PRODUCTION AND PERCEPTUAL REPRESENTATION OF AMERICAN ENGLISH VOWEL SOUNDS BY MONOLINGUAL PERSIAN AND EARLY BILINGUAL AZERBAIJANI-PERSIAN ADOLESCENTS

Naeimeh Afshar
Doctoral School of Applied Linguistics
University of Pannonia, Veszprém
afshar.naeimeh@mftk.uni-pannon.hu

Doctoral (PhD) Dissertation

Production and perceptual representation of American English vowel sounds by monolingual Persian and early bilingual Azerbaijani-Persian adolescents

DOI:10.18136/PE.2022.825

By

Naeimeh Afshar

Supervisor:

Prof. Dr. Vincent J. van Heuven

Multilingualism Doctoral School

Faculty of Modern Philology and Social Science

University of Pannonia

Veszprém, 2022

STATEMENT

This dissertation, written under the direction of the candidate's dissertation committee and approved by the members of the committee, has been presented to and accepted by the Faculty of Modern Philology and Social Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The content and research methodologies presented in this work represent the work of the candidate alone.

Naeimeh Afshar …,…., 2022

Candidate Date

Dissertation Committee:

_____ …,…., 2022

Chairperson Date

_____

First Reader

_____ …,…., 2022

Second Reader Date

The Perception and Production of American English Sounds by monolingual Persian and early bilingual Azerbaijani-Persian Adolescents

Thesis for obtaining a PhD degree in the Doctoral School of Multilingualism of the University of Pannonia

in the branch of Applied Linguistics

Written by Naeimeh Afshar

Supervisor: Prof. Dr. Vincent J. van Heuven

Propose acceptance (yes / no) …….
(supervisor)

…….
(supervisor)

As a reviewer, I propose acceptance of the thesis:

Name of Reviewer: …… yes / no …….
(reviewer)

Name of Reviewer: …… yes / no …….
(reviewer)

The PhD-candidate has achieved ……% at the public discussion.

Veszprém, ……/……. 2022 …….
(Chairman of the Committee)

The grade of the PhD Diploma …..... (……. %)

Veszprém, ……/……. 2022 …….
(Chairman of UDHC)

Table of contents

Chapter 1	**Introduction**	1
1.1	Language situation in Iran	1
1.2	Acquisition of nonnative sounds	2
1.3	Objective of the study	3
1.4	A note on the methodology	4
1.5	Brief comparison of the languages involved	4
1.6	Importance of correct pronunciation as an EFL learning goal	5
1.7	Defining bilingualism	6
1.8	Primacy of vowels	7
1.9	Structure of the dissertation	8
Chapter 2	**Background/literature**	12
2.1	Introduction	12
2.2	Comparing the sound structures of Azerbaijani, Persian and (American) English	14
2.3	Acquisition of third language phonology	17
2.4	Importance of perceptual vowel studies for foreign language learning	19
2.5	Relationship between perception and production of L2 sounds	21
2.6	Language dominance	23
2.7	Research questions and hypotheses	25
Chapter 3	**Language dominance in Azerbaijani/ Persian EFL learners. Analysis of LEAP-Q data**	28
3.1	Introduction	28
3.2	Using the LEAP-Q to establish language dominance	30
3.3	Consistency in perceptual assimilation	36
3.4	Language dominance and PAM consistency	39
3.5	Discussion and conclusions	41
Chapter 4	**Perceptual assimilation Study**	43
4.1	Introduction	43
4.2	Characterization of the vowel systems involved	45
4.3	Methods	47
4.3.1	Materials	47
4.3.2	Participants	48
4.3.3	Procedure	48
4.4	Statistical considerations	49
4.5	Results	51
4.6	Conclusion and discussion	56

Chapter 5	**Mapping perceptual vowel spaces in native and foreign language**	58
5.1	Introduction	58
5.2.	Methods	59
5.2.1	Participants	59
5.2.2	Materials	60
5.2.3	Procedure	61
5.2.4	Data analysis	62
5.3	Results	64
5.3.1	Perceptual representation: centroids and dispersion ellipses	64
5.3.2	Dividing up the vowel space	67
5.3.3	Native and nonnative vowel identification compared in detail	72
5.4	Conclusions and discussion	79
Chapter 6	**Contrastive acoustic vowel analysis**	81
6.1	Introduction	81
6.2	Methods	83
6.2.1	Participants	83
6.2.2	Procedure	83
6.2.3	Statistical analysis	84
6.3	Results	84
6.3.1	Data analysis	84
6.3.2	Location of vowel centroids in F1 by F2 plane	86
6.3.3	Dispersion and overlap of vowel categories in EFL and native AE	88
6.3.4	Vowel duration	89
6.3.5	Inferential statistics for spectral parameters	91
6.3.6	Multivariate analyses	95
6.3.7	Classifying non-native vowels by native models	97
6.4	Conclusions and discussion	100
Chapter 7	**Discussion & Conclusions**	103
7.1	Introduction	103
7.2	Summary of experiments	103
7.3	Answering research questions	105
7.3.1	Perceptual assimilation of English vowels	106
7.3.2	Difference in perceptual assimilation between monolingual and bilingual learners	107
7.3.3	Relationship between language dominance and perceptual assimilation	108
7.3.4	Perceptual representation of AE vowels by monolingual and bilingual EFL learners	109
7.3.5	Difference in perceptual representation of AE vowels between L1 and L2 listeners	110
7.3.6	Native-language interference in perceptual representation	112
7.3.7	Acoustic realization of AE vowels by monolingual and bilingual EFL learners	113
7.3.8	Difference in acoustic realization of AE vowels by L1 and L2 speakers	114

7.3.9	Native language interference in the AE vowel production by EFL learners	115
7.3.10	Predicting incorrect perceptual representation from perceptual assimilation	116
7.3.11	Predicting incorrect AE vowel articulation from perceptual assimilation	118
7.3.12	Correspondence between perceptual representation and production of AE vowels	119
7.4	Insights gained from present research	122
7.5	Limitations and recommendations for future research	124
7.6	Some pedagogical implications	125
	References	128
	Appendices	138
A4.1	Analysis of 22 tokens used in PAM test	139
A4.2	Praat MFC script for stimulus presentation and response collection in PAM test	141
A5.1	Biographic data on 20 native listeners of American English listeners who participated in the control experiment	143
A5.2	Sample oscillograms, spectrograms and formant tracks of synthesized V-test stimuli	144
A5.3	V-test response count for American native listeners	145
A5.4	V-test response count for monolingual Persian listeners	146
A5.5	V-test response count for bilingual Azerbaijani/Persian listeners	147
A6.1	Stimulus materials used to elicit vowel production	148
A6.2A-D	Formant and duration values of vowels produced by monolingual and bilingual English learners in two contexts	149
A6.2E	Formant and duration values of vowels produced by American L1 speakers.	153
A6.3	Confusion matrices of intended vs classified vowels by LDA when trained and tested on non-native tokens	154
A6.4	Confusion matrices of intended vs classified vowels by MLRA when trained and tested on nonnative tokens	155
A6.5	Hierarchical Cluster Scheme for LDA vowel confusions	156
A6.6	Hierarchical Cluster Scheme for MLRA vowel confusions	157
A6.7	Confusion matrices of intended vs classified vowels by LDA when trained on native tokens and tested on non-native tokens	158

List of tables

Table 3.1	Selected LEAP-Q results for bilingual Azerbaijani-Persian and monolingual Persian learners of EFL.	32
Table 3.2	Correlation matrix of eight self-rated performance measures (scales from 0 to 10) for 23 bilingual Iranian participants with Azerbaijani (AZ) as L1 and Persian (PE) as L2.	36
Table 3.3	Overall Response consistency, Goodness rating (on a scale from 1 to 5 = best) and Response latency (ms) on first and second presentation for monolingual Persian and bilingual Azerbaijani/Persian listeners, when assimilating American English vowels to the vowels of Persian (PE) or Azerbaijani (AZ).	38
Table 3.4	Difference scores defined for Azerbaijani (AZ) - Persian (PE) early bilinguals (N = 23).	40
Table 3.5	A correlation matrix of these six variables (non-redundant lower triangle only).	40
Table 4.1A	Perceptual assimilation of eleven vowels of American English to the six vowels of Persian by early monolingual Persian listeners.	52
Table 4.1B	Perceptual assimilation of eleven vowels of American English to the six vowels of Persian by early bilingual Azerbaijani/Persian listeners.	52
Table 4.1C	The results of the bilinguals when instructed to assimilate the English vowels to the nine vowels of Azerbaijani.	52
Table 5.1	Modal vowel response category (with /ɔ, ɑ/ merged) broken down by duration of synthesized vowel for three groups of participants (L1 native listeners, monolingual Persian EFL learners, early bilingual Azerbaijani/Persian EFL learners).	71
Table 5.2	Confusion matrix of all observed responses against modal (‘correct’) response category for 20 American native listeners.	73
Table 5.3	Confusion matrix of all observed responses against AE modal response category for 21 monolingual Persian learners of English listeners as a foreign language.	73
Table 5.4	Confusion matrix of all observed responses against AE modal response category for 27 early bilingual Azerbaijani/Persian learners of English listeners as a foreign language.	75
Table 5.5	Correct (according to L1 AE modal response) vs. confused responses (count plus row percentages) in vowel identification task by three groups of listeners.	76
Table 5.6	Number of responses in each of 11 vowel categories to short vs long vowel duration in synthesized stimuli accumulated across all 43 vowel quality differences, broken down by language background of the listener.	78

