTRASTIVE:		
		All consonants in final position
		/tu/
		/du/
		/si/
		/zi/
		/hu/
		/ni/
		/li/, /lu/, /l/(M,F)
		/wi/, /wu/, /we/, /wo/
		/ji/, /je/

---

[2] Items followed by an asterisk would be considered "contrastive" according to standard Japanese phonological description. See text for further discussion.

	Vowels	Consonants
3. PHONETICALLY CONTRASTIVE:		
	/ɪ/	/f/
	/ɛ/	/v/
	/æ/	/θ/
	/ɔ/	/ð/
	/ʊ/	/r/
	/ʌ/	All consonant clusters

Table 3.2

Contrastive Analysis of English Phonemic Inventory for Hebrew Speakers

---

Position of Segment:
I=Initial
M=Medial
F=Final

	Vowels	Consonants
1. NON-CONTRASTIVE:[3]		
	/i/	/p/(I,M)
	/e/	/b/(I,M)
	/ɑ/	/k/
	/o/	/g/
	/u/	/t/
		/d/
		/s/
		/z/
		/h/
		/ʃ/
		/m/
		/n/
		/ŋ/*
		/l/
		/j/
		/tʃ/(I)
		/dʒ/(I,M)
		/st/, /fr/*, /gr/, /pl/, /br/, /lt/, /ft/
2. PHONOLOGICALLY CONTRASTIVE:		
		/dʒ/(F), /tʃ/(M,F)
		/sk/(F), /nd/(F)

[3] Items followed by an asterisk would be considered "contrastive" according to standard Hebrew phonological description. See text for further discussion.

	Vowels	Consonants
3. PHONETICALLY CONTRASTIVE:		
	/ɪ/	/θ/
	/ɛ/	/ð/
	/æ/	/r/
	/ɔ/	/w/
	/ʊ/	/rks/, /ldz/, /rdz/, /str/
	/ʌ/	

Table 3.3

Specific Targets Extracted for Analysis

Vowels:

/i/, /ɪ/, /e/, /ɛ/, /æ/, /ɑ/, /ɔ/, /o/, /ʊ/, /u/, /ʌ/

Consonants:

/b/, /p/, /d/, /t/, /g/, /k/, /z/, /s/, /ʃ/, /tʃ/, /dʒ/, /f/, /v/, /ð/, /θ/, /h/, /w/, /j/, /r/, /l/, /m/, /n/, /ŋ/

Clusters:

/st/, /fr/, /gr/, /pl/, /br/, /ts/, /sk/, /nd/, /lt/, /ft/, /str/, /rdz/, /rks/, /ldz/

The task items included both nonsense and real words. Nonsense words were included to encourage listeners to pay attention to each segment in the word: They could not guess what they had heard based on the nearest real-word match, if they had not processed all of the segments. Real words were included to make the task less fatiguing for listeners.

Each task item was paired with a rhyming familiar word, such as:

rhyming word	task item
my	ply
hate	nate
day	fray
see	bree

Talkers read the rhyming word and task item, in that order, with a brief pause between them. This method was adopted to ensure that the reading of target items reflected the talker's actual pronunciation of the vowels in a word rather than the irregularity of English orthography. The pronunciation of English consonants was assumed to be orthographically transparent except for /θ/ and /ð/. Where these sounds were intended, talkers were instructed to pronounce the consonant in the same way as they pronounced it in some familiar context. For example:

bead - theed ("th" as in "the")

bit - thit ("th" as in "think")

The instruction sheet and a complete list of the test types, with particular segments of interest underlined, appear in Appendices A1 and A2.

In addition to the 132 test items described above, each talker recorded, in isolation, samples of all the consonant phonemes from his native language followed by all the vowels of his language. This inventory was used for comparison purposes, to determine whether segments mispronounced by the L2 talkers reflected phonetic substitutions of L1 sounds.

The task of this study - reading single words aloud out of context - is clearly not a common language activity, but it was hoped that the talkers performance would reflect, to some large degree, their performance in more familiar contexts. As a check on this, semi-spontaneous sentences produced by the talkers were also presented to listeners, to see whether performance on the single word task correlated with performance on the more spontaneous sentence task. To obtain these semi-spontaneous sentences, the talkers were presented with eighteen simple questions in English, requiring a relatively constrained response. For example:

Who is Ronald Reagan?

What is your favorite sport?

The talkers were asked to respond in full English sentences, and their responses were recorded. The "spontaneous speech" was elicited in this way to reduce the syntactic and semantic errors likely to occur in the fully spontaneous speech of non-native talkers. Although the English speaking listeners were asked to evaluate these sentences with respect to pronunciation alone, deviant syntax or semantics might have biased their judgments. The sentences were evaluated individually and divided according to whether or not they were syntactically correct. Six responses with the most syntactic errors across speakers were discarded and the twelve remaining sentences were included in the experiment. A complete list of the twelve questions which elicited the responses appears in Appendix A3.

Talker responses were recorded on a TEAC 35-2 Tascam Series tape recorder using an Electrovoice 535A microphone (#2605) in a sound proof booth. The frequency response of the microphone coupled with the recorder was measured using a Bruel and Kjaer sweep frequency sine random generator (#1024), a microphone amplifier (#2603) and graphic level recorder (#2305). The recording session for each talker lasted approximately one half hour.

3.2 The Perceptual Experiment

3.2.1 Test Preparation

The recorded tokens were digitized at a sampling rate of 20,000 Hz, using a Nyquist filter (-3dB at 9.7 kHz, -40 dB at 10 kHz, and high frequency pre-emphasis) on the Haskins Laboratories VAX 11/780. Once digitized, the tokens of interest were segmented by hand using the Haskins Laboratories Waveform Editing and Display Program ("WENDY"). Segmentation was done by both visual and auditory criteria.

The extracted words and sentences served as stimuli for the perceptual experiment. 1,968 stimuli (132 words, 12 sentences, and 20 practice tokens X 12 talkers) were converted to analog form and recorded in random order on reel-to-reel Scotch (#206) tapes. These tapes comprised tokens produced by the talkers for the first time; repetitions were held in reserve. For the word tokens, there were twelve different randomized tapes. Each tape included eleven different single word tokens produced by each of the twelve talkers, accounting for 132 items on each tape. Each item consisted of a given token recorded twice in succession with a 500 ms pause between repetitions, followed by nine seconds of silence before the next item. A longer pause of 15 seconds was imposed after every 22 tokens.

Each word tape included only a single production of the 132 tokens, so that, within a tape, talker and token were confounded. This was not deemed to be a problem, however, because several tokens included the target segments of interest. Since the goal of the study was to look at the influence of native language on the production of these targets, the four talkers from each language group can be considered together. Within talker reliability could also be assessed by considering, across tapes, all the tokens produced by a given speaker containing a certain target.

Finally, twelve output tapes of the semi-spontaneous sentences were recorded, with one sentence produced by each speaker on each tape.

3.2.2 Listeners

The 36 listeners were native American English speakers, drawn from a population of instructors of English as a Second Language (ESL). This population was chosen because its members are familiar with non-native speech, and because ESL instructors may be more stringent than the general population in evaluating foreign accents. In addition, most ESL instructors are familiar with phonetic transcription from their graduate programs. The majority of listeners were instructors at the International English Language Institute of Hunter College.

Twelve sets of tapes were prepared. Each set included a 20-item practice task, a randomized 132-item word tape, and a randomized 12-item sentence tape. Three

listeners heard each set of tapes, played in a quiet room on a Pioneer RT-1050 tape recorder over TDH-39 headphones. The frequency response of the recorder did not vary from manufacturer's specifications by more than .8 dB at any frequency. The entire listening task took approximately 45 minutes.

The listeners were given an instruction sheet explaining the task (see Appendix A1). They were told that they would hear nonsense and real words spoken by native and non-native English talkers. Their task was to transcribe what the speaker said, not what they assumed he was trying to say. They were also asked to rate each token on an "accentedness scale" from 1 to 5. A score of "1" meant that the speaker sounded like a native English speaker; "2" meant that the speaker had a slight, almost imperceptible accent; "3" meant that the speaker had a moderate accent; "4," that he had a strong accent, and "5," that he sounded as though he were not producing English sounds. The scale is taken to have face validity, and to yield ordinal values that may be averaged and submitted to standard statistical procedures for rating scales (Stevens, 1951). A sample response form appears in Appendix A4.

A practice listening task included 20 tokens of real and nonsense words that the 12 subjects had produced specifically for the practice task. The listeners transcribed and rated each token, and any misunderstandings about the task were clarified at this time.

In the "sentences" task, the listeners were asked only to rate the sentences on the accentedness scale from 1 to 5.

The native language backgrounds of the talkers were not revealed to the listeners, so that listeners familiar with the phonological systems of Hebrew or Japanese would not be tempted to transcribe based on what they thought the speaker was trying to say rather than on what he actually said. Similarly, the number of talkers was not discussed beforehand so that the listeners would be less likely to categorize a speaker as, say, native English or extremely foreign, and thus give him the same score whenever his voice was heard rather than evaluating each token individually.

Listeners were permitted considerable flexibility in transcription. While IPA symbols were preferred, listeners could use standard alphabetic symbols if they wished. Standard English orthography is adequate for broad transcription of most consonants, so that symbols were required only for vowels and for the consonants /ð/ and /θ/. While providing listeners with a closed response set would have simplified the task, it would have also constrained the possible range of responses. For example, one target of interest was syllable final consonants produced by Japanese speakers, since Japanese has no consonants in this position. The Japanese L2 learner producing English CVC syllables might delete or distort final consonants, or add an additional vowel to create an open syllable. A closed response set requires <u>a priori</u> decisions as to what the errors will be, which may eliminate some relevant information.

Broadly transcribed responses were preferred over narrow transcriptions. The latter offer too fine a level of detail for a study in which the goal was to see whether the sounds produced fell within the range of acceptable English phonemic categories. Non-phonemic deviations were not deemed relevant for this analysis. Diphthongized transcriptions "/eɪ/" and "/oʊ/" were scored the same as "/e/" and "/o/".

3.3 Information Coded for Analysis of Perceptual Results

Targets were classified in terms of General Phonetic Class, roughly corresponding to manner of articulation, as indicated in Table 3.4. Additional items of information for each target segment within a token included whether the segment: (1) was embedded in a real or a nonsense word; (2) occurred word initially, medially, or finally in the word; (3) was consonantal or vocalic.

Each token was produced by each talker, and the following talker-specific information was coded: (1) native language; (2) whether a phonemically identical segment occurs in similar environments in the learner's L1; whether a phonemically identical segment occurs in other environments in the learner's L1, but is phonotactically unacceptable in the current environment; whether no phonemically similar sound occurs in the learner's L1 (i.e., whether the segment belonged in the non-contrastive, phonologically contrastive, or phonetically contrastive categories, described above).

Table 3.4

Classification of Targets in Terms of General Phonetic Class

SPECIFIC TARGET	GENERAL PHONETIC CLASS
/i/	high front vowel
/ɪ/	
/e/	mid front vowel
/ɛ/	
/æ/	low front vowel
/ɑ/	low back vowel
/o/	mid back vowel
/ɔ/	
/ʌ/	
/u/	high back vowel
/ʊ/	
/b/	voiced stop
/d/	
/g/	
/p/	unvoiced stop
/t/	
/k/	
/z/	voiced fricative
/v/	
/s/	unvoiced fricative
/ʃ/	
/h/	
/f/	
/m/	nasal
/n/	
/ŋ/	

SPECIFIC TARGET	GENERAL PHONETIC CLASS
/dʒ/	voiced affricate
/tʃ/	unvoiced affricate
/w/	glide
/j/	
/l/	liquid
/r/	
/st/,/fr/,/gr/	cluster
/pl/,/br/,/ts/	
/sk/,/nd/,/lt/	
/ft/,/str/,/rds/	
/rks/,/lds/	
/ð/	interdental
/θ/[4]	

[4] The voiced and voiceless "th" are considered as "interdentals" in a single group. Listeners' transcriptional responses did not reliably distinguish them even for the native English talkers.

Finally, the following information was noted for each listener's response: (1) whether the carrier utterance (the real or nonsense word) for the target segment was transcribed correctly; (2) whether the target segment was transcribed correctly, and if incorrectly: (a)the specific sound substituted (b) the general phonetic class of the substitution; (3) the accentedness rating for each token.

3.4 Perceptual Data Analysis

Responses were averaged across listeners to give intelligibility and accentedness scores on each token for each speaker; across tokens to give scores on each phonetic class for each speaker; and across speakers to give scores on tokens, or classes of token, for each language. Intelligibility scores were expressed as mean percentages of error in transcription.[5] Scores per token were determined by averaging across the three listeners who had heard that token; for statistical purposes, tokens were deemed intelligible when at least two of the listeners transcribed it accurately. Accentedness scores were expressed as mean ratings on a scale from 1 to 5.

3.5 Acoustic Analysis

For each talker, central portions of 33 vowels (11 types x 3 tokens) were selected for analysis. A 100 ms section from the center of the vowel was used whenever

---

[5] The terms "accurate transcription" and "mistranscription" are used to indicate that a listener's transcription did or did not correspond to the target the talker was asked to produce. Thus, "mistranscription" refers to the listener's (presumed) correct transcription of a talker's (presumed) erroneous production.

possible. In some cases, however, the entire duration of the vowel portion of the token was less than 100 ms, in which case the entire vowel was analyzed. Frequency values for the first and second formants of each vowel were estimated using the Interactive Laboratories System Linear Prediction Coding (ILSLPC) analysis program. The LPC program uses autocorrelation to calculate reflection coefficients on a frame by frame basis. The sampling rate was set to 20 kHz, and the length of the analysis window was set to 20 ms with a 10 ms step size between adjacent frames. The number of predictor coefficients was set to 24, which appeared to provide estimates consistent with expected values for particular vowels.

The LPC analysis program estimated the two lowest resonances of most of the vowel types submitted. Because these values were extracted from CVC contexts, frames close to the consonantal portions included coarticulatory influences that were not the focus of this study. Therefore, values were chosen from a single frame rather than from the average of several adjacent frames. A number of criteria was used in choosing a frame to represent the formants of the vowel. Frames were chosen:

1. near the center of the vowel, as far as possible from the variations associated with the coarticulatory influences of surrounding consonants;
2. with the highest RMS amplitude;
3. for which the lowest two formants remained steady across two or more neighboring frames.

In a number of cases, the first two formants for back vowels were not reliably distinguished. To determine formant values for these vowels, spectrographic analyses were performed with a Kay Elemetrics Digital Sonograph (7800). When the first two formants were still not separable, the following rules were followed:

1. determine from the spectrogram the lowest and highest values for the merged F1 and F2;

2. find the midpoint between these values;

3. take the frequency 1/3 of the distance below the midpoint as the frequency of F1, and 1/3 of the distance above the midpoint as F2.

For the 8 non-native talkers, three tokens of each vowel type from their native languages were extracted from the contexts /s/+vowel#, /b/+vowel#, and #vowel#. These tokens were submitted to the LPC program with the same parameters as the tokens described above. Vowels where F1 and F2 were not easily distinguished were further analyzed using the Spectral Analysis program (SPA) at Haskins Laboratories, which displays the power spectrum resulting from an FFT analysis.

Chapter Four

RESULTS

4.1 Perceptual Results

4.1.1 The Effect on Intelligibility of Consonant Class and Consonant Position in Syllable

One main question this experiment sought to answer was whether certain phonetic classes are more salient than others in interfering with the non-native talker's intelligibility. The major phonetic class division was between consonants and vowels. Consonants were further divided into stops, fricatives, nasals, affricates, liquids and clusters. Differences in voicing and in the position of the segment within the carrier token were also analyzed. The intelligibility scores, that is, the percentage of error for each consonant class averaged across listeners and speakers, is presented in Table 4.1 for each language group. Fricatives, clusters, and liquids have high error percentages for the Japanese talkers; clusters have high error percentages for Hebrew talkers, followed by nasals, liquids, and fricatives.

Intelligibility scores for consonants produced by English speakers were derived as a baseline in evaluating errors of the non-native groups. If English speakers were highly intelligible for certain classes while unintelligible for others, subtracting the English results from the non-native speakers' results would provide a corrected ver-

Table 4.1

Percent Error for Consonant Segments Tabulated by Phonetic Class for the Three Language Groups

Phonetic Class	English	Japanese	Hebrew
stops	4.63	10.49	12.65
fricatives	9.9	31.76	17.34
affricates	3.03	20.45	14.39
clusters	8.33	34.31	27.94
nasals	9.85	12.88	18.18
liquids	5.0	30.42	17.92
Weighted Mean Percent Error:	7.18	24.53	17.75

sion of the non-native speakers' scores. The English speakers' scores were submitted to a repeated measures analysis of variance with "phonetic class" and "position within the carrier token" as within factors. No significant effect was found for either factor, suggesting that the intelligibility of English talkers was homogeneous across phonetic classes and segment position in the carrier token. Therefore, correcting the scores of the non-native talkers by subtracting the error percentages of the English talkers, and using the corrected scores for further statistical analysis, was rejected. The original and corrected versions of consonant class intelligibility appear in Table 4.2, with rank orderings of high to low intelligibility scores in Table 4.3. For Japanese talkers, the orderings shift in the corrected versions only for classes differing by less than 5% in the uncorrected version. Shifts occur for Hebrew talkers only for differences less than 6%.

The uncorrected intelligibility scores of the non-native speakers were submitted to a repeated measures (four speakers per language) three-factor analysis of variance with "phonetic class" and "position in the carrier token" (initial or final) as within factors and native language as the between variable.[6] Results of this analysis appear in Table 4.4. No significant differences were found between Hebrew and Japanese. The non-significant main effect for language suggests that the Japanese and Hebrew groups were well-matched for overall intelligibility level. No statistically significant difference was found between segments based on their position within the token.

---

[6] Arcsine transformations were performed on percentages before all analyses of variance, in order to normalize the distribution of percentage scores around their mean. The formula used to derive the transformed values was:

$$\text{transformation} = 2 \arcsin\sqrt{X\%}$$

Table 4.2

Percent Error for Consonant Segments Tabulated by Phonetic Class for Japanese and Hebrew Talkers, Corrected against English Scores

Phonetic Class	Japanese	Hebrew
stops	5.86	8.02
fricatives	21.86	7.44
affricates	17.42	11.36
clusters	25.98	19.61
nasals	3.03	8.33
liquids	25.42	12.92
Weighted Mean Percent Error For Consonants: (Corrected)	17.35	10.57

Table 4.3

Rank Ordering of Consonants, High to Low Intelligibility for Raw and Corrected Scores of Japanese and Hebrew Speakers

Japanese

Raw Scores	Corrected Scores
stops (10.49)	nasals (3.03)
nasals (12.88)	stops (5.86)
affricates (20.45)	affricates (17.42)
liquids (30.42)	fricatives (21.86)
fricatives (31.76)	liquids (25.42)
clusters (34.31)	clusters (25.98)

Hebrew

Raw Scores	Corrected Scores
stops (12.65)	fricatives (7.44)
affricates (14.39)	stops (8.02)
fricatives (17.34)	nasals (8.33)
liquids (17.92)	affricates (11.36)
nasals (18.18)	liquids (12.92)
clusters (27.92)	clusters (19.61)

Table 4.4

Three-Factor Analysis of Variance

Effect of Consonant Class and Position on Intelligibility

	Sums of Squares	Degrees of Freedom	Mean Square	F	Tail Probability
Language	1.4680	1	1.4680	1.13	.3278
Error	7.7645	6	1.2941		
Position	.2155	1	.2155	.80	.4047
PxL	.002	1	.002	.01	.9353
Error	1.6100	6	.2683		
Phonetic Class	12.2498	9	1.3612	6.24	*.0001[7]
PCxL	4.3445	9	.4827	2.21	.0350
Error	11.7696	54	.2180		
PxPC	5.8744	9	.6527	4.25	*.0003
PxPCxL	1.2511	9	.1390	.90	.5283
Error	8.3014	54	.1537		
Total	54.8506	159			

[7] * = Statistically significant

However, differences among consonant classes were highly significant ($p<.0001$). That is, certain consonant classes were clearly more instrumental than others in disrupting the talker's intelligibility. A significant interaction between position of the consonant within the token and phonetic class indicates that the effect of position was important for some consonant classes but not for others. The interaction between language group and phonetic class was also significant ($p<.05$), indicating that the effect of phonetic class was different for the two non-native groups.

The interaction between position and phonetic class was investigated further through an analysis of simple main effects, holding phonetic class constant . Positional differences were significant for voiced stops ($p<.05$), nasals ($p<.01$), liquids ($p<.05$), and clusters ($p<.05$).

The interaction between language and phonetic class prompted an analysis of simple main effects, holding language constant. Such an analysis for the Japanese speakers confirms a significant effect of phonetic class ($p<.0001$). There is no significant effect for positional differences, but there is a significant interaction between position and phonetic class ($p<.05$). The differences among phonetic classes for Hebrew talkers also reach significance ($p<.05$). The Hebrew talkers, like the Japanese talkers, show no significant effect based on positional differences, with a significant interaction between position and phonetic class ($p<.01$). The results therefore suggest larger differences among the phonetic classes for Japanese talkers

than for Hebrew talkers, while the effect of position on particular phonetic classes is greater for Hebrew talkers.

The differences among the consonantal classes for Japanese and Hebrew talkers appear in Tables 4.5 and 4.6, and are presented graphically in Figures 4.1 and 4.2 , respectively. The graphs are further broken down in terms of whether the target occurred in initial or final position within the carrier token.[8] For consonants produced by Japanese talkers, more variability across the consonant classes occurs in initial than final position, although positional differences do not reach significance. The most dramatic differences (arbitrarily defined as mean differences >15%) between initial and final positions occur for voiced stops, voiced fricatives, and nasals. For voiced stops and nasals, however, initial position was more intelligible,[9] while final position was more intelligible for voiced fricatives. This is consistent with the finding that positional differences were not significant for Japanese talkers, while there was a significant interaction between position and phonetic class. With the exception of voiced fricatives, the most unintelligible consonant classes in both initial and final positions were liquids, clusters, and interdentals. The loss of intelligibility on these classes provides support for the hypothesis that sounds without equivalents in the talker's native phonetic inventory will present the most difficulty.

---

8 It should be noted that initial and final positions were not symmetrical for fricatives, nasals, or clusters. Fricative /h/ does not occur in syllable final position, while nasal /ŋ/ does not occur initially. Clusters /fr/, /gr/, /pl/, /br/ and /str/ occur only initially, while /ts/, /sk/, /nd/, /lt/, /ft/, /rdz/, /rks/, and /ldz/ appear only in final position.

9 The velar nasal /ŋ/, which occurs only in final position, accounts for a large percentage of the loss in intelligibility. When /n/ and /m/ are considered alone, the difference between nasals in initial and final position for Japanese talkers falls to 7.92%; for Hebrew talkers, 8.75%.

Table 4.5

Intelligibility as a Function of Consonant Class in Initial and Final Position for Japanese Talkers

	Initial	Final
voiced stops	93.75	72.22
unvoiced stops	94.79	96.67
voiced fricatives	43.06	75.0
unvoiced fricatives	83.34	84.53
voiced affricates	79.17	75.0
unvoiced affricates	77.08	86.11
interdentals	37.5	52.08
nasals	97.92	82.14[10]
liquids	73.96	60.42
clusters	70.83	61.11

[10] Excluding / ŋ/, 90.0.

Table 4.6

Intelligibility as a Function of Consonant Class in Initial and Final Position for Hebrew Talkers

	Initial	Final
voiced stops	90.62	86.11
unvoiced stops	91.67	86.67
voiced fricatives	87.5	87.5
unvoiced fricatives	86.91	88.1
voiced affricates	83.33	95.83
unvoiced affricates	75.0	94.45
interdentals	58.33	70.83
nasals	93.75	67.86[11]
liquids	90.38	73.71
clusters	86.46	59.26

[11] Excluding /ŋ/, 85.0.

Figure 4.1

Percentage of Error as a Function of Consonant Class in Initial and Final Position for Japanese Talkers

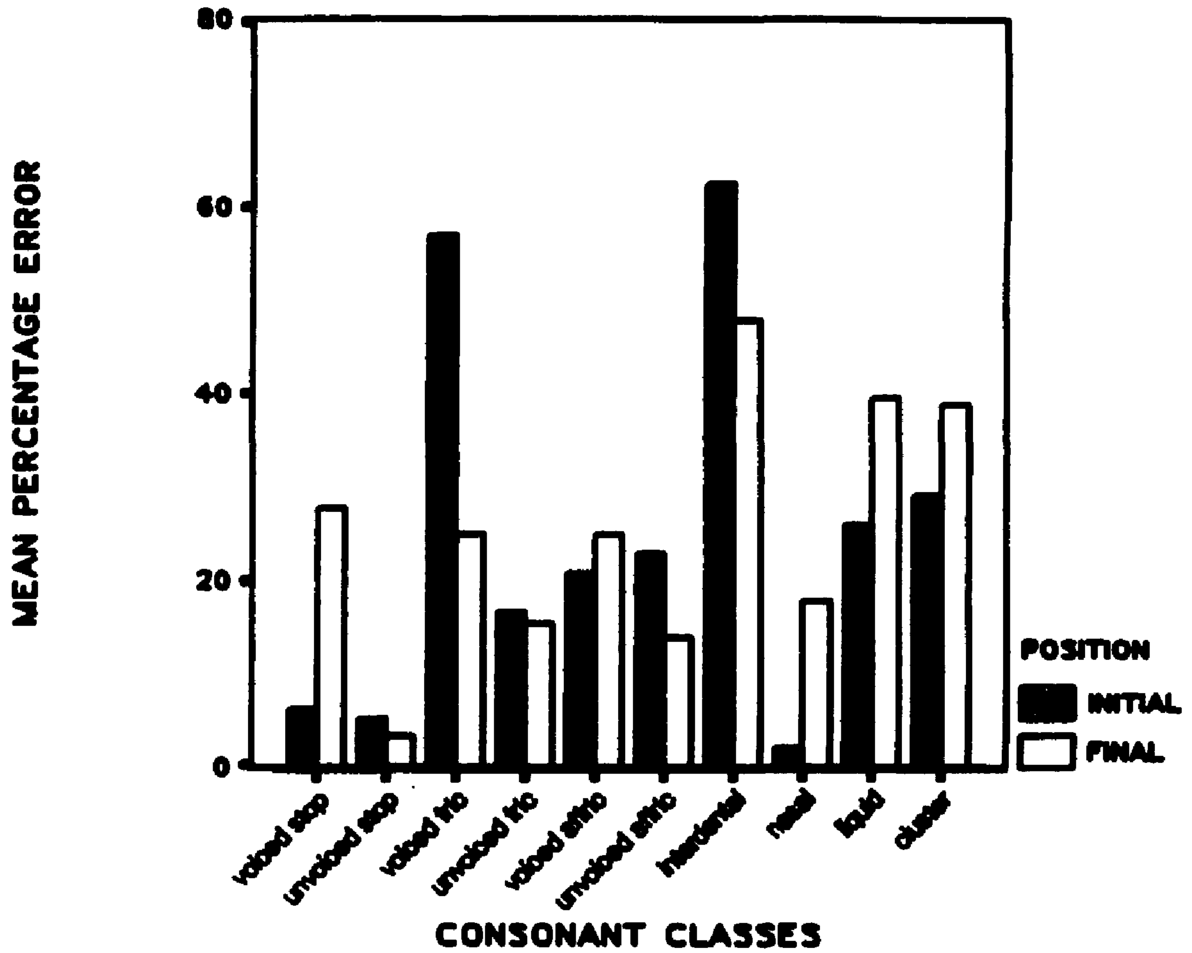
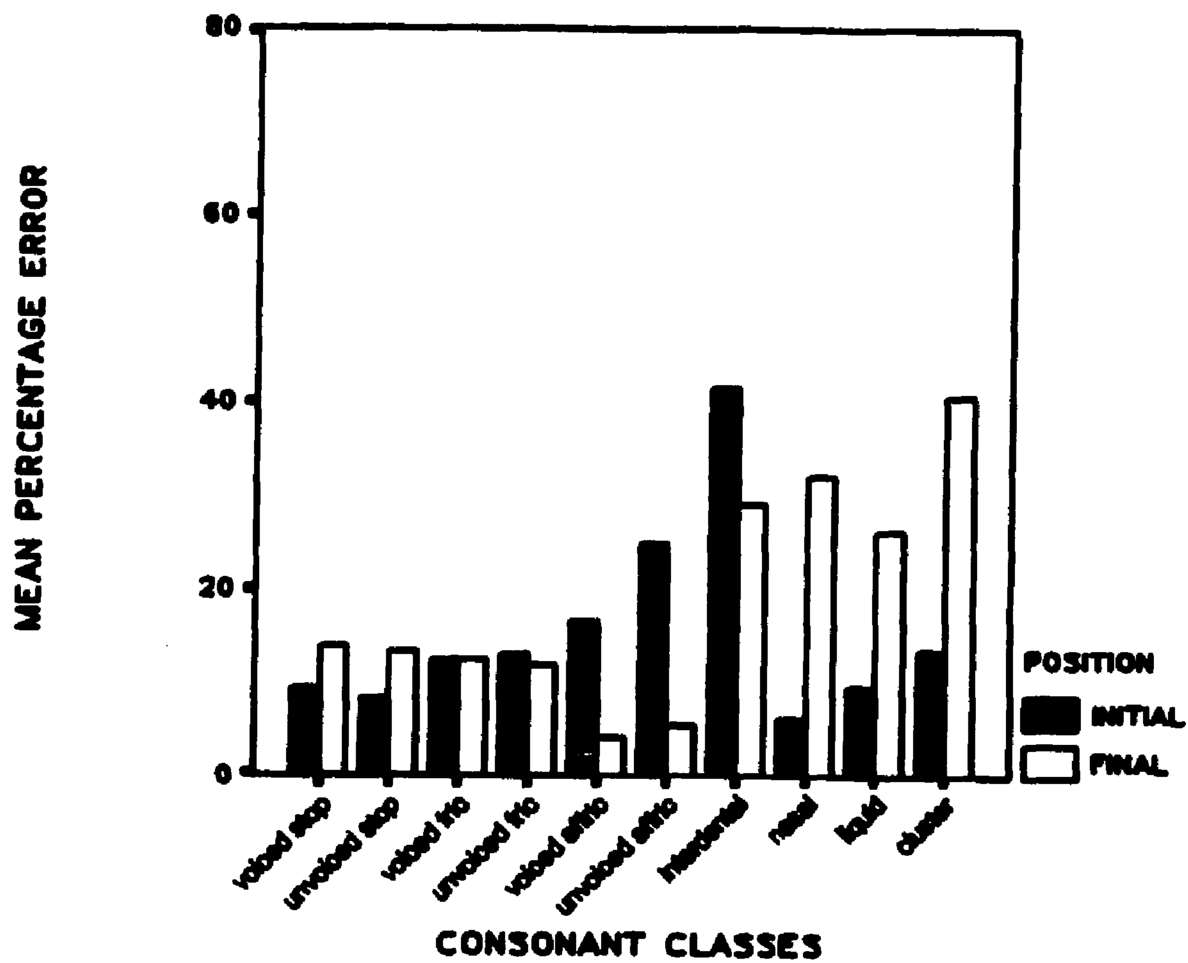


Figure 4.2

Percentage of Error as a Function of Consonant Class in Initial and Final Position for Hebrew Talkers



The non-significant main effect for language limits the interest and generality of the differences between the Japanese and Hebrew populations as a whole. However, the Hebrew talkers in this study did tend to have more intelligible consonants than the Japanese talkers. The mean intelligibility percentage for the Japanese talkers was 75.5, compared with 82.3 for the Hebrew talkers. For the Hebrew talkers, the variability across consonant classes is roughly equivalent for initial and final position. The phonetic classes with positional differences greater than 15% are clusters, unvoiced affricates, and nasals. For clusters and nasals, final position is less intelligible, while initial unvoiced affricates were more difficult than final. It is once again clear that the differences in intelligibility for Hebrew talkers cannot be accounted for in terms of position within the token. The two classes with the lowest intelligibility scores (more than 40% unintelligibility) were final clusters and initial interdentals. Since interdentals do not occur in Hebrew, and final clusters have more limited distribution than English clusters, these findings are consistent with an interference model. Hebrew speakers, like the Japanese speakers, did in fact have greater difficulty producing sounds without equivalents in the their native inventory.

A set of four independent t-tests, a subset of the nine permissible comparisons between consonant classes in each language, confirm the findings depicted in the graphs. The comparisons were chosen (on the assumption that the intelligibility, ease of articulation, and relative frequency of a consonant class are closely related) to reflect the relative difficulty of consonant classes, as predicted by the phonetic in-

ventories of English, Hebrew, and Japanese, and by general developmental and cross-linguistic data. Broadly, these considerations predict that clusters will be more difficult for a speaker (and therefore less intelligible for a listener) than their single consonant components; fricatives than stops; voiceless than voiced in initial position; voiced than voiceless in final position; and final than initial, consonants (Locke, 1983). A subset of these predictions is confirmed by the analysis.

For the Japanese talkers (Table 4.7), stops were significantly better than fricatives in initial position only. Voiceless consonants were significantly more intelligible than voiced consonants in final position, contrary to predictions noted above. Final clusters and liquids were significantly worse than all other consonant classes.

Several of these results fit within an interference model. The increased intelligibility of stops over fricatives is predictable, since stops /p/,/b/,/d/,/t/,/g/, and /k/ occur in Japanese, while fricatives are phonetically and phonologically more contrastive.[12] Consonant clusters do not occur in Japanese in any environment, and consonants never occur syllable finally; final clusters therefore incorporate both types of contrastiveness. Similarly, Japanese liquids differ markedly from American English liquids: a retroflex /r/ does not exist, and /l/ precedes only certain vowels; Japanese liquids also do not appear in syllable final position, so that final liquids

---

[12] In Japanese, initial fricatives are subject to several phonological constraints (see Table 3.1), unlike stops. Syllable final consonants do not occur at all in Japanese, making final stops and fricatives equally contrastive.

Table 4.7

Planned Comparisons Between Consonant Classes in Initial and Final Positions for Japanese Talkers

	Initial Position	Final Position
clusters vs. stops fricatives nasals liquids	NS	p<.05
(Final clusters: significantly worse than other classes)[13]		
liquids vs. stops fricatives nasals	NS	p<.05
(Final liquids: significantly worse than other classes)		
fricatives vs. stops	p<.05	NS
(Initial fricatives: significantly worse than stops)		
voiceless consonants vs. voiced consonants	NS	p<.01
(Final voiceless consonants: significantly better than voiced)		

[13] In this and all subsequent tables, NS = Not significant.

Table 4.8

Planned Comparisons Between Consonant Classes in Initial and Final Positions for Hebrew Talkers

	Initial Position	Final Position
clusters vs. stops fricatives nasals liquids	NS	$p<.01$
(Final clusters: significantly worse than other classes)		
liquids vs. stops fricatives nasals	NS	NS
fricatives vs. stops	NS	NS
voiceless consonants vs. voiced consonants	NS	NS

again present a double contrast for Japanese talkers. Thus, patterns of intelligibility loss are often predictable from the contrastiveness classifications.