Table 6.1	Number of vowel tokens suitable for statistical analysis broken down by gender of speaker and by Language background.	85
Table 6.2	Summary of RM-ANOVA.	92
Table 6.3	Percent correct classification by Linear Discriminant Analysis and by Multinomial Logistic Regression Analysis of 10 American English vowels produced by four groups of Iranian learners of English as a foreign language, and for all groups combined.	96
Table 6.4	Percentage of correct vowel identification by Linear Discriminant Analysis with spectral parameters F1, F2 or with spectral parameters plus vowel duration.	97
Table 7.1	Summary of perceptual assimilation of AE vowels to Persian and Azerbaijani, by monolingual and bilingual EFL learners.	106
Table 7.2	Crosstabulation of discrimination difficulty predicted by PAM (in rows) against confusion in perceptual identification (%, in columns) of 11 American English monophthongs by monolingual Persian and by early bilingual Azerbaijani/Persian EFL learners.	117
Table 7.3	Crosstabulation of discrimination difficulty predicted by PAM (scenarios in rows) against confusions (%, in columns) in automatic classification by LDA of 11 American English monophthongs produced by monolingual Persian and by early bilingual Azerbaijani/ Persian EFL learners.	118
Table 7.4	Discrepancy (in mean squared Euclidean distance in z-transformed F1-by-F2 (Barks) plane) between perceptual representation and production data of nine American English vowels (excluding /ɔ/ and /e/, see text) measured for three groups of speakers (all nine combinations).	121
Table 7.5	Correctly identified vowel tokens (%) by Linear Discriminant Analysis trained on American L1 vowel tokens for nine speaker groups.	126
Table A4.1	Stimulus analysis of 22 vowel tokens used in PAM test. F1, F2 (Hz) and duration (ms) of eleven vowel tokens produced by two male American native speakers in /h..d/ context.	139
Table A5.1	Biographic data on 20 native listeners of American English listeners who participated in the control experiment described in Chapter 5.	143
Table A5.3	Number of responses given to 86 synthesized vowel stimuli by 20 native listeners of American English broken down by vowel duration (short = 200 ms, long = 300 ms) and by formant frequencies (F1 and F2 in Hz).	145
Table A5.4	Number of responses given to 86 synthesized vowel stimuli by 21 monolingual Persian learners of American English broken down by vowel duration (short = 200 ms, long = 300 ms) and by formant frequencies (F1 and F2 in Hz).	146
Table A5.5	Number of responses given to 86 synthesized vowel stimuli by 27 early bilingual Azerbaijani/Persian learners of American English broken down by vowel duration (short = 200 ms, long = 300 ms) and by formant frequencies (F1 and F2 in Hz).	147

Table A6.1	List of stimulus vowels in common keywords (A) and in /hV(r)d/ carrier (B)	148
Table A6.2	A. Mean, standard deviation, range, minimum and maximum for F1, F2 (Hz) and duration (ms) of AE vowels produced by monolingual Persian EFL learners, aggregated and broken down by gender of speaker. Targets were everyday /CVd/ keywords.	149
	B. As Table A6.2A for /hVd/words	150
	C. As Table A6.2A but for bilingual Azerbaijani/Persian EFL learners	151
	D. As Table A6.2C for /hVd/ words	152
	E. As Table A6.2B but for American native speakers	153
Table A6.3	Confusion matrices for intended vowel by predicted vowel. Automatic classification by LDA with leave-one-out cross-validation. The left part of the table uses two predictors (F1, F2), the right part adds vowel duration as a third predictor.	154
Table A6.4	Confusion matrices for intended vowel by predicted vowel. Automatic classification by Multinomial Logistic Regression Analysis.	155
Table A6.7	Confusion matrices for intended vowel by vowel predicted by model trained on native American vowel tokens.	158

List of Figures

Figure 2.1A	IPA vowel diagrams for the vowel inventories of Modern Persian (A, Majidi & Ternes, 1999).	14
Figure 2.1B	IPA vowel diagrams for the vowel of Azerbaijani (B, Ghaffarvand Mokari & Werner, 2016)	14
Figure 2.1C	IPA vowel diagrams for the vowel of American English (C, modified from Manell, Cox & Harrington, 2009).	14
Figure 4.1A	IPA vowel diagrams for the vowel inventories of Modern Persian (A, Majidi & Ternes, 1999).	45
Figure 4.1B	IPA vowel diagrams for the vowel of Azerbaijani (B, Ghaffarvand Mokari & Werner, 2016)	45
Figure 4.1C	IPA vowel diagrams for the vowel of American English (C, modified from Manell, Cox & Harrington, 2009).	45
Figure 4.2	Screens showing the six response categories (in Arabic script) for the Persian version of the perceptual assimilation test (panel A, left). Panel B (right) shows the screen used for the Azerbaijani version of the test, with nine response categories in Azerbaijani orthography.	49
Figure 5.1	Steady-state F1 and F2 values for reference vowels.	61
Figure 5.2	User interface for vowel identification experiment.	62
Figure 5.3	Centroids and dispersion ellipses (± 1 SD) in an F1-by-F2 plane (axes in Barks) for short (panel A) and long (panel B) stimulus vowels, as perceptually labelled by 21 monolingual Persian learners of English.	64
Figure 5.4	Centroids and dispersion ellipses (± 1 SD) in an F1-by-F2 plane (axes in Barks) for short (panel A) and long (panel B) stimulus vowels, as perceptually labelled by 27 bilingual Azerbaijani/Persian learners of English.	65
Figure 5.5	Centroids and dispersion ellipses (± 1 SD) in an F1-by-F2 plane (axes in Barks) for short (panel A) and long (panel B) stimulus vowels, as perceptually labelled by 20 American native listeners.	66
Figure 5.6	Mean duration (ms) of 11 American English vowel types identified in synthesized vowel stimuli, with separate lines for the three participant groups (N = 21 for monolingual EFL learners, 27 for bilingual EFL learners and 20 for native control listeners). Error bars represent the 95% confidence of the mean.	67
Figure 5.7	Modal responses by 20 American native listeners for 43 vowel stimuli differing in F1 (vertically) and in F2 (horizontally) center frequencies.	68
Figure 5.8	Majority responses by 21 Persian learners of English for 43 vowel stimuli	69
Figure 5.9	Majority responses by 27 early bilingual Azerbaijani/Persian learners of English for 43 vowel stimuli.	70
Figure 5.10	Vowel confusion structure of eleven American English monophthongs as identified for 86 synthesized vowel sounds by 20 American native listeners (A) and by 21 monolingual Persian EFL learners (B).	74
Figure 5.11	Vowel confusion structure of eleven American English monophthongs as identified for 86 synthesized vowel sounds by 20 American native listeners (panel A) and by 27 bilingual Azerbaijani/Persian EFL learners (panel B).	76

Figure 6.1 Centroids of the eleven American English monophthongs in an F1 by F2 plane (axes in Barks) as produced in /hVd/ items by monolingual Persian (left, panels A, D) and bilingual Azerbaijani/Persian (mid, panels B, E) adolescent learners of English as a foreign language, broken down by gender of the speaker (upper: male, panels A, B; lower: female, panels D, E). 86

Figure 6.2 Panel A: as Figure 6.1 but averaged over the four Iranian speaker groups. Panel B: as Figure 6.1 but averaged over the male and female native speakers of American English. 87

Figure 6.3 Centroids and dispersion ellipses for eleven American English monophthongs produced by monolingual and bilingual groups of EFL learners (/hVd/ items only), broken down by gender. Ellipses are drawn at ± 1 SD along the first two principal components of the scatter clouds. The right-most panels represent the control data produced by 10 male and 10 female native speakers of American English. 88

Figure 6.4 Duration (ms) for 11 monophthongs of American English produced by male and female monolingual and bilingual adolescent learners. Error bars are the 95% confidence limits of the mean. Vowels are arranged in ascending order of duration as found for all EFL speakers combined. 89

Figure 6.5 Duration (ms) of 11 AE target vowels, averaged over all four groups of EFL learners (red circles, N = 45) and over native speakers of American English (green squares, N = 20). 90

Figure 6.6 F1 center frequency (in Barks) of ten American English monophthongs produced by male (lower panels) and female (upper panels) monolingual Persian (right-hand panels) and early bilingual Azerbaijani/Persian (left-hand panels) EFL speakers. 93