Results for the Hebrew talkers appear in Table 4.8. Final clusters were significantly less intelligible than single consonants; this was the only prediction yielding a significant result. The poor intelligibility of final clusters compared to all other final consonant classes is predictable, since many of the clusters permissible in English do not occur in Hebrew.

4.1.2 The Effect of Vowel Class on Intelligibility

Vowels were classified with respect to their height and location in the front-back dimension. Table 4.9 presents a summary of error percentages for vowel classes for English, Japanese and Hebrew talkers. The English vowels were submitted to a repeated measures analysis of variance to determine whether some vowel classes were significantly less intelligible than others. Results were not significant. This argues against correcting for error by subtracting English talkers' scores from the non-native talkers' scores, since differences among the six vowel classes reflect random variability. Nonetheless, corrected scores are presented for comparison in Table 4.10, and rank ordering shifts resulting from the correction in Table 4.11. The rank orders remain nearly identical for both the Japanese and Hebrew speaking groups. The low front vowel was mistranscribed more than all other front vowels for both non-native language groups. The Japanese talkers produced less accurate back vowels than

Table 4.9

Percent Error on Vowel Classes for the Three Language Groups

Vowel Class	English	Japanese	Hebrew
high front	14.1	30.13	46.15
mid front	9.03	16.67	22.22
low front	8.33	51.39	72.22
high back	20.83	27.78	44.44
mid back	13.89	51.11	48.89
low back	20.83	54.17	56.25
Weighted Mean Percent Error For Vowels:	13.5	36.6	45.4

Table 4.10

Percent Error for Vowel Classes for Japanese and Hebrew Talkers, Corrected against English Scores

Vowel Class	Japanese	Hebrew
high front	16.03	32.05
mid front	7.64	13.19
low front	43.06	63.89
high back	6.95	23.61
mid back	37.22	35.0
low back	33.34	31.55
Weighted Mean Percent Error For Vowels: (Corrected)	22.85	31.55

Table 4.11

Rank Ordering of Vowels, High to Low Intelligibility for Raw and Corrected Error Scores of Japanese and Hebrew Speakers

Japanese

Raw Scores	Corrected Scores
mid front (16.67)	high back (6.95)
high back (27.78)	mid front (7.64)
high front (30.13)	high front (16.03)
mid back (51.11)	low back (33.34)
low front (51.39)	mid back (37.22)
low back (54.17)	low front (43.06)

Hebrew

Raw Scores	Corrected Scores
mid front (22.22)	mid front (13.19)
high back (44.44)	high back (23.61)
high front (46.15)	low back (31.55)
mid back (48.89)	high front (32.05)
low back (56.25)	mid back (35.0)
low front (72.22)	low front (63.89)

front vowels. Both the Japanese and Hebrew talkers produced less intelligible low vowels than high vowels.

Results of a repeated measures analysis of variance with language as the between-subjects variable and vowel type as the within-subjects variable indicate no significant differences between the Japanese and Hebrew speaking groups. The differences among vowel types in terms of accuracy of transcription, however, were highly significant, $p<.001$. Figure 4.3 displays intelligibility percentages of the various vowel types for the Hebrew and Japanese talkers. Tables 4.12 and 4.13 present the results of planned comparisons between vowel groups for the Japanese and Hebrew talkers, respectively. For the Japanese speaking group, glides[14] were significantly more intelligible than vowels; front than back vowels, and high and mid back than low back. For Hebrew talkers, significant differences were found only between glides and vowels. The least intelligible vowel groups for both language groups are the low front vowel, the low back vowel, and the mid back vowels. For the Hebrew speaking group, the high front and high back vowels rank next in terms of difficulty, with similar loss of intelligibility, and mid front vowels presenting the fewest problems. The Japanese talkers also demonstrate similar loss of intelligibility for high front and high back vowels. The Japanese talkers, like the Hebrew talkers, produced highly intelligible samples of the mid front set of vowels.

---

[14] Glides were included in the vowel analysis because they are intermediate between consonants and vowels in terms of their acoustic characteristics. In addition, they were produced in this study in only one position (initial) just as the vowels were produced in only one position (medial). Since the consonant analysis took position within the token into account, it seemed more appropriate to include glides within the vowel analysis.

Figure 4.3

Percent Intelligibility of Vowel Classes for Japanese and Hebrew Talkers

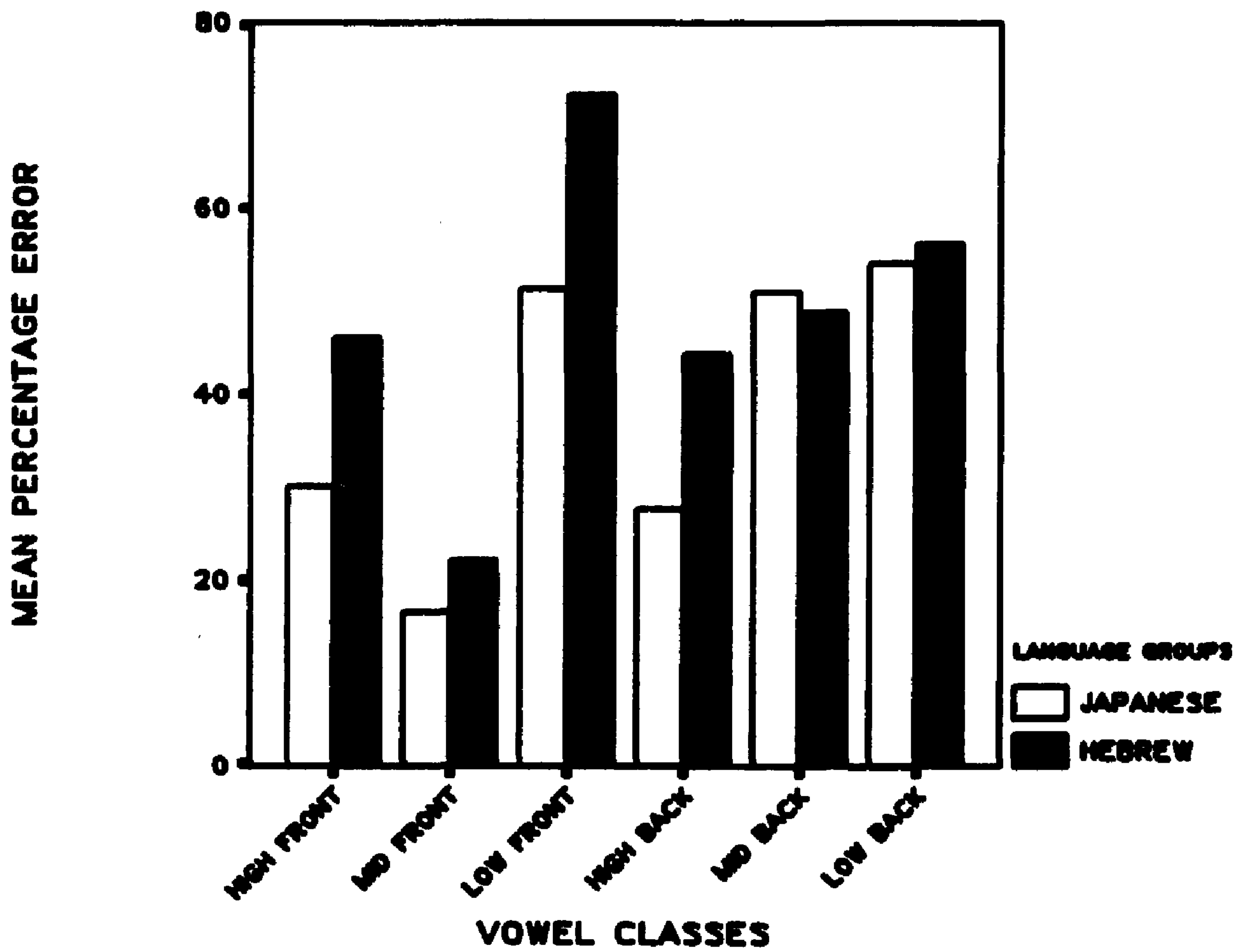


Table 4.12

Planned Comparisons Between Vowel Classes for Japanese Talkers

Comparison	Result
glides vs. vowels	$p<.05$ (glides: better)
front vowels vs. back vowels	$p<.05$ (front vowels: better)
high and mid front vs. low front	NS
mid front vs. high front	NS
mid and high back vs. low back	$p<.05$ (high and mid back: better)
high back vs. mid back	NS

Table 4.13

Planned Comparisons Between Vowel Classes for Hebrew Talkers

Comparison	Result
glides vs. vowels	$p<.01$ (glides: better)
front vowels vs. back vowels	NS
mid and high front vs. low front	NS
mid front vs. high front	NS
mid and high back vs. low back	NS
high back vs. mid back	NS

4.1.3 The Intelligibility of Consonants vs. Vowels

The literature on foreign accents describes vowels as more difficult to produce, and therefore more accented, than consonants. The present study examines whether vowels were less intelligible than consonants; this would also suggest greater difficulty in production. Table 4.14 lists the percentages of mistranscribed vowels and consonants by talker, and the mean percent error for each group. For all talkers, there was a higher percentage of error for vowels than for consonants. These differences are highly significant for all the Hebrew talkers, with Pearson chi squares, $p<.001$. Two of the four Japanese talkers also produced significant differences, $p<.05$. Thus, the disparity between consonant and vowel production is greater for Hebrew than for Japanese speakers.

4.1.4 Specific Segmental Confusions

Confusion matrices were used to determine the segments most frequently mistranscribed, and what they were perceived to be. English talkers also produced sounds that were mistranscribed, though they had much higher intelligibility scores than the Hebrew or Japanese talkers. The high error segments of English talkers provide a frame of reference for evaluating errors of non-native talkers. Clearly native English talkers produce native-sounding targets. High error rates for particular sounds may indicate that those sounds have a wide range of acceptable productions,

Table 4.14

Percent Error: Vowels vs. Consonants, by talker
(Significance of differences: based on Pearson Chi-square)

	Japanese Talkers		Hebrew Talkers	
Consonants	J1	24.1	H1	9.2
Vowels		35.7		36.9
		NS		p<.001
Consonants	J2	37.9	H2	14.1
Vowels		44.6		38.1
		NS		p<.001
Consonants	J3	10.6	H3	20.9
Vowels		26.2		56
		p<.01		p<.001
Consonants	J4	25.5	H4	26.6
Vowels		39.9		50.6
		p<.05		p<.001

so that linguistic context at a higher level than a single nonsense word is necessary to disambiguate them. Tables 4.15 - 4.20 present percent error by segment for native and non-native talkers. Frequently substituted sounds accounting for at least 30% of the error are noted.

The English talkers elicited a high percentage of transcription errors on vowel targets. The segments with particularly large amounts of error were / ɑ/ (20.83%), /ʌ/ (22.22%), and / ʊ/ (36.11%). Among the consonants, there were high percentages of error on /th/ (28.12%), /rks/ (33.33%) and /ldz/ (33.33%) The English talkers did worse than the Japanese talkers and no better than Hebrew talkers in producing /e/. The vowel /e/ has phonemic equivalents in Japanese and Hebrew.

The most problematic vowels for the Japanese talkers are / æ/, / ɑ/,/ ɔ/ and /ʌ/ (>50% error). The only vowels with less than a 25% error rate were /i/,/e/,/ɛ/ and /u/. For the Hebrew talkers, the most error prone vowels were /i/,/ æ/,/ ɑ/,and / ɔ/ with more than 50% error. The only vowel not mistranscribed more than 20% of the time was /e/. The Hebrew talkers had as many or more errors than the Japanese talkers on all vowels but / ɪ/ and /ʌ/.

When individual vowel error rates are examined, the value of classifying vowels into classes becomes questionable. For example, the class of high back vowels was highly intelligible for Japanese talkers (72.22%). In fact, /u/ was the most

Table 4.15

Percent Transcription Errors for Vowels: English Talkers

Vowel	Percent Error	Most Frequent Substitutions[15]
/i/	10.0	/ɪ/, /ɛ/
/ɪ/	16.67	/ɛ/
/e/	16.67	/i/,/ɪ/
/ɛ/	6.48	/æ/
/æ/	8.33	/ɛ/
/ʌ/	22.22	/ʌ/
/ɑ/	20.83	/ɔ/
/ɔ/	16.67	/ɑ/
/o/	1.67	*
/ʊ/	36.11	/ʌ/
/u/	5.56	*

[15] Frequent substitutions are operationally defined as substitutions accounting for 30% or more of the errors; *=Diffuse substitutions, with no single substitution accounting for 30% error.

Table 4.16

Percent Transcription Errors for Vowels: Japanese Talkers

Vowel	Percent Error	Most Frequent Substitutions[16]
/i/	15.0	/ɪ/
/ɪ/	37.5	/i/
/e/	2.78	/æ/
/ɛ/	21.30	/æ/
/æ/	51.39	/ɑ/
/ʌ/	61.11	/ɑ/
/ɑ/	54.17	/ɔ/
/ɔ/	58.33	/o/
/o/	33.33	/ɔ/
/ʊ/	44.44	*
/u/	11.11	/ʊ/

16 Frequent substitutions are operationally defined as substitutions accounting for 30% or more of the errors; *=Diffuse substitutions, with no single substitution accounting for 30% error.

Table 4.17

Percent Transcription Errors for Vowels: Hebrew Talkers

Vowel	Percent Error	Most Frequent Substitutions[17]
/i/	56.67	/ɪ/
/ɪ/	39.58	/i/
/e/	16.67	/ɪ/, /ɛ/
/ɛ/	24.08	/ɪ/
/æ/	72.22	/ɛ/
/ʌ/	36.11	/æ/, /ɑ/
/ɑ/	56.25	/ʌ/
/ɔ/	83.33	/o/
/o/	36.67	/ɔ/
/ʊ/	47.22	/u/
/u/	41.67	/ʊ/

[17] Frequent substitutions are operationally defined as substitutions accounting for 30% or more of the errors; * = Diffuse substitutions, with no single substitution accounting for 30% error.

Table 4.18

Percent Transcription Errors for Consonants: English Talkers

Consonant	Percent Error	Most Frequent Substitutions[18]
/b/	8.33	*
/p/	0	
/d/	9.73	/th/
/t/	0	
/g/	6.25	
/k/	2.08	
/z/	8.33	/s/
/s/	1.67	
/ʃ/	1.39	
/h/	0	
/v/	2.78	
/f/	12.5	/th/
/th/	28.12	/f/
/m/	11.67	/n/
/n/	6.25	
/ŋ/	12.5	*
/tʃ/	3.57	
/dʒ/	2.08	
/l/	7.5	*
/r/	2.5	
/st/	8.33	*
/fr/	0	
/gr/	0	
/pl/	0	
/br/	0	
/ts/	8.33	*
/sk/	0	
/nd/	16.67	*
/lt/	8.33	*
/ft/	0	
/str/	8.33	*
/rdz/	0	
/rks/	33.33	*

[18] Most frequent substitutions are presented only for consonants with error percentages greater than 7.18, which was the mean error percentage for all English consonants by English talkers. Substitutions accounting for more than 30% error are noted; the symbol * indicates that no one substitution accounted for 30% error.

Consonant	Percent Error	Most Frequent Substitutions
/ldz/	33.33	*
/w/	0	
/j/	10.42	*

Table 4.19

Percent Transcription Errors for Consonants: Japanese Talkers

Consonant	Percent Error	Most Frequent Substitutions[19]
/b/	29.17	/p/
/p/	8.33	
/d/	4.17	
/t/	4.17	
/g/	18.75	
/k/	4.17	
/z/	45.0	*
/s/	5.0	
/ʃ/	18.06	
/h/	22.92	
/v/	66.66	/b/
/f/	19.44	
/th/	55.21	
/m/	1.67	
/n/	18.75	
/ŋ/	29.17	/n/
/tʃ/	19.05	
/dʒ/	22.92	
/l/	30.83	/r/
/r/	30.0	/l/
/st/	0	
/fr/	50	*
/gr/	16.67	
/pl/	25.0	
/br/	41.67	*
/ts/	33.33	*
/sk/	0	
/nd/	50.0	/n/
/lt/	58.33	*
/ft/	0	
/str/	33.33	*
/rdz/	50.0	*
/rks/	75.0	*

[19] Most frequent substitutions are presented only for consonants with error percentages greater than 24.53, which was the mean error percentage for all English consonants by Japanese talkers. Substitutions accounting for more than 30% error are noted; the symbol * indicates that no one substitution accounted for 30% error.

Consonant	Percent Error	Most Frequent Substitutions
/ldz/	83.33	*
/w/	27.08	*
/j/	25.0	deleted

Table 4.20

Percent Transcription Errors for Consonants: Hebrew Talkers

Consonant	Percent Error	Most Frequent Substitutions[20]
/b/	16.67	
/p/	16.67	
/d/	8.33	
/t/	15.28	
/g/	10.42	
/k/	10.42	
/z/	20.0	/s/
/s/	10.0	
/ʃ/	2.78	
/h/	18.75	/p/,/k/
/v/	0.0	
/f/	19.44	*
/th/	35.42	*
/m/	13.33	
/n/	8.33	
/ŋ/	50.0	*
/tʃ/	16.67	
/dʒ/	10.42	
/l/	10.0	
/r/	25.83	*
/st/	4.17	
/fr/	0	
/gr/	33.33	/g/
/pl/	25.0	*
/br/	0	
/ts/	16.67	
/sk/	16.67	
/nd/	25.0	*
/lt/	50.0	*
/ft/	33.33	*
/str/	13.33	
/rdz/	83.33	*
/rks/	66.67	*

[20] Most frequent substitutions are presented only for consonants with error percentages greater than 17.75, which was the mean error percentage for all English consonants by Hebrew talkers. Substitutions accounting for more than 30% error are noted; the symbol * indicates that no one substitution accounted for 30% error.

Consonant	Percent Error	Most Frequent Substitutions
/ldz/	75.0	*
/w/	4.17	
/j/	14.58	

intelligible vowel (88.89%), while / ʊ/ ranked much lower (55.56%). Collapsing across the vowel class obscures the high intelligibility of /u/, while the less intelligible / ʊ/ is drawn up.

The highest errors among consonants for the Japanese talkers are for /b/, /z/, /v/, /th/,[21] / ŋ/,/l/,/r/, /fr/,/pl/,/br/,/ts/,/nd/,/lt/,/str/,/rdz/,/rks/,and /ldz/ (>25% error). The Hebrew talkers had an error rate of more than 25% for /th/, / ŋ/,/r/,/gr/,/lt/,/ft/, /rdz/,/ldz/ and /rks/. In addition to having fewer mistranscribed segment types, the Hebrew talkers generally had lower percentages of error than the Japanese talkers.

4.1.5 The Effect of Contrastiveness on Token Intelligibility

A strict interference model predicts that L2 learners will have difficulty with segments dissimilar to sounds in their native languages. Phonetically contrastive, phonologically contrastive, or non- contrastive segments were classified in terms of intelligibility. Intelligibility percentages as a function of contrastive status for all Japanese and Hebrew talkers are presented in Table 4.21. For all talkers in both language groups, as well as the mean across the group, phonetically contrastive sounds present the highest percentage of unintelligibility. The phonologically contrastive set does not appear to differ markedly from non-contrastive sounds;

---

[21] "/th/" incorporates / θ/ and / ð/, but is referred to as a single phoneme since listeners' transcriptions did not reliably discriminate the voiced and unvoiced tokens.

Table 4.21

Percent Target Segments Mistranscribed by Contrastiveness Category

	Japanese Speakers	Hebrew Speakers
phonetically contrastive	39.54	39.06
phonologically contrastive	21.79	15.0
non-contrastive	18.44	19.17

indeed, for some talkers, there is more deviance on non-contrastive sounds. Thus, phonological contrastiveness is not a good predictor of regions of difficulty. For the Japanese talkers, 39.54% of the phonetically contrastive segments were mistranscribed, compared with only 21.79% of the phonologically contrastive group and 18.44% of the non-contrastive segments. The Hebrew talkers also performed worst on the phonetically contrastive tokens, with 39.06% of the segments in this category being mistranscribed. Only 15.0% of the phonologically contrastive group were mistranscribed, against 19.17% of the non-contrastive group.[22] The value of predicting error based on phonetic contrastiveness seems to be supported. A repeated measures two-factor analysis of variance was done to assess differences between Hebrew and Japanese talkers in terms of intelligibility for phonetically contrastive, phonologically contrastive, and non-contrastive segments. No difference was found between the two language groups, though the differences among the three contrastiveness categories proved highly significant, $p<.01$.

"Phonetically contrastive" tokens include vowels, single consonants, and clusters that do not occur in Japanese or Hebrew. As a group, they were less intelligible than segments classified as "non-contrastive." Are these classifications internally homogeneous, or does some subset of the group account for most of the unintelligibility? Tables 4.22 and 4.23 show the loss of intelligibility for contrastive vs. non-contrastive consonants and vowels for Japanese and Hebrew speakers. The

[22] It should be noted that the phonologically contrastive group for Hebrew contained only 40 tokens; that is, 10 tokens per talker.

Table 4.22

Intelligibility of Phonetically Contrastive and Non-contrastive Segments for Japanese Talkers

	Phonetically Contrastive	Non-Phonetically Contrastive
Consonants:		
	>30% Error	>30% Error
	/v/,/th/,/r/,/fr/, /br/,/ts/,/nd/,/lt/ /str/,/rdz/,/rks/,/ldz/	/z/,/l/
	<30% Error	<30% Error
	/f/,/st/,/gr/,/pl/, /sk/, /ft/,/w/,/j/	/b/,/p/,/d/,/t/, /g/,/k/,/s/,/ʃ/, /h/,/m/,/n/,/ ŋ/, /tʃ/,/dʒ/
Vowels:		
	>40% Error	>40% Error
	/ æ/,/ʌ/,/ ɔ/,/ ʊ/	/ ɑ/
	<40% Error	<40% Error
	/ ɪ/,/ɛ/	/i/,/e/,/o/,/u/

Table 4.23

Intelligibility of Phonetically Contrastive and Non-contrastive Segments for Hebrew Talkers

	Phonetically Contrastive	Non-Phonetically Contrastive
Consonants:		
	>25% Error	>25% Error
	/th/,/r/,/rdz/,/rks/, /ldz/	/ŋ/,/gr/,/pl/,/nd/, /lt/,/ft/
	<25% Error	<25% Error
	/str/,/w/	/b/,/p/,/d/,/t/, /g/,/k/,/z/,/s/, /ʃ/,/h/,/v/,/f/, /m/,/n/,/tʃ/,/dʒ/, /l/,/st/,/fr/,/br/, /ts/,/sk/,/j/
Vowels:		
	>40% Error	>40% Error
	/æ/,/ɔ/,/ʊ/	/ɑ/,/i/,/u/
	<40% Error	<40% Error
	/ɪ/,/ɛ/,/ʌ/	/o/,/e/

phonetically contrastive consonants and vowels appear reasonably homogeneous in accounting for loss of intelligibility for the Japanese talkers. Specifically, contrastive consonants /v/, /th/, /r/, and consonant clusters were frequently misheard, as were the contrastive vowels / æ/, /ʌ/, / ɔ/, and / ʊ/.

Vowels were submitted to additional chi square analyses on individual subjects to determine whether contrastive status correlated with vowel intelligibility. The contrastive vowels for both Japanese and Hebrew speakers were / ɪ/, /ɛ/, / æ/, /ʌ/, / ɔ/, and / ʊ/. The intelligibility of these vowels was compared with the non-contrastive set, /i/, /e/, / ɑ/, /o/, and /u/, when produced by English talkers, where the distinction based on contrastive status is irrelevant. No significant differences in intelligibility were found for any of the English speakers. Contrastive vowels were significantly less intelligible than non-contrastive vowels for two of the four Japanese speakers ($p<.05$), and one of the four Hebrew speakers ($p<.05$).

The overall loss of intelligibility on vowels produced by English talkers was 13.5%. The "non-contrastive" set was misproduced 10.4% of the time; the "contrastive" set, 15.5% . The overall loss of intelligibility on vowels produced by Japanese speakers was 36.6%. The contrastive vowels were mistranscribed 43.1% of the time, compared with a 25% error rate for non-contrastive vowels. The Hebrew talkers' vowels were mistranscribed 45.4% of the time. The contrastive set was misheard 46.1% of the time; the non-contrastive set, 43.3%. Thus, only

Japanese talkers exhibited marked differences in production for the contrastive and non-contrastive sets of vowels.

Unlike vowels, consonants classified as "phonetically contrastive" differ for the Japanese and Hebrew groups. The phonetically contrastive and non-contrastive consonants were submitted to chi square analyses on individual subjects to determine whether contrastive status of consonants correlated with intelligibility. A comparable analysis was not performed on clusters, since all clusters were phonetically contrastive for Japanese talkers, and only the three-consonant clusters were contrastive for the Hebrew talkers. Phonologically contrastive consonants were collapsed with non-contrastive segments in this analysis, since they appeared to be equally intelligible. Contrastive consonants were significantly worse than non-contrastive consonants for only one Japanese talker ($p<.001$), though all four Japanese talkers had higher unintelligibility percentages for contrastive segments. Three of the four Hebrew speakers had significantly worse intelligibility for the contrastive segments ($p<.05$); the fourth speaker also had a higher error percentage for the contrastive consonants, though results did not reach significance. The Japanese talkers had an overall error rate for consonants of 24.5%. Of the total number of contrastive consonants attempted, 27.42% were mistranscribed, compared with 9.66% of the non-contrastive consonants. Of the contrastive consonants, 24.36% were transcribed incorrectly, compared with 10.48% of the non-contrastive set.

4.1.6 The Role of Contrastiveness in Accentedness Ratings

The effect of particular segments on accentedness ratings is difficult to assess from the data gathered. Since segments were always embedded in words or syllables, listeners' judgments of accent cannot be ascribed to any one particular sound. However, we can evaluate the relation between the presence of contrastive or non-contrastive segments in a word and perception of accent by, classifying syllables according to the contrastiveness status of their component consonants and vowels. Pearson chi squares were calculated for the data from each Hebrew and Japanese speaker to compare differences in accentedness ratings for the single syllable tokens with: (1) no contrastive segments; (2) contrastive vowels but non-contrastive consonants; (3) one or both contrastive consonants, but non-contrastive vowels; (4) contrastive vowels and consonants. No significant differences were found for any of the Japanese or Hebrew talkers. Thus, the presence of phonetically contrastive segments did not affect accentedness ratings, though overall intelligibility was impaired.

4.1.7 The Relation Between Accentedness and Intelligibility

The accentedness rating scores were reduced to a two-point scale in order to determine the relation between accentedness scores and intelligibility. The intelligibility of the whole word, not the intelligibility of particular target segments, was used for this correlation. Mean accentedness scores of the three listeners who heard any

given token varied between one (all three listeners agreeing that the talker sounded like a native English talker) and five (the three listeners agreeing that the talker was not attempting to produce an English target). Mean scores of 2.50 or less were grouped together (good accent) as were scores between 2.51 and 5 (poor accents). This procedure was used to ensure a sufficient number of scores per cell to perform the chi square analysis. Intelligibility scores varied between 3 (if all listeners perceived the whole item correctly) and 6 (if all listeners perceived the item incorrectly). A token with an intelligibility score of 3 or 4 was classified as intelligible; a token with a score of 5 or 6 as unintelligible. 20.4% of the utterances produced by English talkers were rated as having "poor accents," 46.6% of these were also mistranscribed. 64.9% of the utterances produced by Japanese speakers were rated as poor; of these, 74.3% were also misheard. Utterances of Hebrew talkers were rated "poor" 77.9% of the time; 81.2% were also mistranscribed.

Pearson chi squares were calculated for each speaker to determine the relation between accentedness and intelligibility. All four English speakers were given significantly worse accentedness ratings for mistranscribed tokens ($p<.05$). Two of the four Hebrew talkers, and two of the four Japanese talkers, also had significantly worse accentedness ratings for mistranscribed tokens. For both the Hebrew and Japanese groups, talkers with the worst intelligibility and the most ratings of "poor accent" showed no correlation between accentedness and intelligibility.

4.1.8 Accentedness Ratings for Single Words vs. Sentences

Listeners had two tasks: to transcribe single-word tokens and to rate each token for degree of accentedness. Listeners were therefore basing their judgments of accentedness on single word information only. An obvious concern is whether listeners' ratings of single-word accentedness correlated with their ratings of accent in the more naturalistic full sentence task, particularly since the primary locus of a foreign accent is widely considered to be sentence intonation, yet single words provide no intonational information beyond the word contour. Mean accentedness scores for each talker on the 187 single words were correlated against mean accentedness scores for each talker on the 12 sentences. The correlation coefficient between accentedness scores on the single word task and scores on the naturalistic task was .95, yielding a coefficient of determination of .90. Therefore, there was a close correspondence between listeners' judgments on the two tasks.

4.1.9 Real words vs. Nonsense syllables

Nonsense syllables have distinct advantages in perceptual experiments: they can be phonetically balanced more easily than real words, sounds can be presented in multiple contexts, and single segments can be varied to create minimal pairs. However, listeners' responses may not be the same for nonsense and real words. If listeners were, in fact, listening in some purely phonetic mode, there should be no

significant differences between segments embedded in real words and segments embedded in nonsense syllables. However, significant differences in intelligibility scores[23] were found between real words and nonsense syllables for all of the English speakers, and for three of the four Japanese speakers. The English speakers as a group were mistranscribed 11.6% of the time for real words, and 31.8% of the time for nonsense syllables. Japanese speakers were mistranscribed 42.6% of the time for real words, compared with 63.8% for nonsense syllables. The relation between word status and intelligibility was not significant for three of the four Hebrew speakers. The Hebrew speakers as a group had 54.2% mistranscriptions for real words, compared with 62.4% mistranscriptions for nonsense syllables. For English and Japanese talkers only, listeners' judgments were either biased by whether they were or were not able to identify the target word, or talkers' pronunciations of nonsense words were less phonetically reliable than their pronunciations of real words.

Nonsense syllables were rated as more accented than real words for English talkers; 28.2% of the nonsense syllables were rated as poor, compared with 14.2% of the real words. For Japanese talkers, 68.9% of the nonsense tokens were rated as poor, compared with 61.8% of the real words.

There appears to be no relation between word status and accent for Hebrew talkers. The Hebrew speakers as a group had 78.2% poor ratings for real words,

---

[23] "Intelligibility scores" for this analysis refer to an error on any segment within the token, not only the target segment evaluated in most other intelligibility analyses described.

compared with 77.7% poor ratings for nonsense syllables. The Hebrew talkers also received considerably more accent ratings of "poor" and "very poor " than the Japanese talkers. This suggests that decisions about accent were made in a more absolute fashion for the Hebrew talkers than for the Japanese and English talkers. For the Japanese and English speakers, inability to identify the word as an acceptable English token influenced the judgment of accent; for the Hebrew it did not.

4.1.10 Summary of Perceptual Results

Statistical tests on the listeners' transcriptions of the Japanese, Hebrew, and English speakers suggest the following:

1. Certain consonant classes were significantly less intelligible than others. Specifically, Japanese speakers had less intelligible clusters and liquids than all other consonant classes in final position, and significantly less intelligible initial fricatives than stops. Results for Hebrew speakers were more erratic, but final consonant clusters were significantly less intelligible than all other consonant classes. Several of these findings conform to an interference model predicting difficulty for segments in L2 that have no equivalents in L1.

2. The overall effect of position in the syllable was not significant for either non-native group. Thus, consonants in final position were not significantly more difficult

for Japanese speakers, even though Japanese does not permit syllable-final consonants.

3. Certain vowel classes were significantly less intelligible than others. Both the Japanese and Hebrew speakers were least intelligible in their productions of low vowels / æ/ and / ɑ/, and mid vowel / ɔ/. Glides and mid front vowels were most intelligible for both groups.

4. Japanese speakers had a generally higher error rate for consonants than Hebrew speakers, while Hebrew speakers had less intelligible vowels than the Japanese.

5. Phonetically contrastive segments were mistranscribed approximately twice as often as non-contrastive segments for both Japanese and Hebrew speakers. Phonologically contrastive tokens were not mistranscribed more often than non-contrastive tokens.

6. Contrastive status did not affect accentedness ratings.

7. Ratings in the accentedness measure were correlated with intelligibility scores for all English talkers, two Japanese talkers, and two Hebrew talkers.

8. Listeners' judgments of accent on the single word and sentence tasks were highly correlated.

9. Both English and Japanese speakers were scored unintelligible more often for nonsense syllables than for real words. The differences were not significant for Hebrew speakers.

4.2 ACOUSTIC RESULTS

4.2.1 Overview

Vowels were selected for detailed acoustic analysis because they were the most perceptually confused phonetic class for all three language groups. First and second formant frequency values for the eleven English targets and five native Japanese or Hebrew targets were compared. Analyses of these vowels permit us to address several theoretical issues. First, what is the role of interference from the speaker's native system in accounting for acoustic deviation from the target? Second, do non-native speakers learn a new vowel set as a system of relations among the component vowels, or do they approach each vowel as a separate entity? Third, do speakers of a five-vowel language have more centralized vowels than native English speakers? Finally, do speakers tend either to centralize or to increase the acoustic range of vowels in L2 compared to the vowels of their native language?

4.2.2 Reliability and Validity of Acoustic Results

Most of the acoustic and perceptual results reflect mean values across the four speakers within each language group. A comparison of the mean formant values for the four English talkers of this study with those reported by Peterson and Barney (1952) based on 33 male talkers (Table 4.24) yields good agreement. Each of the Spearman rank order correlations between F1, F2 and (F2-F1) for the present study and Peterson and Barney is .98 ($p<.01$). The percentage differences in formant values on the two studies are 10% or less for seven of the nine vowels on F1 (/ ɔ/:16%; /ʌ/:20%) and less than 10% for seven of the nine vowels on F2 (/ ɔ/:15%, / ʊ/:16%) (Table 4.25). These values fall comfortably within the ranges of variability across speakers found by Peterson and Barney.

The results described above suggest that the small samples used for the three language groups in this study provide a tolerably good estimate of the mean vowel formant frequencies to be expected in the populations from which they were drawn. They also provide some validation of the formant measurement procedures described above (see section 3.4). Despite the diversity of syllabic contexts for the vowels of this study, the mean formant values closely approximate those observed by Peterson and Barney in fixed /h__d/ environment.