Figure 6.7 Center frequency of F2 (Barks) for 10 English monophthongs pronounced in /hVd/ words and in rhyming everyday keywords (/CVd/) by Iranian EFL learners, broken down by Gender and by Language background (monolingual Persian vs bilingual Azerbaijani/Persian). 94

Figure 6.8 Vowel duration (ms) for ten American English monophthongs produced by four groups of Iranian learners of English as a foreign language. 95

Figure 6.9 Correct classification (%) of ten American English vowels by Linear Discriminant Analysis trained on native vowel tokens and tested on the same tokens (20 speakers, circles), and on EFL tokens produced by monolingual Persian (21 speakers, triangles) and bilingual Azerbaijani/Persian (24 speakers, squares) learners. 98

Figure 6.10 Vowel confusion structure for classification by LDA of ten American English monophthongs produced by and tested on 20 American native speakers. Predictors were F1, F2 and vowel duration. 99

Figure 6.11 Vowel confusion structure for classification by LDA of ten American English monophthongs produced by monolingual Persian (panel A, left) and for early bilingual Azerbaijani/Persian (panel B, right) learners of English as a foreign language. 100

Figure 7.1 Perceptual representation of the vowel quality (location in F1-by-F2 plot in Barks) of the 11 American English monophthongs entertained by three groups of listeners. 111

Figure 7.2 Location of centroids (F1 and F2 center frequencies, in Barks) of 10 AE vowels produced by monolingual Persian, bilingual Azerbaijani-Persian and American L1 speakers of English. 113
Figure 7.3 Vowel quality (F1 by F2 in Barks) of ten monophthongs of English (excluding /ɔ/) in the perceptual representation and in speech production by 22 Persian monolinguals, 27 Azerbaijani/Persian bilinguals and 20 American native speakers. 120

Figure A4.1 Vowel tokens of Table A1 plotted in the acoustic vowel space defined by F1 (top to bottom, Barks) and F2 (right to left, Barks). 139
Figure A4.2 Duration (ms) of 11 American English monophthongs produced by two male native speakers. Vowel types are plotted from left to right in descending order of the duration realized by speaker 1. 140

Figure A5.2 Oscillograms (amplitude against time), spectrograms and formant tracks (frequency against of time, gray shades represent intensity) of selected synthesized /mVf/ stimuli for Chapter 5. 144
Figure A6.5 Hierarchical tree structures for vowel confusion determined by Linear Discriminant Analysis. 156
Figure A6.6 Hierarchical tree structures for vowel confusion determined by Multinomial Logistic Regression Analysis. 157

Acknowledgement

First of all, I would like to express my sincere gratitude to Prof. Dr. Vincent J. van Heuven for his dedicated support, invaluable feedback, and wise guidance during the running of this project. Prof. van Heuven continuously provided encouragement and was always willing and enthusiastic to assist in any way he could throughout the research project.

Furthermore I would like to acknowledge Khazra high school and Raze Danesh language institute in Marand for their participation and engagement in the study that enabled this research to be possible.

I am indebted also to Prof. Maghsoud Esmaili Kordlar for his collaborative effort during data collection and Prof. Mohammad Sadeq Naebi for providing initial sources at the beginning of the project.

Moreover, thanks are due to the University of Pannonia, especially Prof. Dr. Judit Navracsics, who assisted me in achieving an Erasmus grant as well as stipend from the Új Nemzeti Kiválóság Program (ÚNKP), which provided me with the financial means to complete this project.

Finally, I would like to thank my husband, parents, and numerous friends who endured this long process with me, always offering support and love during the compilation of this dissertation.

Naeimeh Afshar
Pannon Egyetem, Veszprém
6 August 2022

Chapter 1

Introduction

1.1. Language situation in Iran

Nowadays, English is the most extensively studied foreign language throughout the world. Since English has become a dominant international language and various applications in different social media (specifically Twitter, Facebook and Instagram) have recently emerged, the number of motivated learners has increased steadily.

Iranian people are no exception to this phenomenon and many of them, especially students, attempt to learn English at a high level so that they may interact with foreigners all over the globe. Iran (also known as Persia) is located in Western Asia and according to the UN Population Division, has a population of about 83 million inhabitants (Gerland et al., 2019). Iran is bordered in the northwest by Armenia, the Republic of Azerbaijan, and the Azerbaijani exclave of Nakhchivan; in the north by the Caspian Sea; in the northeast by Turkmenistan; in the east by Afghanistan and Pakistan; in the south by the Persian Gulf and the Gulf of Oman; and in the west by Turkey and Iraq. The main and official language of the country is Persian (also called Farsi) but since Iran is a multicultural country comprising numerous ethnic and linguistic groups such as Persians, Turks, Gilaks, Kurds, Lors, Armenians, Arabs, Baluchis, Turkmen, Assyrians, and Georgians (among others), many other languages are spoken among these groups in distinct areas of the country as well, so that bilingualism and multilingualism are common phenomena.

According to Aghazada et al. (2021), 30-35 million of the population of Iran is the Azerbaijanian Turks. Iranian Azerbaijanis are a Turkic-speaking people of Iranian origin, who live mainly in the northwestern historic region of Azerbaijan (i.e., Iranian Azerbaijan). Due to their historical, genetic and cultural ties to the Iranians, Iranian Azerbaijanis are also often associated with the Iranian peoples. They are the second largest ethnicity in Iran as well as the largest minority group. For Azerbaijani/Persian bilinguals, English counts as a third or foreign language. Therefore, studying the sound structure of English by bilingual Azerbaijani/Persian learners is an interesting project to carry out. The current study will make an effort to answer the question: does early bilingualism influence the acquisition of the sound structure of English as a third language?

1.2. Acquisition of non-native sounds

When we learn to speak a foreign language after the age of puberty, the way we pronounce the sounds of the foreign language is generally reminiscent of the sounds of our native language: we speak the foreign language with a specific foreign accent. A strong foreign accent will compromise the efficient decoding of the message and increase the risk of communication breakdown (e.g., Trofimovich & Baker, 2006; Munro & Derwing, 2008; Cutler, 2012). Most courses on English as a foreign language (EFL) contain modules which aim to improve the learner's pronunciation of English, i.e., get the learner to pronounce the English sounds more like a native speaker of English would pronounce them. There is ongoing debate among experts on the question how native-like the foreigner should pronounce the English sounds (Celce-Murcia et al., 2010; Morley, 1991; Walker, 2001, but it is generally agreed that all contrasts with a high functional load should be properly made by the foreign speaker (Jenkins, 2000, 2002; Howlader, 2010). High functional load means that there are many minimal pairs that depend on the particular contrast (e.g., Brown, 1991). If a contrast is needed to differentiate only between a handful of minimal pairs, missing the contrast will not impede the speaker's intelligibility.[1]

It is not the case that any particular sound (or contrast between two sounds) in English constitutes a learning problem *per se*. The sound system of the learner's native language determines which English sounds will be difficult to pronounce and which ones will be easy. A systematic comparison of the sounds and sound structures of the learner's native language (L1) and the target language (L2, here English) may be used to generate predictions of which particular sounds in the target language will be difficult and which ones will be easy. The relevance of contrastive analysis for the acquisition of a non-native language has remained unchallenged since the original proposal was made by Lado (1957) in his Transfer Theory, although the ideas have diversified considerably in more recent decades, which saw the light of – among others – the Speech Learning Model (SLM, Flege, 1987, 1995), the Perceptual Assimilation Model (Best, 1995; Best et al., 2001), the Second Language Linguistic Perception model (L2LP, Escudero, 2005), and the Markedness Differential Hypothesis (Eckman, 1977, 1985). These models and their more recent developments will be discussed in some detail in chapter 2 of this study.

[1] The concept of functional load was introduced by the Prague School of Linguistics, and later taken over by American structuralists (Hockett, 1955).

1.3. Objective of the study

Accurate pronunciation of English sounds is often a problem for foreign-language learners. It has been argued that speech sound differences between L1 and L2 are an important source of pronunciation problems (e.g., Wang (2007) and references therein, see also next section). For instance, articulating vowels such as schwa /ə/ is a problem for monolingual Persian speakers due to the absence of schwa /ə/ in Persian (so that a full vowel /e/ will be substituted in Persian-accented English). This may be different for Azerbaijani/Persian bilinguals since Azerbaijani has an unrounded high vowel /ɯ/ that might be a more reasonable substitute for schwa. Foreign languages are typically acquired by adolescents or young adults in a school setting, after the age of 12, i.e., after that the acquisition of the first language has been completed. When one has learned a first language, speech sounds in foreign languages are typically perceived in terms of the (phoneme) categories of the learner's native language. These native categories were shaped during the first 12 months after birth (e.g., Kuhl & Iverson, 1995). If a learner can no longer perceive the difference between a foreign sound and its nearest equivalent in his native language, it will be very difficult to learn the correct pronunciation of the foreign sound. Nevertheless, there are indications that at least some adults are able to learn to pronounce a foreign language in a way that cannot be distinguished from that of born and bred native speakers, despite the fact that the learning process did not involve early L2 exposure (e.g., Bongaerts, Van Summeren, Planken & Schils, 1997).[2]

It is the primary purpose of this study to see how English learners with Azerbaijani and/or Persian as their native language pronounce and perceive the vowels of English, and to compare this to the sound structure of the native languages. We will test the hypothesis that the perception and production of the English vowels will reflect properties of the vowels in either Azerbaijani, Persian or both, and that the Azerbaijani or Persian influence will be stronger or weaker depending on the dominance of Azerbaijani over Persian in the bilingual learner. The performance of these early bilingual learners will be compared with that of monolingual Persian learners of English (matched for age, gender and education) in order to establish whether English as a third language is easier than learning English as a second language.