Table 4.24

Mean Formant Values for the Four English Speakers Compared with those for the 33 Speakers of Peterson and Barney (1952)

Vowel	F1		F2		F2-F1	
/i/	297	270	2351	2290	2054	2020
/ɪ/	420	390	1812	1990	1392	1600
/ɛ/	499	530	1710	1840	1211	1310
/æ/	656	660	1602	1720	947	1060
/ɑ/	703	730	1181	1090	478	360
/ɔ/	481	570	910	840	429	270
/ʊ/	444	440	1188	1020	744	580
/u/	311	300	944	870	633	570
/ʌ/	515	640	1188	1190	673	550

Table 4.25

Percentage Differences for English Speakers' Vowels (Percentaged against Peterson and Barney's Values)

	F1		F2	
Vowel	Absolute (Hz)	%	Absolute (Hz)	%
/i/	27	10	61	3
/ɪ/	30	8	-178	9
/ɛ/	-31	6	-131	7
/æ/	-4	1	-112	7
/ɑ/	-27	4	91	8
/ɔ/	-89	16	130	15
/ʊ/	4	1	168	16
/u/	11	4	74	9
/ʌ/	-125	20	-2	0

4.2.3 Statistical Comparison of F1 and F2 for English and Non-native Speakers: Contrastive vs. Non-contrastive Vowels

Both an interference model and a model involving natural simplifications predict greater difficulty for the vowels /ɪ/, /ɛ/, /ʌ/, /ʊ/, /ɔ/, and /æ/ (i.e., contrastive vowels) than for /i/, /e/, /ɑ/, /o/, and /u/ (i.e., non-contrastive vowels). The mean F1 value for each vowel across all Japanese speakers was subtracted from the mean F1 value for the corresponding vowel across all English speakers; F2 values for Japanese speakers were similarly subtracted from corresponding English F2 values. The procedure was repeated for Hebrew and English speakers. Differences derived reflect the distances of the non-native talkers from the mean English values with respect to F1 and F2. These differences were submitted to Mann-Whitney U-Tests to determine whether the distances for the non-contrastive set were significantly closer to the target English values than for the contrastive group. The results were not significant for F1 or F2 for either the Japanese or Hebrew speakers.

4.2.4 Vowel-by-Vowel Comparison of Formant Values for Native and Non-native Speakers

Since the statistical tests comparing F1 or F2 differences from English targets as a function of contrastive status were not significant, the mean values of each English vowel target produced by the Japanese and Hebrew speaking groups were examined vowel by vowel. Mean formant values for English vowels produced by English,

Japanese, and Hebrew talkers appear in Appendices B1-B3. Mean formant values for Japanese and Hebrew vowels appear in Appendices B4 and B5. Observations in the following discussion are based on the graphs shown in Figures 4.4 and 4.5. These include the F1 and (F2-F1) values for the English targets as produced by the native English talkers,[24] the mean formant values for the same targets produced by Japanese speakers, and the mean formant values for the five vowels of Japanese (Figure 4.4). Figure 4.5 displays the same data for the Hebrew speaking group.

In the following discussion on vowel shifts, "minimal deviation" will be operationally defined as formant differences of greater than 10% for F1 and for F2. Using this criterion, Japanese speakers' mean formant values differ minimally from English speakers for /i/, /ɛ/, / æ/, / ɔ/, /o/, and /u/. Hebrew speakers differ minimally for /ɛ/, /ʌ/, /o/, and /u/.

Interference can be manifest in several ways. In the clearest case, non-native speakers can substitute a native phone for an unfamiliar L2 target. Speakers may also produce a sound intermediate between a similar native sound and the actual target, suggesting that they are aware of a distinction, but that the native sound still exerts a pull. Alternatively, non-native speakers may be acutely aware of the distinction between their native vowel and an English target, and respond by "anti-interference" : overshooting the actual distance between native and target. No

---

[24] The value of F2-F1 was chosen over F2 because it models articulatory differences among vowels in the front-back dimension more accurately than F2 (Ladefoged, 1982).

Figure 4.4

English and Native Vowels Produced by Japanese and English Talkers

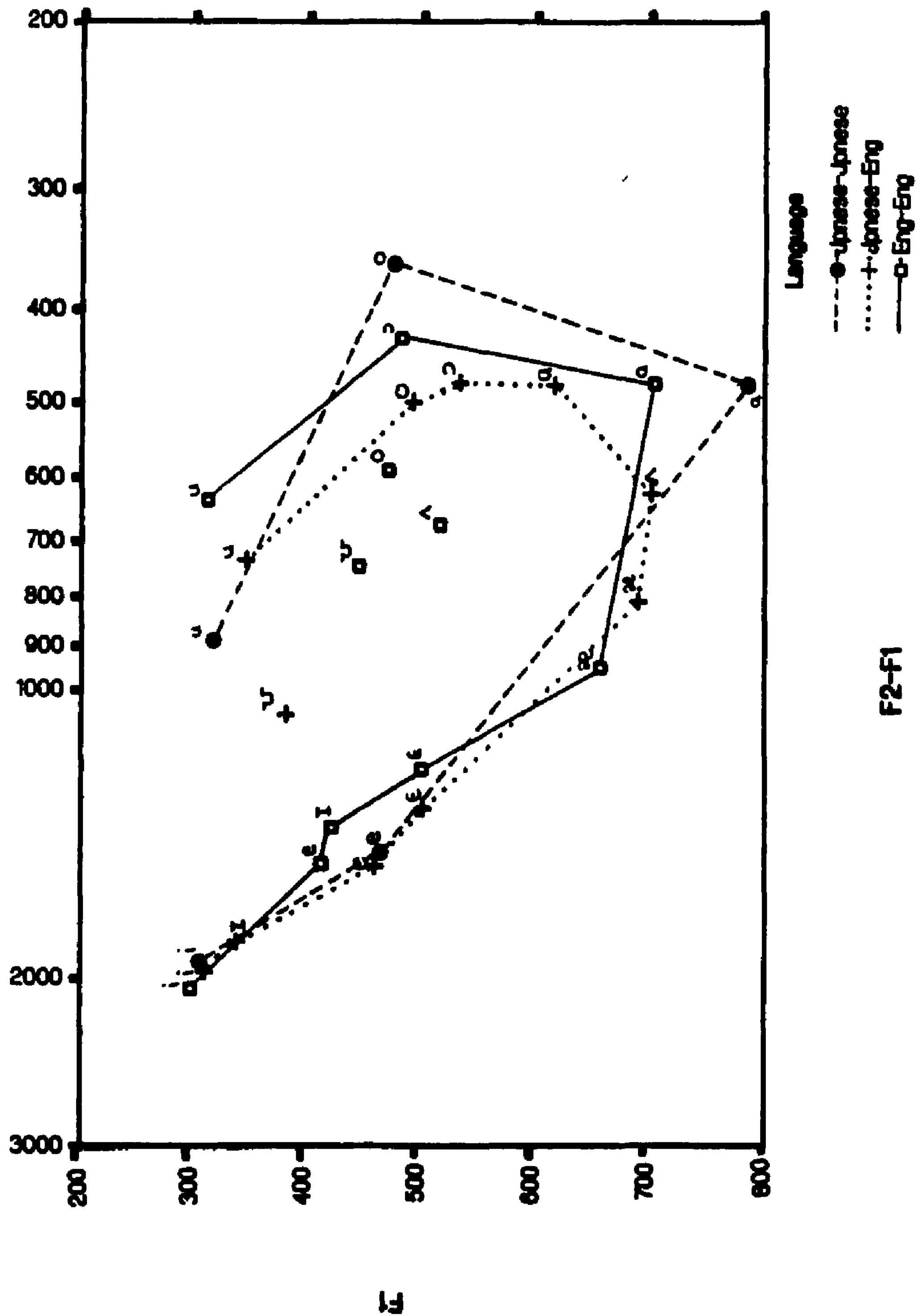


Figure 4.5

English and Native Vowels Produced by Hebrew and English Talkers

conclusions can be drawn in cases where a native vowel and English target are already acoustically close. Finally, no clear relation may exist, requiring some model other than interference to explain what is happening.

Consider, first, the front vowels for Japanese speakers. As shown graphically in Figure 4.4, Japanese formants for English /i/ are close to the /i/ values for native English speakers as well as for native Japanese. The speakers apparently had little difficulty with this familiar vowel. By contrast, although the unfamiliar / ɪ/ target is produced with higher F1 and lower (F2-F1) values than /i/, thus preserving the correct relation between /i/ and / ɪ/, interference appears to play a role: the unfamiliar new sound / ɪ/ is drawn towards the familiar native target /i/. The native Japanese /e/ occurs at a higher F1 value than the English /e/ target. Japanese talkers produce the English /e/ target at a point close to their native /e/, but slightly displaced backwards on the (F2 -F1) dimension. Native vowel placement seems to be interfering with accurate production of the new vowel. Japanese speakers may detect a difference only in the front-back dimension rather than in height. The English vowel /ɛ/ differs from /e/ by a higher F2 and lower (F2-F1) values. This distinction is maintained by the Japanese talkers, though the difference between the two vowels is less than for native English speakers. The front vowel / æ/ has no Japanese equivalent. Nonetheless, Japanese speakers speakers produce a front low vowel close to the native English production of / æ/.

The back set of English vowels produced by Japanese speakers deviates noticeably from the English targets with respect to (F2-F1). On the whole, the Japanese version of English back vowels is more centralized than the English speakers' productions. Each back vowel mean will be examined individually, in relation to English target vowels and the native language system.

Native Japanese / ɑ/ is produced with a higher F1 than any other vowel produced by either the Japanese or English speakers. Their rendition of / ɑ/ in English, however, is considerably higher than the target produced by English speakers. While the (F2-F1) dimension remains the same across the three mean / ɑ/ vowels in Fig. 4.4, F1 distances from native / ɑ/ to English target to Japanese production of / ɑ/ in English are roughly equal. It appears that Japanese speakers have recognized that the English vowel is slightly higher than their native / ɑ/, but in attempting to replicate the new form they "overshoot" the mark. By contrast, though the shift from / ɑ/ to / ɔ/ is in the appropriate direction in terms of F1, this time, the Japanese speakers undershoot the target, resulting in a lower / ɔ/ than the English mean / ɔ/ value. Native Japanese /o/ has a lower (F2-F1) value than any other vowel in Fig 4.4. The Japanese speakers approach the shift towards English /o/, but they traverse only part of the distance. English /o/ also has a lower F1, which Japanese speakers have not replicated. The Japanese value for / ʊ/ is entirely anomalous, compared to both the English target and any other native Japanese vowel. The centralized / ʊ/ they produce suggests that they have not associated English target / ʊ/ with any other vowel in their own system or English, and respond

by producing a centralized vowel not far from schwa. Finally, the Japanese /u/ matches F1 for English target /u/, but has a higher (F2-F1) value. Their English production of /u/ shifts appropriately from the Japanese /u/ along the (F2-F1) dimension, but inappropriately with respect to F1.

Figure 4.5 displays formant values for English targets produced by English and Hebrew speakers, as well as formants of native Hebrew vowels. Hebrew speakers are obviously producing an /i/ value in English similar to Hebrew /i/. Both vowels are produced with a higher F1 and lower (F2-F1) value than the English target. The Hebrew speakers mark the /i/-/ɪ/ distinction with a lower (F2-F1) value for /ɪ/, while English speakers couple this with a rise in F1 as well. Surprisingly, native Hebrew /e/ is closer to English target /e/ than the value the Hebrew speakers actually produce in English. The English target values for /e/ appear not to influence the Hebrew speakers at all in their production of /e/, and the shift from the native Hebrew /e/ values cannot be explained with respect to English. The absolute placement of /ɛ/ also appears unrelated to any vowel in English or native Hebrew. The vowel /æ/ has formant values closer to target English /ɛ/ than English /æ/. In addition, the orientation of /æ/ and /ɛ/ is skewed, with a difference in the (F2-F1) dimension rather than F1.

The low and back vowels produced by Hebrew speakers in English differ markedly from the native English set in terms of the overall use of the vowel space. The Hebrew talkers clearly centralize this set of English vowels, in a way not predictable

from their native vowel inventory. The results are puzzling with respect to an interference model. The / ɑ/ vowel in Hebrew, for example, is close to the English / ɑ/ in terms of F1. Instead of substituting their native equivalent, the Hebrew speakers produce a vowel considerably displaced in the F1 dimension. The vowel / ɔ/ is produced with considerably lower F1 and also centralized in a way that does not appear related to any other vowel in their native system or English. The remaining back vowels - /o/, / ʊ/, and /u/ - are produced in appropriate relation to one another. The Hebrew speakers' /o/ is produced closer to the English target /o/ than native Hebrew /o/; no explanation is evident for the marked lowering of F1. Unfamiliar / ʊ/ has been pulled towards the familiar anchor /u/; while attempts at English /u/ result in values identical to native Hebrew /u/ for F1, but overshot in the (F2-F1) domain.

4.2.5 Relation between Target Intelligibility and Formant Deviation

Rank order correlations were computed to determine the relation between intelligibility and deviation from the target. Intelligibility rankings were abstracted from results in Tables 4.9 and 4.10. Target deviation was calculated by subtracting mean F1 and F2 values for each vowel produced by the non-native groups from the mean F1 and F2 values when produced by native English speakers. In addition, a total formant deviation value (F1+F2 Deviation) derived by summing the absolute values of the differences in F1 and F2, was correlated against the intelligibility scores. No

significant correlations were found between F1, F2, or F1+F2 Deviation and degree of intelligibility for the vowel target.

4.2.6 Acoustic Variability of Vowels

Since the vowels measured in this study were extracted from multiple contexts, the variation in coarticulatory forces across the vowels is great. Variability across vowels was measured by calculating standard deviations of F1 and F2 for each vowel produced by each talker across the three contexts. It is generally hypothesized that the L2 learner's phonological system is more variable than that of the native speaker, which would result in larger standard deviations across repetitions. However, the L2 learners may coarticulate less than native speakers, in which case the vowels would appear more static across the three contexts. The standard deviations for F1 and F2 were submitted separately to a 2-factor analysis of variance to determine whether standard deviations differed as a function of language. No statistically significant differences were found across the three language groups. This suggests that the phonological systems of the non-native talkers in this study are as stable as the systems of native talkers. The variability of the contexts, however, resulted in large standard deviations for all talkers, which may mask variability differences across the three language groups.

4.2.7 Five Vowel Systems and Distances Among Vowels

Two predictions can be made about the distribution of vowels in a five-vowel system. One hypothesis is that the vowels will fall close to the peripheral point vowels that form the universal anchors of vowel systems. Alternatively, vowels may drift towards the center, since confusability of one vowel with another is less of a danger than in a more crowded vowel space. Comparing the vowel spaces of Japanese and Hebrew, both 5-vowel systems, may shed some light on the issue of whether the size of the vowel space is determined by the number of vowels in the system. Measuring vowel spaces of Japanese and Hebrew speakers in English will show whether speakers from 5-vowel systems expand the size of their native vowel space appropriately to accomodate the larger number of English vowels.

The ranges between highest and lowest values on F1 and F2 values for Japanese speakers are presented in Table 4.26. For Japanese speakers producing native vowels, the mean F1 range was 481 Hz, while the mean F2 range was 1394 Hz. Thus, large F1 and F2 distinctions occur even when relaxing these differences in Japanese would not result in a loss of intelligibility or distinctiveness across the vowels. When producing English vowels, the F1 range averaged across the Japanese talkers is 391 Hz, while for English talkers the range is 406 Hz. This indicates that the vowel space is roughly equivalent in the vowel height dimension. The front-back dimension is roughly approximated by F2. The mean F2 range for Japanese talkers

Table 4.26

Ranges of F1 and F2 Across All Vowels in Hz for the Three Language Groups

		English	Japanese	Hebrew
English	F1:	406	391	241
Vowels:	F2:	1441	1287	1118
Native	F1:	-	481	352
Vowels:	F2:	-	1394	1317

is 1287 Hz, while that same distance for English talkers is 1441 Hz. Japanese speakers exhibit some inward shift of English vowels from their canonical values.

The mean F1 and F2 ranges for native Hebrew vowels are in fact greater than their distances for native English vowels, although these distinctions would seem to be perceptually more critical to maintain the integrity of the English vowel space: the F1 and F2 ranges for Hebrew vowels were 352 Hz and 1317 Hz respectively. The mean F1 range for Hebrew talkers producing English vowels is 241 Hz, more than 150 Hz smaller than the F1 difference for English talkers. The F2 range is 1118 Hz, more than 300 Hz less than the English talkers' distance. Vowel systems can fall towards the center for articulatory ease when this does not interfere with perceptual adequacy. Hebrew speakers seem to be applying that strategy erroneously to a system which is too large to allow it.

4.2.8 Summary of Acoustic Results

Evaluation of formant values for vowels produced by English, Japanese, and Hebrew speakers indicates the following:

1. Comparisons of F1 and F2 differences between English and non-native speakers producing the "contrastive" vowel sets, and F1 and F2 differences for the non-contrastive vowel sets, were not statistically significant.

2. Formant deviation from English targets is described. Japanese speakers deviate least for vowels /i/, /ɛ/, /æ/, /ɔ/, /o/, and /u/. Hebrew speakers deviate minimally for /ɛ/, /ʌ/, /o/, and /u/. Interference from native vowels seems to occur for Japanese speakers producing /ɪ/, /e/, /ɔ/, and /o/. For Hebrew speakers, interference seems to account for production of English /i/, /ɛ/, and /ʊ/.

3. Overall differences between vowels produced by native and non- native speakers in terms of F1, F2, and F1+F2 Deviation did not correlate significantly with intelligibility scores for each vowel.

4. Standard deviations for F1 and F2 were compared for native and non-native speakers. No statistically significant differences were found across the three language groups.

5.The overall size of the Japanese speakers' English vowel space remains roughly similar to the English speakers' quadrilateral. However, the Hebrew speakers' English vowels, particularly back vowels, drift towards the center.

Chapter Five

DISCUSSION

The principal goals of this study were to explore the role of L1 interference in disrupting the pronunciation of L2 learners; to determine which sounds most impede intelligibility and account for perception of accent; to assess the stability of the L2 learner's phonetic targets; to determine phonetic criteria resulting in perception of accentedness and reduced intelligibility; and to compare perception of accent across tasks.

5.1 Models Accounting for L2 Learners' Phonological Distortions

5.1.1 L1 Interference and Unintelligible Segments

Based on a strict interference model (CAH), speakers should have most difficulty with sounds not occurring in their native inventories. For Japanese speakers, these include six of the eleven vowels examined in this study, the liquid /r/, certain fricatives, and all clusters. Considered as classes, intelligibility errors of the Japanese speakers conform perfectly to the model (see Tables 4.1 and 4.14). The Hebrew speakers should have errors on the same six vowels as the Japanese speakers, as well as 3-consonant clusters, /w/, /r/, /θ/ and /ð/. Again, intelligibility errors conform rather well to the CAH, with vowels and consonant clusters being least intelligible.

Japanese has a complex set of phonological constraints unlike Hebrew or English, and it seemed possible that these constraints would interfere with their English productions. Contrary to expectations, phonological contrastiveness did not contribute to reduced intelligibility. That is, Japanese speakers had no particular difficulty with sequences phonotactically unacceptable in Japanese, including syllable final consonants. This finding is inconsistent with the view that positional differences must be taken into account when describing regions of difficulty for L2 learners (Greenberg,1983) Thus, segments that never occur in L1 present more difficulty to L2 learners than sounds which can occur only in restricted environments.

Considering the contrastive segments individually mars the conformity somewhat. Vowels will be considered first. For Japanese speakers, / ɑ/ is unintelligible 54.17% of the time, making it the third least intelligible vowel in their English inventory, even though it is classified as non-contrastive. For the remaining four non-contrastive vowels, intelligibility data conform more closely to predicted results. The Hebrew speakers have less intelligible vowels overall, and it is more difficult to account for loss of intelligibility on contrastive principles alone.

The Japanese and Hebrew speakers performed differently when producing contrastive and non-contrastive vowels. As noted earlier, the Hebrew speakers' vowels were mistranscribed more often than Japanese speakers' vowels (45.09% vs. 36.61%). Also, the Japanese speakers had considerably more difficulty with

contrastive vowels than with non-contrastive vowels, while the Hebrew speakers performed poorly on both. This supports the hypothesis that Japanese speakers' productions of English vowels were more influenced by their native phonetic inventory than the Hebrew speakers'. By associating the five English vowels with the five "matched" vowels in their native set, Japanese speakers were more likely to be intelligible at least on four of the five near-equivalent vowels in English. The Hebrew speakers appear to have applied a strategy of centralization which has the effect of reducing intelligibility for all their vowels. Consequently, the five vowels that would have been more intelligible had the Hebrew speakers simply substituted sounds from their native inventory are no more intelligible than the remaining vowels.

When consonants are broken down individually, CAH still predicts errors rather well for Japanese speakers, with /v/, /th/, /r/, and most clusters highly unintelligible. However, there are many additional consonants with poor intelligibility scores - frequently higher, in fact, than the contrastive set. For example, /b/ resulted in an error rate of 29.17%, although it does not qualify as either phonetically or phonologically contrastive. Additional sources of error, discussed in section 5.1.2, need to be investigated to explain these findings. Hebrew speakers' consonant errors also conform well to CAH predictions, with 3-consonant clusters, /th/, and /r/ highly unintelligible. Again, there are unpredictably high error rates, such as for /p/, /b/, and /z/, that cannot be accounted for by appeal to contrastiveness.

5.1.2 Error Patterns and Alternatives to CAH

One possible alternative to CAH that might account for error patterns bases predictions on developmental data and linguistic universals, assuming these to reflect natural simplification processes. How well does such an analysis predict, for Japanese and Hebrew talkers, which sounds cause difficulty and how talkers handle that difficulty? Certain sounds may be inherently more difficult than other sounds for L2 learners irrespective of the native language inventory. Learners do not necessarily map English phonetic targets onto their native phonology and then struggle only with the group left over, even when this strategy could serve them well. The Hebrew speakers provide several examples of this. For example, they have a high percentage of unintelligible 2-consonant clusters, despite the fact that many of the 2-consonant clusters occur in their Hebrew phonetic inventory. Clusters are physiologically more difficult than simple consonants, however, because they require rapidly changing gestures. The effort of speaking a new language may distract a speaker's attention from aspects of speech that would be easily controlled under less stressful conditions. This would be analogous to findings in first language acquisition, where the addition of new content categories is often complemented by simplifying some other parameter, such as utterance length or phonetic accuracy.

Oller and Eilers' (1982) comparison of Spanish and English infants suggests the following preferences[25] for consonant-like sounds: singletons (vs. clusters); initial

[25] "Preference" is defined in terms of frequency of occurrence.

consonants (vs. final consonants); initial plosives (vs. initial fricatives and affricates); final voiceless obstruents (vs. final voiced obstruents); prevocalic glides (vs. liquids). Similar preferences have been reported for many other languages (Locke, 1983). Do the data of the present study conform to these findings from phonological acquisition in L1? The planned comparisons between consonant classes shown in Tables 4.3 and 4.4 address such questions.

For Japanese speakers, singletons were in fact significantly easier than clusters, but only in final position. While initial vs. final consonants were not explicitly compared, the lack of a position effect for either language group suggests that this finding was not confirmed. Voiceless consonants were better than voiced consonants in final position, conforming to the preference list above. While glides vs. liquids were not statistically compared, the error percentage for glides were 26.04, for liquids, 30.42, suggesting that they were equally difficult.

The Hebrew speakers, like the Japanese speakers, had significantly worse clusters than singletons only in final position. Again, the non-significant effect of position overall does not conform to the developmental finding that final consonants are more difficult than initial consonants. Fricatives were not significantly worse than plosives. Voiceless consonants were not significantly better than voiced in final position. Glides were unintelligible only 9.3% of the time, compared with 17.92% for liquids, conforming to the developmental data.

Universal processes frequently reflect articulatory constraints. Ohala (1983) cites several cross-linguistic studies indicating that velar stops are least compatible with voicing, while labial stops are most compatible. Ohala hypothesizes that the greater compliance of the oral cavity for labial stops allows passive oral cavity expansion, accomodating more glottal airflow before transglottal airflow falls below the level necessary to maintain voicing. Only the pharyngeal walls and part of the soft palate can yield to air pressure to prolong voicing for velar stops. One might therefore predict more voicing errors on alveolar and velar stops than bilabials. The Japanese speakers had a high error rate for /b/ (often heard as /p/) as well as /g/ (often heard as /k/). Contrary to expectations, the Hebrew speakers primarily had voicing confusions with /b/ (heard as /p/) and /p/ (heard as /b/). "Universal processes" therefore fail to predict the surprisingly high error rate for voiced bilabial stops. Overall, predictions based on natural simplification processes are not well supported by the data on consonant production gathered in this study on L2 phonological acquisition.

5.1.3 Predicting the Nature of Errors: CAH and Natural Processes

When interference predicts difficulty for a given sound, can it also predict how the speaker will produce that sound? The data gathered in the present study address this issue in terms of vowel production only, since the acoustic analyses obtained for vowels provide a stronger basis for inference about what speakers actually did than do the transcriptions available for the consonants. The role of L1 in producing vowel

errors may be more usefully discussed for the Japanese than for the Hebrew speakers, since the latter had serious problems with the whole English vowel set. In commenting on Figure 4.4, it was noted that contrastive / ɪ/ is drawn toward familiar /i/; contrastive /ɛ/ towards familiar /e/. Contrastive / ʊ/, however, is centralized and appears unrelated to any other vowel in Japanese or English. Thus, the Japanese speakers do not appear to have a single strategy for contrastive vowels. The Hebrew speakers, as Figure 4.5 shows, tend to centralize their vowels. This does not relate to L1 interference; but it does perhaps reflect a natural simplification process. Developmental data provided by Lieberman (1980) show that as children grow (from the time of their first words, around the 12th month, through the third year) their formant frequencies fall and the overall size of their vowel quadrilateral expands. This suggests that a reduced vowel space is in some sense easier. The tendency of Hebrew speakers to centralize English vowels may reveal a natural strategy when confronted with a difficult vowel system.

In summary, L1 interference better predicts which English vowels are difficult for Japanese speakers than do natural simplification processes; neither interference nor natural simplification processes predicts the phonetic form of the difficult vowels. L1 interference does not predict problem vowels for the Hebrew speakers, but natural processes may account for how Hebrew speakers simplify vowels in English.

5.2 Contrastiveness, Intelligibility, and Accentedness

5.2.1 Relation between Contrastiveness and Intelligibility

Listeners provided accentedness judgments as well as transcriptions of each single-syllable token in this study. The accentedness ratings were less useful than the transcriptional data in isolating regions of difficulty for the L2 learners. Had particular vowels or consonants been largely responsible for low accentedness ratings, the repeated occurrence of these segments across tokens would have increased (worsened) the accentedness scores of any tokens in which they appeared. No particular segments or phonetic classes were found to be singularly responsible for accentedness ratings. Since listeners always heard no less than a single syllable, it was unclear exactly what was accounting for poor accentedness ratings. The data were therefore broken down further with respect to number and type of contrastive segments within the syllable and accentedness ratings. The failure to find differences based on the number of contrastive segments suggests that contrastive segments, though responsible for reduced intelligibility, are no more responsible for accent than non-contrastive segments.

5.2.2 Relation between Accentedness and Intelligibility

Listeners' judgments of accentedness correlated highly with intelligibility for all the English talkers and for four of the eight non-native talkers. That is, the mistran-

scribed utterances were given poor accentedness ratings more often than correctly transcribed utterances.

Two hypotheses emerge from this finding. First, the features interfering with a talker's intelligibility may be the same as those accounting for accentedness, with accentedness the more sensitive measure of the two. Despirated word-initial /t/ would provide an example of this phenomenon. The non-English sounding /t/ might be correctly transcribed as /t/, yet be clearly recognized as non-native (i.e., accented). The lack of aspiration could also interfere with intelligibility, and the token might be confused with /d/.

The second and less interesting interpretation is that listeners were not able to make independent decisions. When they could not confidently transcribe the token, they automatically assumed it was more accented. For example, listeners unsure of whether the talker produced /a/ or /ʌ/ may have automatically given the talker a high (poor) accentedness rating. Since native English talkers were also misunderstood frequently when producing these two vowels, the confusions suggest something about the difficulty listeners have in making fine distinctions between acoustically similar vowels without disambiguating context, rather than speaker errors. In that case, judgments of accentedness would be indices of the listeners' confidence in the transcription: When listeners had difficulty in transcribing the token, they rated the speaker as having a worse accent. This strategy provides relevant information in cases where speakers produced serious distortions of consonant seg-

ments that were still easily identifiable. The accentedness rating is the only clue that a speaker's production was in some way unacceptable as an English target, even though it did not cross phonemic borders and become confusable with some other segment. Using confidence in transcription as an overall strategy for deciding on degree of accentedness, however, reduces the meaning of accentedness to a mirror image of intelligibility, without independent status. An L2 learner may, after all, produce an easily identifiable item with a strong accent, and this would not be reflected in listeners' responses.

It was noted earlier (section 4.1.9) that real words were given better accentedness ratings than nonsense syllables. Listeners may be more confident of their transcriptions, and therefore give the token a better accentedness rating, when they have transcribed a real word. The difference between English and Japanese talkers on the one hand, and Hebrew talkers on the other, may reflect greater distortion in segments produced by Hebrew talkers. The degree of distortion may have been such that listeners could not benefit from the presence of real words, never feeling confident in the accuracy of their transcriptions, or, indeed, whether they had heard a real word.

5.2.3 Features Determining Accentedness

What determines perception of accentedness? The finding that contrastive segments within the syllable do not result in worse ratings of accent implies that

contrastive status of a segment is not the determining factor. The finding that perception of accent in sentences correlates with perception of accent in single words suggests that segmental information is relevant to our overall judgment of accentedness. So, while segmental information is relevant, the information does not reside primarily in contrastive segments.

Loss of intelligibility, unlike accentedness, clearly fell more on contrastive than non-contrastive segments. Japanese talkers in particular were less intelligible for liquids, contrastive fricatives, clusters, and contrastive vowels. Thus, given two tokens produced by a Japanese talker, one highly contrastive (e.g., /br(ɪ)v/) and one not (e.g., /ko/), intelligibility should be impaired only for /br(ɪ)v/; accentedness scores may be equally good or poor for both.

Second language learners, confronted with sounds in L2 without phonemic equivalents in L1, may use any of the following hypothetical strategies:

(1) substitute another similar L2 sound (e.g., produce English /d/ for both /d/ and /ð/);

(2) substitute a similar sound from their L1 inventory (e.g., produce native /d/ for English /ð/;

(3) attempt to produce the new target, but distort it.

Applying strategy (1) would impair intelligibility scores but not accentedness ratings. Strategy (2) would result in unintelligible productions, which may also sound accented, depending on the distance of the native substitution from any sound in the

L2 inventory. Strategy (3) would impair intelligibility only if the distortion were sufficiently great; perceived accentedness would probably be worse than (2).

When L2 learners attempt to produce sounds with L1 phonemic equivalents, strategy (3) may again apply. Alternatively, speakers may employ strategy (4):

(4) substitute the L1 phonemic equivalent.

Intelligibility of the segment, as well as perceived accent, would then depend on how phonetically close the L1 equivalent is to the L2 target.

How, then, can we explain the correlation of intelligibility with contrastive status, but the lack of correlation between accentedness and contrastiveness? If we hypothesize that L2 learners use strategies (2) or (3) for contrastive segments, intelligibility and accentedness should be worse for contrastive sounds. But if L2 learners then apply strategy (4) to non-contrastive sounds, intelligibility of the non-contrastive group may be acceptable while accentedness remains great. Thus, no correlation would be found between accentedness and contrastive status, since distortion - accounting for perceived accent - occurs on both contrastive and non-contrastive segments.

A puzzle remains, however, in trying to account for the correlation between accentedness and intelligibility for four of the eight non-native talkers, where strongly accented tokens were more often unintelligible than mildly-accented tokens. The preceding argument suggests a high correlation between accentedness and intelligibility for contrastive sounds, and a low correlation for non-contrastive sounds.

One would therefore predict that accentedness would not correlate with intelligibility. An extremely high correlation for the contrastive set may be responsible for the correlation between accentedness and intelligibility for four of the non-native speakers. The low correlation between accentedness and intelligibility for non-contrastive sounds may not have been sufficiently great to weaken the overall correlation.

5.3 Intralingual Vowel Distances and L2 Transfer

What accounts for distances between vowels in a language? It would seem that vowel distributions would be determined, at least in part, by the number of vowels in a particular system (Liljencrants and Lindblom,1971). A system with many vowels needs to mark regions of tongue constriction within the oral cavity more carefully than one with few vowels in order to retain articulatory and perceptual distinctions. A five vowel system, for example, can allow phonologically contrastive vowels to fall phonetically towards the center of the vowel quadrilateral without losing distinctiveness. Language specific differences may also exist, so that two languages with an equal number of vowels may distribute vowels differently, although the overall area should be roughly equal (Manuel and Krakow, 1984). The data of the present study provide some insights on this issue. First, speakers of Japanese and Hebrew, both 5-vowel languages, used roughly the same overall area in their native languages as did English speakers in English. This is inconsistent with the prediction of inward drift in smaller systems, for which tolerable phonological distinctions might

be maintained even with moderately centralized vowels. Second, the two non-native groups did not use the available English vowel space in the same way: The Japanese talkers used a slightly smaller area than they do in native Japanese, while the Hebrew speakers used a markedly smaller area. Thus the number of vowels in the English system did not seem to influence talkers' vowel spaces as a simple function of their native systems.

5.4 Intelligibility of Consonants over Vowels

The lower intelligibility for vowels than for consonants across all language groups results from fundamental differences between the two major phoneme classes. In terms of production, consonants require some degree of closure at particular points of articulation, monitored tactilely by the speaker, though points of closure are relatively grossly defined. For example, a voiceless stop with tongue tip closure anywhere from the teeth to the hard palate will result in a /t/-like sound, even if slightly distorted by production at an atypical point for a given language. Therefore, consonants have more discrete targets than vowels, both articulatorily and perceptually.[26] Vowels, by contrast, are produced more continuously, and the articulators receive considerably less tactile feedback. Small shifts in tongue height or along the front-back dimension in the oral cavity can result in markedly different vowels.

---

[26] Despite the relative ease of consonant over vowel place of articulation, non-native speakers produced enough consonantal distortion to result in low intelligibility. Therefore, other sources of error must be sought. Consonants, unlike vowels, have voicing and manner distinctions, marked primarily by differences in timing and degree of constriction. These issues have not been addressed in the present study.

5.4.1 Perception and Production of Vowels

Certain vowels are presumably more difficult than others both to perceive and produce. (Peterson and Barney, 1952; Singh and Black, 1966). The notion of simplicity of sounds according to "natural" or "universal" processes is appealing, but would be more appealing if it could be identified in physiological terms, such as type and degree of feedback available. Vowels that provide more feedback may be easier to execute than vowels that provide less.

Perusal of the perceptual errors for all language groups reveals considerably worse perception for low and mid than for high vowels. For non-native vowels, some of this perceptual difficulty may arise from the talkers' lack of articulatory skill and the resulting acoustic deviance. High vowels require constriction close to the palate, so that a talker receives tactile, as well as kinesthetic and auditory, feedback. Tongue movement for low and back vowels, however, provides only kinesthetic and auditory feedback, since the tongue does not make contact with other articulators. In addition, the tongue tip is more densely innervated, and therefore more sensitive, than the tongue root, the constriction region for / ɑ/. Reduced feedback could result in less accurate articulatory placement, and this could be a source of the tendency among Japanese, and particularly Hebrew talkers, to centralize their English back vowels.