[2] In Bongaerts et al. (1997) the English learners' native language was Dutch, which is closely related to English. It is unclear at this time if native pronunciation of English can be attained by learners whose native language is not related to English.

1.4. A note on the methodology

The research will be carried out to discover primarily what mental conception monolingual Persian and early bilingual Azerbaijani/Persian learners of English have of the English vowel system in terms of vowel quality (color) and vowel duration compared with native speakers of American English. To estimate the relative strength of the two languages, used by the early bilinguals and monolinguals in my study, the LEAP-Q (Language Experience and Proficiency Questionnaire) is administered to estimate language dominance of the participants in each group.

I would have liked to also compare the performance of monolingual Azerbaijani leaners of English as a control group but unfortunately that is not possible. There are monolingual Azerbaijani speakers (with an old and ancient accent) to be found in the villages around Marand (located in North-West of Iran) but these are mainly above 60 years old and have no knowledge of English (nor do they speak or understand Persian). Finding monolingual Azerbaijani speakers in the country of Azerbaijan is also impossible since all Azerbaijanis are bilingual as well (in Azerbaijani and Russian); pure Azerbaijani monolinguals have the same problem as the Iranian monolingual Azerbaijani speakers: they are in the age bracket over 60 and never learned English.

1.5. Brief comparison of the languages involved

In Chapter 2, a comparison between English, Azerbaijani, and Persian syllable structures and sound systems will be made. As a result of this comparison, the problematic areas that may be responsible for pronunciation difficulties of bilingual Azerbaijani/Persian speakers and monolingual Persian speakers of English will be identified. In order to understand the role of the first language (L1), emphasis will be given on studies that have focused on the differences between English, Azerbaijani and Persian phonological systems.

Regarding the difference between Turkish and Azerbaijani, Salehi and Neysani (2017) state that Azerbaijani and Turkish are typologically, genealogically and geographically close languages within the Öguz branch of the Turkic languages. Due to many factors, both linguistic and nonlinguistic, these languages have been expected to enjoy a high degree of mutual intelligibility. In this regard, Öztopcu (1993), by comparing the most prominent features of Turkish and Azerbaijani including basic linguistic features such as: phonology, morphology, vocabulary and syntax, has identified differences and similarities between these languages which all lead to an expectation of a strong intelligibility between these languages. Öztopcu concludes that differences between the two languages are not that numerous.

In addition to similar linguistic features as a cause for raising the potential level of intelligibility level, there are also some extra-linguistic reasons which might lead to strengthening this mutual understanding. The most important source of exposure to the Turkish language is the Turkish TV programs in Iran and Azerbaijan. Turkish satellite TV programs are very popular among Azerbaijanis, whether in the republic of Azerbaijan or living in northwestern of Iran.

There is a lack of research in the field of sound structures of English as a foreign language as acquired by Azerbaijani/Persian bilinguals in Iran. Therefore, I decided to address this subject and study it in detail. In order to understand the role of the L1 in the phonological acquisition of the L2, emphasis has been given to the studies that have focused on the similarities and differences between the phonology of Azerbaijani and English phonological system as well as between Persian and English.

1.6. Importance of correct pronunciation as an EFL learning goal

Flege (1988) states that pronunciation is a crucial element of human interaction because speech carries affective and social meaning in addition to referential meaning. Flege argues that people seldom speak their own native language with an accent they themselves judge to be unacceptable. However, many individuals speak a foreign language with an undesirable accent, or hear their native language spoken with a foreign accent. Moreover, learners with a foreign accent may be unintelligible to a degree that they are often misunderstood, or they may be intelligible but understanding them requires more effort (e.g., Hall, 2007 and references therein). Intelligibility is the most desirable objective for foreign-language learners. In phonetics it is customary to differentiate between intelligibility and comprehensibility (of a speaker or of a spoken message). Intelligibility is related to speech recognition, i.e., the recognition of linguistic units (such as morphemes and/or words) in the order in which they were pronounced by the speaker. Comprehension (or understanding) is the result of a higher-order process in which the meanings of the recognized units and of their order are integrated and the intention of the speaker is reconstructed (e.g., Gooskens et al., 2010; Gooskens & Van Heuven, 2021 and references therein).[3]

[3] There are also useless circular definitions of intelligibility in terms of speech understanding (or comprehension) Intelligibility is the degree to which a speaker can be understood, e.g., Kenworthy (1987: 13) “intelligibility is being understood by a listener at a given time in a given situation.”

1.7. Defining bilingualism

In the case of bilingual learners, it is better to start by defining bilingualism and its role in learning English as a third or foreign language. Some of the definitions related to bilingualism are as follows. According to Webster's dictionary (1961), *bilingual* is defined as 'having or using two languages especially as spoken with the fluency characteristics of a native speaker; a person using two languages especially habitually and with control like that of a native speaker' and *bilingualism* as 'the constant oral use of two languages' (Hamers & Blanc, 1989). Some linguists, such as Bloomfield (1933), defined bilingualism as 'the extreme case of foreign language learning where the speaker becomes so proficient as to be indistinguishable from the native speakers round him.' Haugen (1972) preferred a more lenient definition, namely that bilingualism is the 'knowledge of two languages' regardless of the degree of competence and without any need 'for a bilingual to use both his languages.' So, it can be said that bilingualism is simply the property of a speaker that s/he commands two (instead of one) languages. Typically, one language will be the native language (also mother tongue) while the other language is acquired in a later stage in life. However, two (or even more languages) may also be learned (almost) simultaneously in the early stages of one's life, in which case we may speak of early bilingualism. Bloomfields's definition would apply to the phenomenon which we would call 'perfect bilingualism'. Merrikhi (2011) argues that becoming bilingual is a way of life. Your whole person is affected as you struggle to reach beyond the confines of your first language and into a new language, a new culture, a new way of thinking, feeling, and acting (Brown, 1994). Kluge (2007) agrees with this view that bilingualism is a social phenomenon that occurs as a result of language contact. According to Raymond et al. (2002), bilingualism as both a cognitive and social feature of a person is influenced by the details of the individual's life and also has effects on language education and related domains.

Part of the current study will be to determine the degree of language dominance of Azerbaijani versus Persian on the part of the English learners in my experiments. Language dominance will be determined by administering a questionnaire, in which the participants of my study will be asked to estimate their experience with, exposure to, and time spent on Azerbaijani and Persian in various stages of their life. They will also be asked when (at what age and in which order) they learned each of the two languages.

1.8. Primacy of Vowels

According to the International Phonetic Association (IPA), Azerbaijani has nine vowels: four high vowels /i, y, ɯ, u/, three mid vowels /e, œ, o/, two low vowels /æ, ɑ/, no tense-lax vowel contrast or neutral vowel (schwa), no length contrast and no diphthongs. Azerbaijani word stress is fixed and word-final. Azerbaijani, similar to Turkish, has a symmetrical vowel harmony system. That is, the vowels in the stem (or root) of the words do not alternate, and the suffix vowel(s) agree with the harmonic feature, i.e., [back] and/or [round], of the nearest non-alternating vowel (Clements & Sezer, 1982). Since the nearest non-alternating vowel in the stem determines the suffix form, it is called the trigger, while the vowel(s) in the suffix is/are referred to as the target(s) of the harmony pattern (Gafos & Dye, 2011). The direction of the harmony in Azerbaijani is left to right, i.e., the vowels to the right of the trigger vowel agree with it in terms of the harmonic feature (Walker, 2012).

Persian has an even smaller vowel inventory, also without a tense-lax contrast or schwa. Persian has six monophthongs: /i, e, a, u, o, ɑ/. The structure of this vowel system is typologically common with three degrees of vowel height (high, mid, low), and two constriction places (front, back). Lip rounding is unmarked (back = rounded, front = spread). Persian word stress is stem-final (rather than word-final) and its rhythm is syllable-timed (Windfuhr, 1979: 529). A more detailed discussion related to the vowel system of these languages, including comparative acoustic and perceptual data will be presented in Chapter 4.

The World Atlas of Linguistic Structures (WALS, Haspelmath et al., 2005), vowel systems with 5 or 6 monophthongs are of average size. Languages with 7 to 14 monophthongs are classified as having a rich vowel inventory. American English is usually analyzed with a vowel inventory of 11 monophthongs, which can be split into a long (sometimes called ‘tense’) subset comprizing 7 vowels /i, e, æ, ɑ, ɔ, o, u/ and a short (‘lax’) subset of 4 vowels /ɪ, ɛ, ʌ, ʊ/. Moreover, vowels in English are reduced to either schwa [ə] or [ɪ] in unstressed syllables, while the position of the word stress is governed by complex (quantity-sensitive, and morphologically conditioned) rules and is often unpredictable/exceptional. In contrast to this, the consonant inventory of English, comprizing 17 members, is of average size, and is not more complex than the inventory of either Persian or Azerbaijani, with 24 and 25 consonants, respectively. It follows from these considerations that it will be relatively easy for Azerbaijani and Persian learners of English to find a consonant in their native sound inventory that can be substituted for an English target consonant, without dramatically reducing the EFL speaker’s intelligibility. In terms of vowel inventories, however, Azerbaijani and Persian (the latter even more so) are under-differentiated relative to English, so that adequate substitute

sounds will be difficult to find in the native vowel inventories. We will assume, therefore, that EFL learners with Persian and/or Azerbaijani will benefit most if they learn to improve their production and perception of the vowels of English, rather than improving their consonants. For this reason, the present dissertation concentrates on the production and perception of the vowels of (American) English by monolingual Persian and early bilingual Azerbaijani/Persian learners. Consonants, consonant clusters, and connected speech materials were also recorded in the early stages of the research. These will be kept on record for future work, but will not be analyzed in the present dissertation.

1.9. Structure of the dissertation

The structure of the dissertation is as follows. In **Chapter 2**, I will outline and summarize a number of theories and models of the acquisition of the sounds of a foreign language, and provide a more detailed overview of the sound systems of the languages involved in the research, i.e., Azerbaijani, Persian and English. One model in particular will guide my work. This is the Perceptual Assimilation Model (PAM), from which specific predictions can be derived as to which sounds and sound contrast in a foreign language will constitute a learning problem. The perceptual assimilation task asks listeners (in our case with an Azerbaijani and/or Persian native background) to decide with which vowel in their native language (or languages in the case of the early bilinguals) they identify each of the vowels of (American) English, and to state how good the match is between the foreign and the native sound. Depending on the results of this matching task, the prediction will be that some contrasts between English vowels will be easy to perceive (if each English vowel is matched with a different vowel in the native language) or that the contrast will be more difficult (if the learner matches two different vowels in English with the same vowel category in their native language(s). In the final part of chapter 2, I will describe what materials were recorded from the participants in my study. This description will be limited to the recordings of the vowels. The materials recorded for the consonants, clusters and connected speech, which will not be analyzed in the present dissertation, will be relegated to the Appendix.

Chapter 3 examines in detail the language background of my Persian EFL learners. Using the Language Experience And Proficiency Questionnaire (LEAP-Q, Marian et al. 2007), I will determine the experience and proficiency (by self estimation) of my EFL learners with their native language(s), i.e., Persian and Azerbaijani, as well as with English (and any other language they are familiar with). The results of the LEAP-Q will confirm that the monolingual Persian group has no experience with or proficiency in Azerbaijani, whereas the early bilinguals

are proficient in both languages, even though all of these participants indicate that their learned Azerbaijani (the home language) before they learned Persian, and that their spoken (but not their written) language skills are slightly better in Azerbaijani than in Persian. The difference in scores for Azerbaijani and Persian will allow me to define participant-individual measures of relative language dominance, which I will correlate with each other, and with the consistency with which the participants carried out the perceptual assimilation task in chapter 4.

Chapter 4 describes the perceptual assimilation experiment done with two groups of Iranian adolescent learners of English as a foreign language in Iran. One group comprizes monolingual learners with Persian as their only native language. The second group is composed of EFL learners, whose first native language is Azerbaijani (the home language) but who acquired Persian from the age of 4 onwards in their (pre-)school years. These participants can be considered early bilinguals. The two groups are matched in terms of age an education. In this part of the project, we examine the way the monolinguals and the bilinguals identify the pure vowels of American English as instances of vowels in their native language. The monolinguals match the English vowels only with the vowels of Persian, the bilinguals do the assimilation task twice, i.e., once with the six vowels of Persian and a second time with the nine vowels of Azerbaijani. This experiment serves a dual goal. First, the results will tell us how easy (or difficult) it will be for the listener to notice differences between English vowels. For instance, if two English vowels, such as tense /u/ and lax /ʊ/ are both identified as good or at least acceptable tokens of the listener's native /u/ category, we predict that the /u~ʊ/ contrast will be a problem for the Iranian EFL learner. Azerbaijani has (three) central vowels, where Persian has peripheral vowels only. Familiarity with central vowels may prompt the bilingual EFL learners to match the central /ʌ/ vowel of English to one of the central vowels in their native inventory, so that the English contrast between /ʌ/ and its non-central competitors (e.g., /æ, ɑ, ɔ, ʊ/) will be easy to perceive and learn. Second, the perceptual assimilation task will be performed for each American English vowel token twice (in different random orders) so that we will be able to examine the consistency with which the EFL learners make their decisions. We will subsequently test the hypothesis, for the bilingual EFL learners only, that the degree of task consistency will correlate positively with the relative degree of language dominance of Azerbaijani over Persian. If so, the task consistency can be used in future research as a measure of language dominance in bilinguals.

In **Chapter 5**, I will report an experiment which was set up to map out the perceptual representation of the monophthongs of American English that is entertained by the members of my two groups of EFL learners. Knowing the perceptual representation will allow me to do

two things: (i) check predictions by the Perceptual Assimilation Model about which contrasts between vowels in American English will be compromized (relative to the perceptual representation found for American native listeners), and (ii) derive more specific predictions as to what errors will be found when my participants have to actively pronounce the vowels of American English. The perceptual mapping was done by asking the participants to identify each of 86 synthesized vowel sounds (systematically differing in degree of jaw opening, in backness/lip rounding, and in duration) as one of the 11 vowels of American English, and then comparing the results with the task performance by native listeners. Such perceptual mapping of an entire vowel system has not been attempted very often in the literature. The artificial vowel set we used was developed specifically for the purpose of my project, and can be seen as an innovative research tool. The results will show that the mental representation of the vowels of American English is seriously flawed, and often in ways that are predicted well by PAM. Specifically, American native listeners rely much less on vowel duration as a correlate of the tense-lax distinction than the Iranian EFL learners do. However, no indications will be found that the three extra (central) vowels in the inventory of Azerbaijani offer an advantage over knowing only the six vowels of Persian, for Iranian learners of English as a foreign language.

Chapter 6 reports the results of a large-scale acoustic analysis of the EFL vowels produced by the same individuals who participated in the earlier experiments. I will compare the results of these vowel measurements with data from an earlier study of the vowels produced by American native speakers (Wang & Van Heuven, 2006). The center frequency of the lowest resonance (called first formant of F1) was measured as a correlate of vowel height (degree of jaw opening), while the second lowest resonance (second formant or F2) was measured as a correlate of the degree of backness and lip-rounding. Vowel duration was measured as a third distinguishing property. Again, the EFL proved to be seriously compromized. Lack of acoustic contrast was observed between the high front vowels, the high back vowels, as well as among the members of the low back vowels (including the unrounded, centralized low vowel /ʌ/ as in *but*). Only the low front vowel /æ/ was acoustically distinct from all other vowels. The results confirm several predictions made from the PAM study in chapter 4. Automatic classification of the vowels was performed by two self-learning algorithms, i.e., Linear Discriminant Analysis (LDA) and Multinomial Logistic Regression Analysis (MLRA). It was shown that the F2 was the most successful predictor of the vowel category intended by the speaker, followed by F1, and with vowel duration last. On the basis of the two spectral parameters (F1, F2) correct vowel classification was between 59 and 73% for the LDA method, and between 61 and 75% for the MLRA. Adding vowel duration as the third predictor increased the percentage of correct

classification by 2 to 8 points for the LDA and by 5 to 11 points for MLRA. Ideally, all the EFL vowel tokens should be presented to native listeners for identification, in order to determine which contrasts are and are not properly upheld in the EFL speech. Since it is undoable, in practice, to present as many as 45 (speakers) × 11 (vowel types) × 3 repetitions = 1485 vowels for perceptual identification to a group of American native listeners, the native listeners were simulated through the LDA and MLRA classification algorithms. By training the algorithms with native speaker data, adequate models were obtained for each of the 11 vowels. Forcing the algorithms to classify the EFL vowels subsequently showed which vowel categories in the EFL speech were incorrectly produced. The confusion structure revealed largely the same that was found in the perceptual representation in chapter 3, which strengthens our claim that correct production of L2 vowels presupposes a correct perceptual representation, i.e., correct targets. Finally, the results bear out that the incorrect production of EFL vowels was virtually the same for the monolingual Persian and for the early Azerbaijani/Persian bilinguals, so that – again – the conclusion follows that the additional three central vowels in Azerbaijani offer no advantage over the Persian L1 vowel system for Iranian EFL learners.