The relatively poor intelligibility for low and mid vowels of English speakers, however, suggests that perceptual factors are also at work, since native speakers presumably do not require multiple sources of feedback to maintain intelligibility. Vowel positions differ from one another in height as well as in frontness and backness; distinctions within front and back vowels are often described only in terms of differences in height. From a scan of figures 4.4 and 4.5, however, it is immediately clear that vowels in the so-called front group vary considerably with respect to frontness and backness as well as height. The (F2-F1) difference between /i/ and /æ/ for English speakers is 1107 Hz, while for /u/ and /ɑ/ it is only 155 Hz. Accordingly, to produce distinctions among the front vowels, speakers vary the parameter of tongue frontness-backness over a greater range than for back vowels. Articulatorily, this could mean that back vowels are easier than front vowels, because they require attention to a single parameter rather than two. Perceptually, however, the front set may be easier for listeners to sort out, because the acoustic index of front-back variation (F2-F1) has a wider range than for the back vowels, for which indeed the two lower formants may be impossible to separate acoustically by standard measurement procedures. This suggests that listeners will be more tolerant of shifts from the English targets for the front set than the back set, since the front set is better specified acoustically. In general, the worst vowels (perceptually) for both non-native language groups were, indeed, the back set. For English speakers, the only two vowels with more than 30% unintelligibility were /ɑ/ and /ʊ/.

5.4.2 Vowel Acoustic Deviance and Loss of Intelligibility

Statistical tests described in Chapter Four failed to find significant relations between degree of formant deviance and degree of intelligibility. Two conclusions could be abstracted from this finding. First, formant information may not account for deviance in perception. Extensive documentation on the importance of F1 and F2 in vowel perception, however, makes this possibility unlikely (Delattre, Liberman, Cooper, and Gerstman, 1952). The second, more plausible conclusion is that the vowels are not shifted uniformly within a language group. That is, speakers of a given language did not shift F1 or F2 in a single, consistent direction across all vowels. In addition, listeners may be perceptually more sensitive to small acoustic shifts in some regions of the vowel space than in others.

The vowels produced by non-native speakers can be analyzed as a system rather than as a series of independent vowels. Two questions are appropriate for a system analysis. First, does the overall "envelope" for the L2 learner's English vowel space differ in size with respect to either a native English speakers' vowel space or to his own native vowel space? Centralization of vowels was discussed earlier as a possible strategy for speakers with a small and highly distinctive vowel inventory. It may also be relevant in explaining how L2 learners handle unfamiliar vowels. Centralization could serve as a strategy in producing vowels that the speaker cannot "anchor" to specific tongue positions. This would probably have a more devastating effect on intelligibility than producing vowels no less acoustically deviant, but at the periphery.

For example, both non-native groups were highly intelligible for their productions of English /e/. This did not necessarily reflect more accurate production of /e/ than some other, less intelligible vowel. Rather, the vowel /e/ seems to occur in a region that permits laxer production before loss of intelligibility results. Had non-native speakers responded to their uncertainty about /e/ production by centralizing and producing a more schwa-like vowel, perceptual results would presumably have suffered.

A second question appropriate for a system analysis addresses the relations among subsets of vowels. The differences between back vowels produced by L2 learners and the English target is so great that talkers may seem to be randomly stabbing at the vowel targets. However, if the F1 and F2 relations among the vowels are maintained, then the speaker has obviously internalized the systematic differences among the vowels phonologically, and is having difficulty with precise phonetic execution. Alternatively, the speakers may have no sense of how to produce the vowel or where it belongs in the overall system.

In the following discussion, vowel articulation is inferred from the formant information, so that F1 is considered an indicator of tongue height, and (F2-F1) a measure of tongue frontness and backness. Fig. 4.4 indicates that Japanese talkers have internalized appropriate relations among the front vowels, using English talkers' vowels as criteria. Their front vowel system is somewhat broader than that of English speakers, and this is probably attributable to L1 interference. The vowels /ɪ/,

/e/, and /ɛ/ are pulled towards native Japanese /i/ and /e/, but the Japanese talkers have apprehended the differences among them with respect to tongue height and frontness-backness. The fact that / æ/ is a low front vowel to be distinguished from /ɛ/ has also been internalized, but the Japanese speakers overshoot the mark in moving from /ɛ/ to / æ/ with respect both to tongue lowness and to tongue backness.

The strategies for Japanese speakers' production of back vowels in English differ from those for front vowels. First, the height relations among /ʌ/, / ɔ/, and / ɑ/ are not preserved, so that /ʌ/ is produced as the lowest vowel of the three. The interference from similar vowels in L1 is less obvious here, with / ɑ/ actually closer to English / ɑ/ than native Japanese / ɑ/, and / ʊ/ unrelated to any target in Japanese or English. Also, the back vowels tend to shift towards the center in a way that the front vowels do not. As already noted, this phenomenon can be accounted for in terms of either production or perception. A production-based analysis notes differences between the constriction regions of the tongue for back and front vowels. The front vowels, produced near the palate, may provide more tactile feedback for the talker. Reduced tactile feedback for back vowels may result in overall reduction of accuracy, manifested as a drift towards the center of the oral cavity. Another explanation for the phenomenon of centralization is perception rather than production based. The back vowels may be inherently more difficult to distinguish from one another perceptually. Note that even English speakers had more than 20% loss of intelligibility for vowels / ʊ/ , /ʌ/, and / ɑ/. L2 learners unable to sort out the

differences may respond by allowing the whole system to drift inward, in a schwa-like fashion.

The Hebrew speakers are less phonologically accurate than the Japanese speakers. As shown clearly in Fig. 4.5, they do not maintain distinctions in the appropriate dimensions for /i/-/ ɪ/, /e/-/ɛ/-/ æ/, / ɑ/-/ʌ/, or /o/-/ ɔ/-/ ʊ/. The lack of distinctiveness therefore is not linked specifically to back vs. front vowels, or high vs. low ; it cuts across the entire system.

Another interesting difference between the Hebrew vowel system in English and the native English and Japanese groups is the dramatic inward shift, both in the height and front-back dimensions. This phenomenon is limited to the low and back vowels. It is noteworthy that the English system produced by the Hebrew speakers is in fact narrower than their native Hebrew system, despite the fact that this compresses an already crowded vowel space. Using the entire available vowel space would be a more efficient strategy, as the vowels would be less likely to collide. Ironically, unfamiliarity with the phonology of a new vowel system may result in reduction of the space the talker uses.[27]

---

[27] Systemic "centralization" would refer to a reduction in the acoustic distances among all vowels, describing a system where vowels are pulled from the periphery towards the center. The vowel formants of the Hebrew talkers do not conform to such a strict definition of centralization, since their vowels are not uniformly pulled inward.

5.5 Perceptual Consequences of Vowel Acoustic Deviance

As we have seen, the L2 learner's English productions sometimes seem to be influenced by the nearest-matching vowel in their native inventory. This has different consequences, depending in part on how divergent the native vowel is from the English target.

Comparing the acoustic distances between English vowel formant targets and the formant values produced by the L2 learner provides insight into the relation between acoustics and perception. The Japanese speakers' English vowels /i/ and /u/ are produced close to their native Japanese equivalents. No other native English vowel is produced in that region so that, even with some deviance, /i/ and /u/ are easily identified as the intended targets. The English target /ɪ/ is produced with a higher F1 and lower (F2-F1) values than the English target /i/. While a phonological distinction is perhaps maintained, the Japanese speakers do not produce a sufficiently large acoustic distinction between /i/ and /ɪ/, so that a high percentage of /ɪ/ is perceived as /i/.

The error rate for /e/ perception is very low, considering how much lower the F1 value is for Japanese than for English speakers. The Japanese error rate for /e/, in fact, is 2.8%, compared with 17% for native English speakers. The Japanese speakers may have used some additional cue beyond F1 and F2 information, such

as duration or diphthongization (formant movement).[28] Or, an /e/-like vowel in the front-back dimension, coupled with a high F1, may serve as an unequivocal universal cue for /e/ - a cue that the native English speakers did not provide as unequivocally. This suggests the interesting possibility of canonical vowel types that are perceptually less ambiguous as English targets than the targets actually produced by native English talkers. The low error rate for /e/ for all language groups, despite acoustic deviation from the English speakers' target, provides support for the hypothesis that certain regions of the vowel space are relatively more impervious to minor deviations than others. The vowel /e/ appears to occupy a large portion of the vowel space, so that any vowel produced within that general area is accepted as an intelligible token of /e/.

The vowel / æ/ appears to have a "floor" in terms of how far low and back it may be produced before being confused with / ɑ/; this limit was apparently passed by the Japanese speakers. Similarly, / ɑ/ must have a "ceiling" after which it is confused with some higher vowel, even when that higher vowel requires rounding specification absent in / ɑ/. Japanese talkers distinguish /o/ and / ɔ/ in the height dimension, while English talkers produced /o/ with a fronter point of constriction than / ɔ/. The Japanese strategy was apparently unsuccessful, particularly for / ɔ/, which was most often misheard as /o/. Finally, the centralized / ʊ/ was frequently

---

[28] Intrinsic durational differences among vowels, and how durations vary depending on consonantal context, were not considered in the present study.

misheard, and, as might be predicted, it was not consistently misheard as any one sound.

The formant values for Hebrew speakers were frequently deviant and unpredictable; this corresponds with the perceptual finding of high error in terms of vowel intelligibility. As noted in Table 4.10, Hebrew speakers had high error rates for /i/ and /ɪ/, where one was confused with the other. Both vowels are produced with considerably higher tongue position (determined by F1) than the English /ɪ/; the formant values would therefore predict that both vowels be misheard as /i/. The error rate in the opposite direction suggests that some additional cue for /i/, such as increased duration over /ɪ/, was not produced, resulting in confusions with /ɪ/. The vowel /e/ remains robust and intelligible despite its deviance from English target /e/, further supporting the hypothesis that /e/-like vowels are generally easy to perceive even when they do not match the English formant target. The relatively good intelligibility for acoustically deviant /ɛ/ is mysterious. The vowel /æ/, which was produced in a region more appropriate to /ɛ/ than /æ/, was often misheard as /ɛ/. Listeners seem to have taken the region where /æ/-/ɛ/ were produced by Hebrew talkers as an /ɛ/-region, resulting in higher intelligibility for /ɛ/ than would be predicted. The vowel with the highest error rate is /ɔ/, which could be predicted given its unorthodox F1 and F2 values. The poor intelligibility for the remaining back vowels is puzzling, particularly for /u/, which is closer to the native target /u/ than Japanese speakers' English /u/ production. An attractive explanation would be that the centralized English back vowel system of the Hebrew speakers did not allow easy

separation of the vowel targets. Since the perceptual task was not presented in block design, however, listeners did not have the opportunity to draw conclusions about a speaker's system, and decisions about vowel type had to be made on a vowel-by-vowel basis.

In light of the above discussion, the failure to find significance in statistical tests matching perception and acoustics is not surprising (see 4.2.5). The acoustic deviations from the English target norms are not consistent across vowels; for example, Japanese speakers do not consistently shift F1 upward, or F2 downward. The shifts seem rather to involve interactions among several strategies, such as attempts to make the L2 target similar to some native language equivalent, or shoot for the new target too energetically and overshoot the mark, or complete confusion resulting in a centralized vowel. The different strategies are reflected in different F1 and F2 deviations in different directions, and lack of statistical significance becomes predictable.

Significant correlations were expected in particular between F1+F2 Deviation and loss of intelligibility, since this measure incorporates deviance from the target with respect to both F1 and F2; but the correlation was not found. The hypothesis that certain regions require greater phonetic accuracy than others was posited earlier, and may be offered as an explanation. If the hypothesis is valid, minor deviations in F1 and/or F2 from the native target are more critical to the intelligibility of some sounds than others. The F1+F2 Deviation might be described as a "blind measure"

that weights deviance equally wherever it occurs in the vowel space. It would seem, then, that an absolute distance metric applied across the entire vowel space is not useful in predicting breakdown of intelligibility.

5.6 Stability of L2 Learners' Targets

Are L2 learners with poor intelligibility trying to achieve the English targets, and missing? Are they aiming for the wrong target, and achieving it consistently? Or are they randomly and inconsistently stabbing at unfamiliar targets? The first two questions, while interesting, cannot be addressed from this experiment, since perceptual data from the L2 learners were not obtained. The third possibility - that is, learners making random stabs at some target - was tested by comparing standard deviations across vowels for English, Japanese, and Hebrew speakers. The non-native speakers were no more variable than English speakers. Thus, the L2 learners sampled here did have some definite target in mind, and performed in a consistent fashion when trying to produce it. A question for future study is whether they have perceived the critical features of the English targets that they need to replicate. To answer this question, L2 learners' perception of English must be tested as well as their production. For example, an identification task where L2 learners hear "canonical" and "deviant" L2 sounds could establish whether learners have internalized correct targets.

When talkers produce a target incorrectly, they may either substitute or distort the intended segments. Clear substitutions would not have been detected by listeners in this task, since unpredictable nonsense words were included. Thus, substitutions of other phones in the target language would have reduced intelligibility scores, but would not have resulted in poorer accentedness scores. The close association between poor accent and mistranscribed segments implies that talkers in this study distorted rather than substituted mistranscribed segments. The fact that the non-native speakers distorted target segments may provide information about their perceptual capacity in English. Substituting /i/ for /ɪ/ could reflect a production error, or it could mean that learners did not detect differences between the two phones (perception error). Distorting /ɪ/, however, and producing a sound intermediate between /i/ and /ɪ/, implies that L2 learners are aware of differences. The problem might still be perceptual, however, if they detect some, but not all the features distinguishing the two sounds. In the latter case their production of the target is correct to the extent that it includes the features that listeners find perceptually salient.

5.7 Accentedness Judgments: Citation Forms vs. Longer Speech Samples

Listeners did not have access to prosodic information beyond the word in the single word task, yet their judgments of accentedness matched their judgements in a sentence task, where prosodic information was clearly present. This close correspondence between accentedness scores on single words and sentence materials

suggests two hypotheses. First, listeners may have based their decisions about accentedness primarily on segmental information in both tasks, ignoring the additional intonation information in the sentences. This rejects the notion that prosodic information is the critical cue in assessing foreign accents. Alternatively, listeners may have depended heavily on prosodic information in the sentence task, where it was available. However, the prosodic information may correlate highly with the segmental information. While the listeners' judgments may have been based on different sets of information in the two tasks, the two sets may correlate highly with each other. The talker who produced native-sounding segmental information may have produced native-sounding intonation; those with poor segmentals may have similarly poor prosodic control. Both hypotheses are interesting areas for further research.

5.8 L2 Learner Variability

Various methodological problems arise in a study of this nature. First, measuring multiple tokens per speaker reduces the number of speakers one can reasonably include in the study. As noted in 4.2.2, the small number of English speakers sampled provided reliable data when compared with data derived from Peterson and Barney (1952). The non-native speakers, however, cannot be easily compared with standardized data. In fact, the non-native speakers probably have more built-in variability. Several studies suggest that L2 learners use different strategies at different stages of L2 acquisition (Lobo and Yoshida, 1982; Wode, 1980), with L1 interference

dominant at early stages and other processes emerging later. English proficiency of L2 learners was controlled by choosing students at similar ranks in English Language Institutes. However, this reflects more about their syntactic competence than their phonological abilities. These competences need not develop in parallel; a student with mastery of L2 syntactic structure can have poor control of phonological features in L2. Thus, some of the speakers may have been at a phase of interlanguage where L1 interference played the major role in their pronunciation errors; other speakers may have been influenced more by natural simplification processes. The problem of comparing learners at different IL phases could be avoided by doing single-subject studies, or by using some metric determining the learner's phonological stage of interlanguage before including him in such a study. One way to ensure phonologically homogeneous learners is to obtain accentedness scores for speakers before including them in a phonological study. It would then be possible to average across speakers more confidently, without obscuring relevant inter-speaker differences.

Chapter Six

SUMMARY

Adult non-native speakers generally exhibit phonetic deviance in English even when they control syntactic and semantic features reasonably well. The present study analyzed second language (L2) learners' difficulties with English sound patterns by examining a representative sample of English syllables produced by native Japanese and Hebrew speakers. The following questions were addressed:

(1) Which model better accounts for pronunciation errors of L2 learners: first language (L1) interference or natural simplification processes?

(1.1) Are there systematic differences between Japanese and Hebrew speakers reflecting differences in their L1 phonologies?

(2) Which classes of sound most significantly interfere with phonetic intelligibility and/or account for perception of accentedness in the L2 learner?

(2.1) What is the acoustic basis of poor intelligibility?

(2.2) Does intelligibility correlate significantly with the extent of acoustic deviance from the target?

(2.3) Does the L2 learner have stable, but incorrect targets, or appropriate targets with high variability in production?

(2.4) Are the L2 learner's phonetic difficulties rooted in problems of perceiving L2 targets, producing them, or both?

(3) What is the relation between accentedness and intelligibility?

(4) Do judgements of accentedness for isolated, citation form syllables and for longer speech samples correlate significantly?

Four native Japanese, four native Hebrew, and four native English male speakers (controls) were recorded producing 132 English syllables and disyllables, and twelve simple sentences. A phonologically representative sample of syllables from his native language was also recorded by each Japanese and Hebrew speaker, as a baseline for acoustic comparisons with his English utterances. The test syllables of all speakers were randomly ordered on tapes, and played to 36 experienced listeners, who were asked both to transcribe each utterance (thus providing a measure of its intelligibility) and to rate its degree of accentedness. Only accentedness judgments were obtained for the sentences. The first and second formant frequencies of English vowels were measured and compared with those of the English speakers, and with the nearest equivalent vowel in the speakers' native phonetic inventories.