Chapter 7 summarizes the main findings of the dissertation, and systematically answers the questions I raised in the introductory chapter. Weaknesses in the experimental setup will be identified, and recommendations for future research will be made.

Chapter 2

Background/literature

2.1 Introduction

It is well known that second language (L2) learners have great difficulty when attempting to learn L2 sounds. This difficulty is clearly observed in the phenomenon commonly known as 'foreign-accented speech', which seems to be characteristic of most adult L2 learners. Typically, adult learners are outperformed by infants and young children when the task is to learn the sounds of a language. That is, every child learns to produce and perceive ambient language sounds resembling adult performance in that language. In contrast, adult learners struggle to acquire native-like performance and commonly maintain a foreign accent even after having spent many years in an L2 environment. This paradoxical situation has sociological consequences since the general abilities of adult L2 learners are commonly judged on the basis of their language skills. Moreover, if their speech is (strongly) 'accented' (close to unintelligible), it may impede communication and even prevent integration into the community of native speakers (Escudero, 2005)

From being a relatively neglected area in the study of second language learning, the acquisition of second language speech has emerged over the last decades as an important research field with a wide range of approaches; the traditions of articulatory, acoustic, perceptual, phonetic, phonological, and psycholinguistic investigation contribute a rare interdisciplinarity to this area of linguistic inquiry (Leather & James, 1991). Moreover, some researchers and scholars have developed theories and models which treat L2 speech acquisition as a subfield within cognitive science.

According to Escudero (2005), the three most influential phonetic models that aim to explain L2 sound perception are Best's Perceptual Assimilation Model (PAM), Kuhl's Native Language Magnet (NLM) model, and Flege's Speech Learning Model (SLM). PAM seeks to account for the observed performance in the perception of non-native sound contrasts. It proposes that adult listeners have no mental representations or mental perceptual mappings for sound perception, and that they directly seek and extract the invariants of articulatory gestures and gestural constellations from the speech signal. This proposal is based on the

frameworks of Articulatory Phonology (cf. Browman & Goldstein, 1989) and the ecological approach to speech perception, also called direct realism (cf. Best, 1984; Fowler, 1986).

On the other hand, Kuhl's NLM model attempts to explain the development of speech perception from infancy to adulthood. It argues that complex neural perceptual maps underlie sound perception and that such neural mappings result in a set of abstract phonetic categories. Adult perception is seen as language specific because it is shaped by earlier linguistic experience (cf. Kuhl, 2000: 11854). Unlike the PAM proposal, NLM claims that what is stored in memory are perceptual representations. Perceptual mappings differ substantially for speakers of different languages so that the appropriate perception of one's primary language is completely different from that required for other languages (cf. Iverson & Kuhl, 1995; Iverson & Kuhl, 1996). Kuhl emphasizes that perception is language specific, claiming that "no speaker of any language perceives acoustic reality; in each case, perception is altered in service of language" (2000: 11852).

As for Flege's SLM, it has been primarily concerned with the ultimate attainment of L2 production (cf. Flege, 1995: 238) though it has recently begun to show an interest in the ultimate attainment of L2 perception (cf. Flege, 2003). SLM aims to predict the abilities of non-native speakers to perceiving or producing L2 sounds. Accordingly, the aim of Flege's research is to understand how speech learning changes over the life span and to explain why "earlier is better" as far as learning to pronounce a second language (L2) is concerned. Flege (1995) makes an assumption that the phonetic systems used in the production and perception of vowels and consonants remain adaptive over the life span, and that phonetic systems reorganize in response to sounds encountered in an L2 through the addition of new phonetic categories, or through the modification of old ones (Escudero, 2005).

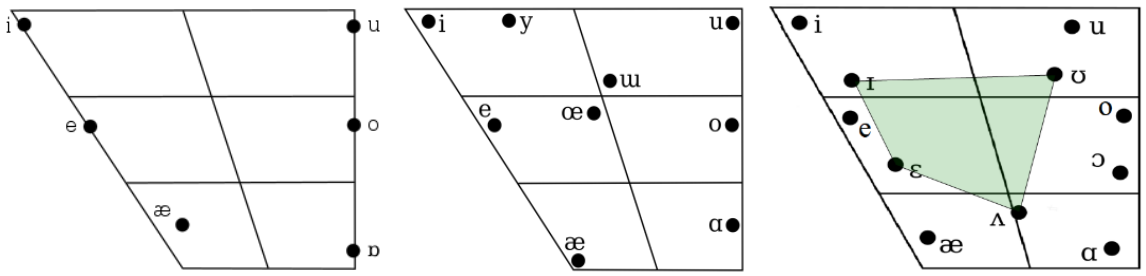
The PAM and the NLM models are mainly interested in explaining the L2 initial state through non-native perception but they still offer suggestions as to how the L2 development and end state can be achieved. SLM mainly deals with the end state but its claims regarding why L2 learners may not have a native-like end state are directly connected to an explanation of the initial and developmental states in L2 sound perception (Escudero, 2005).

In the current study, Best's Perceptual Assimilation Model (PAM) has been used to seek a learning problem in monolingual Persian listeners and their bilingual Azerbaijani peers (see Chapter 4 for more details).

2.2. Comparing the sound structures of Azerbaijani, Persian and (American) English

In the current study, a comparison between English, Azerbaijani, and Persian syllable structures and sound systems has been made. As a result of this comparison, the problematic areas that may be responsible for pronunciation difficulties of bilingual Azerbaijani/Persian speakers and monolingual Persian speakers of English will be identified. In order to understand the role of the first language (L1), attention will be given to studies that have focused on the differences between English, Azerbaijani and Persian phonological systems.

The monophthongal vowel system of Persian is rather straightforward, with three degrees of height (high, mid, low) and two degrees of backness (front, back). Lip rounding is unmarked, i.e., typologically normal, such that front vowels are pronounced with spread lips and back vowels with rounded lips. Persian has no contrast based on vowel duration (short, long) or tenseness (lax, tense). The approximate positions of the six vowels in the IPA vowel chart are shown in Figure 2.1A (copied from Majidi & Ternes, 1999).



A. Persian B. Azerbaijani C. American English

Figure 2.1. IPA vowel diagrams for the vowel inventories of Modern Persian (A, Majidi & Ternes, 1999), Azerbaijani (B, Ghaffarvand Mokari & Werner, 2016) and American English (C, modified from Manell, Cox & Harrington, 2009). The shaded quadrilateral connects the four phonetically lax vowels.

The vowel system of Azerbaijani is almost the same as that of Persian as far as the peripheral vowels (also called edge vowels) is concerned but it is augmented with three vowels in the central region of the vowel space, yielding a total of nine, as shown in Figure 2.1B (copied from Ghaffarvand Mokari & Werner, 2016: 509). The coupling of backness and lip rounding is more complex in Azerbaijani in that the three central vowels have a-typical lip rounding. Phonologically, Azerbaijani /y/ and /œ/ are front vowels (as they are in Turkish) but with the marked presence of lip rounding. The phonologically high /ɯ/ is a back vowel with marked spread lips. Like Persian and Turkish (closely related to Azerbaijani with a fair degree of mutual intelligibility), Azerbaijani has no length or tenseness contrast in the vowels.

The pure (monophthongal) vowels of American English comprise a more complex system than either Persian or Azerbaijani. Although considerable regional variation exists,

varieties distinguish eleven vowels that are normally analyzed as monophthongs, as illustrated in Figure 2.1C, which is based on Manell, Cox & Harrington (2009) and Ladefoged & Johnson (2011: 197). This system has five front vowels (all unrounded) and five back vowels (all rounded, except /ɑ/), with four degrees of height: high, high-mid, low-mid and low. The monophthongs can be split into a group of seven long vowels /i, e, æ, ɑ, ɔ, o, u/, and a smaller group of four short vowels /ɪ, ɛ, ʌ, ʊ/ (Lehman & Heffner, 1940; Peterson & Lehiste, 1960; House, 1961; Wang & Van Heuven, 2006; Celce-Murcia et al., 2010), which not only have shorter durations, but also a rather more centralized vowel quality, and no diphthongization. Because of the more centralized vs peripheral vowel articulation, the short-long contrast is sometimes called lax vs. tense (House, 1961; Celce-Murcia et al., 2010).[4] The long/tense vs. short/lax properties distinguish between the members of the high-mid vowel pairs /e, ɪ/ and /o, ʊ/. There is just one central monophthong: mid-low /ʌ/. The (mid-) low back vowels /ɔ, ɑ/ as in *law* /lɔ/ and *father* /fɑðɚ/ are analyzed as long (tense) vowels. Low front /æ/ is a long vowel in American English (see references above, see also Strange et al. 2004). The high-mid tense vowels /e/ and /o/ are semi-diphthongs in most varieties of English, including American English. I group them with the monophthongs because the slight diphthongization is not essential for their identification, and when pronounced as monophthongs (as they are in some varieties, e.g., Scots English) they remain distinct from each other and from all other vowels – which is not the case for the full diphthongs /ai, ɑu, ɔi/. Here I follow the analysis adopted by, among others, Celce-Murcia et al. (2010: 115-116) and Yavaş (2011: 77–79). Also, in line with his analysis, we exclude all vowels that only occur as positional allophones before coda /r/, such as [ɚ], which is listed among the monophthongs by Ladefoged and Johnson (2011).

In this dissertation, I only consider the vowels of English in stressed syllables. A number of interference phenomena will therefore not be studied. For instance, one important source of L1 interference can be found in the difference in rhythmic structure between languages. Languages can be arranged along a rhythm dimension that ranges between stress-timed and syllable-timed (Abercromby, 1967; Dauer, 1983). In strict syllable timing, every syllable takes up the same amount of time, so that stressed and non-stressed syllables will not differ in duration. In a strictly stress-timed language, the time-intervals between stressed syllables are constant, no matter how many unstressed syllables intervene between two

[4] There is even some support that the so-called tense vowels require greater muscular effort on the part of the speaker (Raphael & Bell-Berti, 1975).

successive stresses. The more syllables there are between stresses, the shorter they are (Lehiste, 1977; Fowler, 1981).

Syllable-timed languages have simple syllable structures such as CV, V, VC and VCV. They have no length contrast, no diphthongs, and no vowel reduction in unstressed syllables. These properties conspire to keep syllables of (roughly) equal length. Stress-timed languages, however, allow complex syllable structures with up to three onset consonants and up to four consonants in the coda. Stress-timed languages may have both short and long vowels as well as diphthongs, and reduce vowel quality in unstressed syllables. Complex syllables are typically stressed, while the simple syllables tend to be unstressed (Dauer, 1983).

English is often mentioned as the prototypical example of a stress-timed language, Turkish and Persian have been classified as syllable-timed languages (Yavaş, 2012: 191 and 204, respectively. Azerbaijani is said to be of a mixed rhythm type, and is 'partially stress-timed'. Its most complex syllable structure is CVC, so that Azerbaijani is probably more syllable-timed than stress-timed.

In English, only the two shortest vowels, /ɪ/ and schwa (/ə/), are permitted in unstressed syllables, while full vowels and diphthongs can only occur in stressed syllables. Pronouncing reduced vowels /ə, ɪ/ in unstressed syllables is a major challenge for any EFL learner. Persian and Azerbaijani EFL learners typically pronounce full vowels in unstressed English syllables, which disrupts the stress-timed rhythm and compromizes word recognition (e.g., Field, 2005). Since vowel reduction is stress-related, i.e., a prosodic phenomenon, it falls outside the scope of the present dissertation.

Regarding the difference between Turkish and Azerbaijani, Salehi and Neysani (2017) state that Azerbaijani and Turkish are typologically, genealogically and geographically close languages within the Öguz branch of the Turkic languages. Due to many factors, both linguistic and nonlinguistic, these languages have been expected to enjoy a high degree of mutual intelligibility. In this regard, Öztopcu (1993), by comparing the most prominent features of Turkish and Azerbaijani including basic linguistic features such as: phonology, morphology, vocabulary and syntax, has identified differences and similarities between these languages which all lead to an expectation of a strong intelligibility between these languages. Öztopcu concludes that differences between the two languages are not that numerous.

In addition to similar linguistic features as a cause for raising the potential level of intelligibility level, there are also some extra-linguistic reasons which might lead to strengthening this mutual understanding. The most important source of exposure to the Turkish language is the

Turkish TV programs in Iran and Azerbaijan. Turkish satellite TV programs are very popular among Azerbaijanis, whether in the republic of Azerbaijan or living in north-western of Iran.

There is a lack of research in the field of sound structures of English as a foreign language as acquired by Azerbaijani/Persian bilinguals in Iran. Therefore, I decided to address this subject and study it in detail. In order to understand the role of the L1 in the phonological acquisition of the L2, emphasis has been given to the studies that have focused on the similarities and differences between the phonology of Azerbaijani and English phonological system as well as between Persian and English.

2.3. Acquisition of third language phonology

The great majority of (experimental) studies on the acquisition of the phonetics and phonology of a foreign language have been done on the assumption that the target language is the second language the learner tries to acquire. This assumption is probably valid in many cases, especially when the learner's native language is one of the major global languages such as Mandarin, Hindi, English or Spanish – for whom learning one foreign language is typically the maximum the school curriculum offers. However, the assumption is more often incorrect, since the majority of the world's population is bilingual or even multilingual. This makes it a legitimate question to ask whether knowing more than one language is or is not an advantage when learning yet another language in a later stage of life.

Until recently most research done on the possible advantage of knowing multiple languages in third (or later) language acquisition (TLA), was concentrated on the acquisition of reading and writing skills, where correct use of vocabulary, morphology and syntax are of primary importance. The general finding in these domains was that learning is faster and more effective in TLA than in SLA. Spoken language skills in TLA, however, have not become a research topic until the beginning of the 21st century (e.g., Beach et al., 2001).

In TLA, one can ask quite generally whether knowing multiple languages gives the learner an edge when acquiring yet another language, or whether the advantage is limited to only specific features of the earlier languages that can transfer positively to the new language. One line of research has been to compare the speed and accuracy with which monolinguals, bilinguals and multilinguals learn to discriminate between the sounds of a new language (unknown to the learner). Early studies suggest that bilinguals (and multilinguals) enjoy a general cognitive advantage when learning a new language, as well as better auditory sensitivity to unfamiliar sound contrasts, but no language-specific transfer phenomena were observed (see Tremblay & Sabourin, 2012 for a literature review). Other, more recent, studies

nevertheless show that bi- and multilinguals were more sensitive to specific sound contrasts in an unknown language, only if at least one of their additional languages, but not the language they shared with monolingual controls, employed the specific type of sound contrast. Beach et al. (2001) report that monolingual British-English listeners could only discriminate between voiceless and aspirated plosives in Thai, whereas bilingual Greek/English listeners could also discriminate the Thai prevoiced vs voiceless contrast. This suggests that experience with specific contrasts in both languages of the bilinguals were used ('narrow transfer') when they were asked to discriminate the Thai prevoiced-voiceless-aspirated plosives. At the same time, this study does not rule out the alternative explanation that that the bilinguals simply benefited from the general bilingual advantage ('broad transfer). A more complete experiment was done by Patihis et al. (2015), who tested monolingual English listeners, early bilingual English/Spanish and English/ Armenian listeners, and early trilinguals with English and two out of 14 different other languages, two of which have a ternary VOT distinction (voiced, voiceless, aspirated) in the plosives, i.e., Armenian (*n*=3) and Thai (*n*=1) while the other 12 languages have a binary contrast, e.g., Persian (*n*=5), Spanish (*n*=4), French (*n*=4), Russian (*n*=2), Ukranian (*n*=2). Monolingual English and early bilingual English/Spanish participants discriminated the ternary VOT contrast in the test language (Korean) at 66% and 63% correct, respectively, which contradicts the idea of a general bilingual advantage.[5] Early bilingual English/Armenian participants did significantly better (74%), as did all trilinguals lumped together (74%). Apparently, familiarity with the Armenian ternary VOT contrast transferred positively to Korean. However, since only 4 of the 14 trilinguals had a ternary VOT opposition in one of their languages, the other 10 trilinguals must have benefited from an overall trilingual benefit. This suggests that knowing more than two languages yields a 'broad' benefit for perceiving unfamiliar sound contrasts, but that 'narrow' transfer is obtained even for bilinguals. In fact, the contribution of broad and narrow transfer seems additive, given that the 14 English/Armenian bilinguals and the 4 trilinguals with the ternary VOT contrast in one of their languages, scored best of all: 76% correct.

[5] Pahitis et al. (2015) assume that English has no voiceless vs. aspirated contrast in its stop consonants, and therefore is like Spanish. This assumption is wrong. Spanish uses pre-voiced vs voiceless, while English uses voiceless vs aspirated (e.g., Lisker & Abramson, 1964, 1970; Delattre, 1965; Docherty, 1987). What is crucial here is that early English/Spanish bilinguals appear unable to integrate the two binary contrasts into one ternary VOT contrast. Early bilinguals with Armenian as one of the languages, however, positively transfer the ternary Armenian VOT contrast to Korean. This suggests that contrast systems transfer as a whole, but denies the creation of new multivalued contrasts by combining two existing binary contrasts in separate languages. This, again, conflicts with Beach et al. (2001) who showed that English/Greek bilinguals were able to unify two binary VOT contrasts into one ternary contrast in their discrimination of Thai stops.

As shown in Figure 2.1, the Azerbaijani and Persian vowel systems have approximately the same six vowels along the front and back edges of the IPA diagram. However, Azerbaijani has three additional vowels in the (mid) high centre portion of the vowel space. One of these vowels, /ɯ/, might be a reasonable substitute for American English lax /ʊ/, which our early bilingual EFL learners have recourse to in one of their native languages. Moreover, the existence of three central vowels in the Azerbaijani system may help the bilinguals to grasp the distinction between central /ʌ/ and its front and back neighbours in the AE vowel space. Since our participants are either monolinguals or early bilinguals but not early trilinguals, we expect 'narrow' transfer from the additional central vowels of Azerbaijani, but no 'broad' multilingual advantage.