The least intelligible phonetic classes were, in order, vowels, clusters, fricatives, and liquids for Japanese speakers; vowels and clusters for Hebrew speakers. When consonants were grouped by phonetic class, L1 interference adequately predicted regions of low intelligibility. When individual segments were considered, however, predictive adequacy broke down: Several English segments with L1 phonological equivalents received poor intelligibility scores. Predictions based on natural simplification processes were also not well supported.

Acoustic and perceptual measures on vowels, and comparisons with vowels from native Japanese and Hebrew, indicated that L1 interference accounted for the vowels that Japanese speakers produced poorly; but neither interference nor natural simplification processes predicted the phonetic form of the difficult vowels. For Hebrew speakers, L1 interference did not predict regions of difficulty, since the entire vowel system was deviant. Natural simplification processes may provide a better account of how the Hebrew speakers handled the English vowel system.

Intelligibility of non-native speakers' English vowels did not correlate with degree of acoustic deviance, defined as F1 and F2 distance from the English speakers' vowel productions. It appears that more acoustic drift is permissible in some regions of the vowel space than in others without resulting in confusions with some other vowel. Although the L2 speakers in this study had a high percentage of unintelligible vowels in English, their vowels were not significantly more variable than vowels produced by native English speakers. This suggests that their L2 systems were stable.

If the non-native speakers had substituted other English phones for the intended targets, intelligibility scores, but not accentedness ratings, would have suffered. The close correspondence between poor accent and mistranscribed segments therefore implies that talkers in this study distorted target segments rather than substituted other English sounds. This finding, coupled with acoustic evidence that many poorly perceived vowels were nonetheless shifted from the values predicted on the basis of

L1 interference toward the correct target, suggests that the L2 learners perceived more accurately than they spoke.

For most of the speakers, listeners' judgments of accentedness correlated highly with intelligibility. We can explain this in two ways. First, perhaps listeners were not able to make independent decisions: When they could not confidently transcribe a token, they automatically assumed it was more accented. A more interesting alternative is that the features interfering with a talker's intelligibility are the same as those accounting for accentedness.

A high correlation was found between accentedness ratings on sentences and on single words. Since single syllables, unlike sentences, provide minimal prosodic information, we may interpret this as meaning that listeners based their accentedness ratings primarily on segmental information in both tasks. Alternatively, perhaps talkers with native-sounding segments had native-sounding intonation, while talkers with poor segments also had poor prosody.

APPENDIX A1

INSTRUCTIONS

This is a list of words for you to read. Some of the words are real English words, but some of them are not. All of the words have a "rhyming word " next to them to help your pronunciation.

Example: cat - lat
hot - sot

Look at the words on the left hand side of the page and make sure you are familiar with them. (This is not a vocabulary test. You won't need to give definitions). Read each word on the left with its rhyming pair on the right. Please read this list slowly. The first few words are for practice only.

feet - keet

go - low

way - nay

hate - tate

Now, please continue with the real list.

APPENDIX A2

WORD LIST[29]

pot - rot	borrow - sorrow	men - ten
carry - harry	could - hood	pot - jot
Bob - tob	bed - ched	book - pook
owed - sode	holds - molds	be - dee
bet - net	come - bum	edge - hedge
paws - faws	sing - ping	more - tore
ash - mash	saw - paw	pack - gack
let - bet	fear - hear	roof - zoof
bean - wean	soak - choke	know - stow
suit - hoot	be - she	sit - bit
book - took	sun - mun	look - dook
them - shem	car - tar	too - you
as - fazz	seen - mean	both - toth
marks - tarks	bug - kug	song -kong
case - dase	tooth - sooth	pawed- sawed
took - chook	paws - tawz	which- itch
could - zould	be - yee	pet - met
school - vool	book - wook	soak - woke
bet - wet	smooth - toothe	pin - min
seat - leet	laughed - taft	feel - veal
give - live	sock - gock	hats - mats
hit - mit	pot - hot	page - sage
pillow - willow	end - tend	map - strap
he - chee	full - vull	follow - hollow

[29] Target segments are underlined.

breathe - eathe

off - zoff

law - straw

me - bree

birds - herds

pass - dass

saw - raw

hate - zate

catch- batch

be - ree

ham - sham

sir - ter

me - stree

hate - nate

bees - keeze

me - zee

keep - eep

ask - bask

man - tan

bone - tone

walk - thawk (/θ/)

blue - chew

push - mush

hat - rat

mess - dess

my - ply

rib - mib

home -tome

beg - keg

was - fuzz

be - tea

bet - let

bee - see

will - mill

them - yem

tip - ip

built - dilt

save - lave

fast - bast

rub - mub

meet - beat

cut - but

bit - thit

code - thode (/ð/)

come - hum

then - men

hum - yum

day - fray

ma - sha

rough - zuff

hit - nit

them - bem

ham - tam

blue - grew

web - teb

paid - jade

bead - theed (/ð/)

him - bim

beach- each

go - ko

APPENDIX A3

Questions presented to elicit sentences:

1. What time is it?

2. What street do you live on?

3. What color is your shirt?

4. How many brothers do you have?

5. What is your favorite food?

6. How many chairs are there in this room?

7. What color is the Statue of Liberty?

8. Who won the democratic nomination?

9. What sport do you like to do?

10. How many blocks do you live from the subway?

11. What's the temperature outside today?

12. Where did you go last Sunday?

APPENDIX A4

SAMPLE OF LISTENER'S RESPONSE SHEET

KEY

1. - Speaker sounds like a native American English speaker
2. - Speaker has slight, almost imperceptible accent
3. - Speaker has moderate accent
4. - Speaker has strong accent
5. - Speaker sounds like he is not producing English sounds

Transcribe what the speaker says:	Circle the number that best corresponds to the speaker's accent:
1.________________	______1______2______3______4______5______
2.________________	______1______2______3______4______5______
3.________________	______1______2______3______4______5______
4.________________	______1______2______3______4______5______
5.________________	______1______2______3______4______5______
6.________________	______1______2______3______4______5______
7.________________	______1______2______3______4______5______
8.________________	______1______2______3______4______5______
9.________________	______1______2______3______4______5______
10.________________	______1______2______3______4______5______

APPENDIX B1

Formant Values of Averaged English Vowel Data for English Speakers

	F1	F2	F2-F1
/i/:	297	2351	2054
/ɪ/:	420	1812	1392
/e/:	411	1931	1520
/ɛ/:	499	1710	1211
/æ/:	656	1602	947
/ɑ/:	703	1181	478
/ʌ/:	515	1188	673
/ɔ/:	481	910	429
/o/:	470	1060	590
/ʊ/:	444	1188	744
/u/:	311	944	633

APPENDIX B2

Formant Values of Averaged English Vowel Data for Japanese Speakers

	F1	F2	F2-F1
/i/:	309	2277	1969
/ɪ/:	337	2169	1832
/e/:	458	1986	1528
/ɛ/:	500	1831	1331
/æ/:	689	1496	808
/ɑ/:	615	1095	480
/ʌ/:	700	1323	623
/ɔ/:	532	1008	477
/o/:	491	990	500
/ʊ/:	380	1440	1061
/u/:	345	1077	732

APPENDIX B3

Formant Values of Averaged English Vowel Data for Hebrew Speakers

	F1	F2	F2-F1
/i/:	330	2143	1813
/ɪ/:	328	2013	1685
/e/:	482	1796	1314
/ɛ/:	533	1436	903
/æ/:	540	1634	1094
/ɑ/:	524	1113	586
/ʌ/:	569	1332	763
/ɔ/:	381	1145	765
/o/:	438	1042	603
/ʊ/:	373	1025	652
/u/:	335	1027	692

APPENDIX B4

Formant Values of Averaged Japanese Vowel Data

	F1	F2	F2-F1
/i/:	304	2227	1923
/e/:	463	1942	1479
/ɑ/:	785	1264	479
/o/:	474	833	359
/u/:	316	1204	888

APPENDIX B5

Formant Values of Averaged Hebrew Vowel Data

	F1	F2	F2-F1
/i/:	330	2182	1852
/e/:	443	1812	1369
/ɑ/:	682	1362	680
/o/:	488	918	430
/u/:	345	865	521

REFERENCES

Altenberg, G. P. & Vago, R. M. (1983). "Theoretical implications of an error analysis of second language phonology production." Language Learning, 33, 427-447.

Anderson, J. I. (1982). "Syllable simplification in the speech of second language learner's." Paper presented at TESOL conference, Hawaii, 1982.

Brennan, E., Ryan, E. B., & Dawson, W. E. (1975)."Scaling of apparent accentedness by magnitude estimation and sensory modality matching." Journal of Psycholinguistic Research, 4, 27-36.

Briere, E. (1966)."An investigation of phonological interference." Language, 42, 768-798.

Broselow, E. (1984)."An investigation of transfer in second language phonology." International Review of Applied Linguistics, 22, 253-269.

Broselow, E. (1983)."Non-obvious transfer: on predicting epenthesis errors." in L. Selinker & S. Gass, (eds.) Language Transfer in Language Learning. Rowley, Massachusetts: Newbury House.

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

Cairns, C.E. (1969). "Markedness, neutralization, and universal redundancy rules." Language, 45, 863-885.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. G., & Carbone, E. (1973). "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals." Journal of the Acoustical Society of America, 54, 421-428.

Chayen, M. J. (1973). The Phonetics of Modern Hebrew. The Hague: Mouton.

Davidian R. & Flege, J. E. (1984)."Typological hierarchies and learning difficulty hierarchies in adult second language phonology: is there a correlation?." Department of Linguistics, Northwestern University. Unpublished manuscript.

Delattre, P., Liberman, A.M., Cooper, F.S., & Gerstman, L.J. (1952). "An experimental study of the acoustic determinants of vowel color: Observations on one- and two-formant vowels synthesized from spectrographic patterns." Word, 8, 195-210.

Dickerson, L. (1975). "The learner's interlanguage as a system of variable rules." TESOL Quarterly, 9, 401-408.

Dickerson, L. & Dickerson, W. B. (1977)."Interlanguage Phonology: Current research and future directions" in S. P. Corder & E. Roulet (eds.) The Notions of Simplification,Interlanguages, and Pidgins and their Relation to Second Language Pedagogy. Faculte des lettres, Neuchatel.

Eckman, F. (1977). "Markedness and the contrastive analysis hypothesis." Language Learning, 27, 315-330.

Flege, J.E. (1980). "Phonetic approximation in second language acquisition." Language Learning, 30, 117-134.

Flege, J. E. & Davidian, R. D. (1984). "Transfer and developmental processes in adult foreign language speech production." Applied Psycholinguistics, 5, 323-347.

Flege, J. E. & Hammond, R. M. (1982). "Mimicry of non-distinctive phonetic differences between language varieties." Studies in Second Language Acquisition, 5, 1-17.

Flege, J. E. and Port, R. (1981). "Cross language phonetic interference: Arabic to English." Language and Speech, 24, 125-146.

Gass, S. (1983)."Development of speech perception and speech production abilities in adult second language learners." A revised version of a paper presented at AILA in Lund, Sweden.

Greenberg, C. (1983). "Syllable structure in second language acquisition." CUNY Forum Papers in Linguistics, 9, 41-64.

Guiora, A. Z., Beit-Hallahmi, B., Brennan, R. Dull, C. & Scovel, T. (1972)."The effect of experimentally induced changes in ego states of pronunciation ability in a second language." Comprehensive Psychiatry, 13, 421-428.

Hecht, B. & Mulford, R. (1982) "The acquisition of a second language phonology: Interaction of transfer and developmental factors." Applied Psycholinguistics, 3, 313-328.

Hodson, B. W. & Paden, E. P. (1981). "Phonological processes which characterize unintelligible and intelligible speech in early childhood." Journal of Speech and Hearing Disorders, 46, 369-373.

House, A.S., Williams, C.E., Hecker, M.H.L., and Kryter, K.D. (1965). "Articulation-testing methods: consonantal differentiation with a closed response set." Journal of the Acoustical Society of America, 37, 158-166.

Ingram, D. (1976). Phonological Disability in Children. New York: Elsevier Press.

Johansson, F. A. (1973). Immigrant Swedish Phonology. A study in multiple context analysis. Lund, Sweden: CWK Gleerup.

Kohler, K. (1971)."On the adequacy of phonological theory for contrastive studies" in G. Nickel, (ed.) Papers in Contrastive Linguistics. Great Britain: Cambridge University Press.

Krashen, S.D., Long, M.A., & Scarcella, R.C. (1979). "Age, rate and eventual attainment in second language acquisition." TESOL Quarterly, 13, 573-582.

Labov, W. (1969). "The study of language in its social context." Studium Generale, 23, 30 - 87.

Ladefoged, P. (1982). A Course in Phonetics. N.Y: Harcourt, Brace, Jovanovich Inc.

Lado, R. (1957). Linguistics Across Cultures. Ann Arbor: University of Michigan Press.

Lenneberg, E.H. (1967). Biological Foundations of Language. New York: John Wiley and Sons.

Lieberman, P. (1980). "On the development of vowel production in young children", in G.H. Yeni-Komshian, J. F. Kavanagh, & C.A. Ferguson, (eds.) Child Phonology (Production). New York: Academic Press.

Liljencrants, J. and Lindblom, B. (1972). "Numerical simulation of vowel quality systems: The role of perceptual contrast." Language, 48, 839-862.

Lobo, F. & Yoshida, K. (1982). "The perceptual acquisition of English phonology by Japanese students." Georgetown University Roundtable.

Locke, J. L. (1983). Phonological Acquisition and Change. Academic Press.

Mack, M. (1982)."Voicing dependent vowel duration in English and French: monolingual and bilingual production." Journal of the Acoustical Society of America, 71, 173-178.

MacKain, K., Best, C. & Strange, W. (1981). "Categorical perception of /r/ and /l/ by Japanese bilinguals." Applied Psycholinguistics, 2, 369-390.

Miyawaki,K., Strange, W. Verbrugge, R., Liberman, A., Jenkins, J., & Fujimura, O. (1975). "An effect of linguistic experience: The discrimintation of /r/ and /l/ by native speakers of Japanese and English." Perception and Psychophysics, 18, 331-340.

Manuel, S. & Krakow, R. (1984). "Universal and language particular aspects of vowel-to-vowel coarticulation." Haskins Laboratories Status Report on Speech Research, SR-77/78.

Martin, S. (1952). "Morphophonemics of standard colloquial Japanese" Supplement: Language Dissertations no. 87, 28, part 2.

Monsen, R. B. (1978)."Toward measuring how well hearing-impaired children speak." Journal of Speech and Hearing Research, 21, 197-219.

Neufeld, G.(1980). "On the adult's ability to acquire phonology." TESOL Quarterly, 14, 285-298.

Oller, D.K. and Eilers, R.E. (1982). "Similarity of babbling in Spanish- and English-learning babies." Journal of Child Language, 9, 565-577.

Ohala, J.J. (1983). "The origin of sound patterns in vocal tract constraints." in P.F MacNeilage, (ed.) The Production of Speech. New York: Springer-Verlag.

Oyama, S. (1976). "A sensitive period for the acquisition of a non-native phonological system." Journal of Psycholinguistic Research, 5, 261-284.

Peterson, G.E. & Barney, H.L. (1952). "Control methods used in a study of the vowels." Journal of the Acoustical Society of America, 24, 175-184.

Richards, J. C. (1971)."A non-contrastive approach to error analysis." English Language Teaching, 25, 204-219.

Rubin, Judith (1983). Static and Dynamic Information in Vowels Produced by the Hearing Impaired. Indiana University Lingusitics Club.

Scholes, R. (1968)."Phonemic interference as a perceptual phenomenon." Language and Speech, 11, 86-103.

Seliger, H. (1978). "Implications of a multiple critical periods hypothesis for second language learning." in W.C. Ritchie, (ed.) Second Language Acquisition Research: Issues and Implications. New York: Academic Press.

Selinker, L. (1982)."Interlanguage." International Review of Applied Linguistics, 10, 209-231.

Singh, S. & Black, J. W. (1966)."Study of twenty-six intervocalic consonants as spoken and recognized by four language groups." Journal of the Acoustical Society of America, 39, 372-387.

Stevens, K., Liberman, A., Studdert-Kennedy, M., & Ohman, S. (1969). "Cross-language study of vowel perception." Language and Speech, 12, 1-23.

Stevens, S.S. (1951) Handbook of Experimental Psychology. New York: John Wiley and Sons.

Stevick, E. (1976). Memory, Meaning, and Method. Rowley, Mass.: Newbury House.

Tahta, S, Wood, M., & Loewenthal, K. (1981). "Age changes in the ability to replicate foreign pronunciation and intonation." Language and Speech, 24, 363-372.

Tarone, E. (1978). "The phonology of interlanguage." in J.C. Richards, (ed.) Understanding Second and Foreign Language Learning. Rowley, Mass.: Newbury House.

Tarone, E. (1980). "Some influences on the syllable structure of interlanguage phonology." International Review of Applied Linguistics, 18, 139-152.

Tarone, E. (1982). "On the variability of Interlanguage systems." Applied Linguistics, 4, 142-163.

Taylor, B. (1975). "The use of overgeneralization and transfer learning strategies by elementary and intermediate students in ESL." Language Learning, 25, 73-107.

Williams, L. (1979). "The modification of speech perception and production in second language learning." Perception and Psychophysics, 26, 95-104.

Wode, H. (1980). "Phonology in Second Language Acquisition." in S.W. Felix (ed.) Second Language Development: Trends and Issues. Tubingen: Gunter Narr Verlag.