2.4. Importance of perceptual vowel studies for foreign language learning

In studies on the phonetics of vowel systems, the usual procedure is to record a number of speakers of the language variety of interest and then measure the lowest two to four resonances in the speaker's vocal tract as an indication of how each vowel is pronounced in terms of tongue height and backness. Here, the center frequency of the lowest resonance, called first formant or F1, corresponds to the openness of the vowel (openness is inversely related to vowel height), while the second-lowest resonance (F2) is an indication of the vowel backness (Ladefoged & Johnson, 2010; Ladefoged & Disner, 2012). More precisely, the F2 reflects the length of the oral tract, which is determined not only by the constriction point (where the body of the tongue most closely approximates the palate or backwall of the throat) but also by lip protrusion (or rounding). Measuring the F1 and F2, and plotting the coordinates in a two-dimensional map, then gives a good impression of the general organization of the vowel system. Since individual speakers have different shapes and sizes of their vocal tracts and of the cavities therein, there is considerable variability in the exact location of the vowels on such maps. In practice, the mean (also called centroid) of the dispersion cloud of each vowel is taken as the most representative or typical realization of the particular vowel type. Vowel duration is often added as a third parameter to define the vowel space of the language (variety). Such representations of the vowel system in a universal vowel space are an important tool in the teaching of the pronunciation of a language to non-native learners.

By comparing the system of the target language with the learner's native language, differences and similarities in the organization of the respective vowel systems can be illustrated, potential learning problems can be identified, and specific instructions can be formulated to explain to the learner how s/he should modify the native vowel category so as

to articulate a more authentic vowel in the target language. It is insufficiently realized in the teaching of the phonetics of foreign languages that studying the acoustics of the vowel systems per se does not reveal the full organization of a vowel system, and – more importantly – does not reveal the (often incorrect) perceptual representation of the vowel system of the target language. What is needed to appreciate the representation of the vowel system in the mind of the learner (and of the native speaker) is a perceptual mapping. Using perceptual techniques allows the researcher to establish so-called trading relationships between the parameters that define the individual listener's vowel space.

As a case in point, Van Heuven (1986) studied the mental representation of the vowel system of Dutch with native Dutch listeners and with Turkish immigrants who had lived in the Netherlands for eight years or longer. Dutch contrasts tense and lax vowels in pairs, the members of which are rather close to one another in the spectral space but differ in duration by a 2-to-1 ratio. In one vowel pair, /a/ is articulated as a long front vowel, which is contrasted with a short back vowel /ɑ/. For native Dutch listeners, a short low vowel is still perceived as the tense (long) counterpart if the articulation is fronted (i.e., by raising the F1 and F2 frequencies). Conversely, a front [a] is identified as the lax /ɑ/ for durations below 150 ms. So, in the Dutch /ɑ/~/a/ contrast, backness (as cued by low F1 and F2 values) and duration are in a trading relationship, i.e., insufficient backness for /ɑ/ can be compensated by extra short vowel duration, and insufficient fronting for /a/ can be compensated for by adding length. For the Turkish learners of Dutch no such trading (or compensation) was found: these participants only used the vowel duration in their decision whether the vowel was tense or lax; the quality difference between /a/ (approximately cardinal vowel #4) and /ɑ/ (approximately cardinal vowel #5) played no role whatsoever. As a result of the incorrect mental representation of the /a/~/ɑ/ contrast, about 50% of the vowel tokens were incorrectly classified by the Turkish learners of Dutch. It was argued that the incorrect mental representation was also the reason why the Turkish L2 speakers of Dutch did not differentiate between the two vowels in their speech production.

In this dissertation, I will test the predictions of learning difficulties based on the results of the perceptual assimilation test (Chapter 4) on the results of an experiment (Chapter 5) in which I asked monolingual Persian and bilingual Azerbaijani/Persian learners of English to identify each of 86 artificially generated sample vowels as tokens of the 11 vowel types of American English. The results of the identification task will inform us of the relative importance of vowel quality and vowel duration in the perceptual representation of the English vowels entertained by the learners. Then, in Chapter 6, I will acoustically analyze the

AE vowels that were produced by the same Iranian (monolingual and bilingual) EFL learners, and match the degrees of acoustic contrast (or the lack of it) among the 11 vowels with the predicted learning difficulties.

2.5. Relationship between perception and production of L2 sounds

When it comes to learning spoken language, whether a first language or a second language, perception necessarily precedes production. The precedence of perception starts even before birth. When still *in utero*, the human embryo develops a working hearing system in the last three months of gestation that allows it to resolve sounds produced by the mother or which pass through from outside the mother's body (Querleu et al. 1988). Ample experimental evidence shows that, immediately after birth, human infants are able to distinguish their mother's voice from other voices (e.g., DeCasper & Fifer 1980, Querleu et al. 1984, Hepper, Scott & Shahidullah 1993), and recognize melodies and rhythms of the mother's language as different (less attractive) than melodies and rhythms of other, unrelated languages (e.g., DeCasper & Spence 1986, Ramus, Nespor & Mehler 2000, Ramus 2002). There are also indications that the way newborn infants cry is related to the basic melody of the mother's language (Mampe, Friederici, Christophe & Wermke 2009). There are no indications that unborn infants have an occasion to practice the articulation of sounds in the mother's womb, so that the conclusion follows that the language-specificity in the infant cries is caused by the infant's trying to approximate perceptual targets that were established by listening *in utero*.

The primacy of establishing perceptual targets before being able to pronounce the sounds of a language can be generalized to the learning of a second (or later) language. Most of the L2 sound acquisition models base their prediction of learning problems on the difficulty to perceive differences between the sounds of the L2 target language and the learner's native language. Flege's SLM (see § 2.1) assumes that the L2 sounds are classed as equivalent to the sounds of the native language, and that, over time, the learner discovers that the assumption of equivalence is incorrect in some cases, and then sets up 'new' sound categories. Alternatively, the differences between the L2 sound and the closest L1 sound are so subtle that they escape the learner's attention, so that no new category is formed but the L1 and L2 categories are merged ('similar' sounds). These processes are basically perceptual in nature. It hardly ever happens (outside a school context) that native interactants of the target language point out to the learner that the realization of the target sound is (slightly) incorrect. Escudero's LPL2 model assumes that the first stage of L2 acquisition is that the learner considers the phonologies of the L1 and L2 to be identical, and gradually discovers that new boundaries have to be instated, or that

existing boundaries have to be shifted ('optimized') to accommodate the target language. Similarly, PAM predicts L2 learning difficulties from the learner's (in)ability to perceptually map the foreign sounds onto existing categories in the native language.

At the same time, however, PAM embraces the 'direct-realist' approach to speech perception (based on Fowler 1986), which claims that the listener (and L2 learner) does not just perceive sound patterns but also intuitively knows the articulatory structures and dynamics by which the sound phenomena are produced.[6] In this view, hearing a difference between sound categories is (almost) the same as knowing how to produce the difference between the sounds, which renders it virtually impossible to test the hypothesis that perception precedes production. The recent SLM-r (r for revision) abandons the idea that (improved) perception necessarily precedes (improved) production, and instead holds that perception and production evolve in parallel and in interaction with one another (Flege & Bohn 2021), which – again – renders the earlier claim of perceptual primacy untestable.

One way that has been suggested as a test of the 'perception-leads' hypothesis is to compare the effectiveness of competing methods of L2 training. If it is true that a correct mental representation of the target sounds is required for correct production of the L2 sound, just learning how to differentiate between two (or more) nonnative sounds from each other and/or from the vowels in the learner's native system, should speed up the process of learning to adequately pronounce the target sounds actively. It should, in fact, be a more effective way than asking the learner to pronounce the target sounds authentically without prior perceptual training. This idea has been put to the test many times, in different ways, but – unfortunately – the results of such experiments are not easy to interpret. Nagle and Baese-Berk (2021: 18) review the available evidence as follows:

> According to Sakai and Moorman's (2018) meta-analysis of perception training studies, perception training leads to small, yet significant, gains in posttest production accuracy. At the same time, they found that gains in perception were not significantly correlated with gains in production, and individual perception training studies have yielded a wide range of results. Researchers have also investigated the effect of production training and integrated perception-production training paradigms on perceptual learning. Here too, a range of effects have been observed: production training leading to medium gains in

[6] The direct-realist view of speech perception can be seen as a modern version of the 1950s idea of the motor theory of speech perception (e.g., Liberman, Cooper, Shankweiler & Studdert-Kennedy (1967). The existence of 'mirror neurons' (e.g., Rizzolatti & Craighero 2004), has been advanced as an explanation of how seeing or hearing actions performed by humans activates the neural circuitry needed to produce the same phenomena oneself.

perception (Sakai, 2016); no influence of production training on perception (Thorin et al., 2018); and disruption of perceptual learning (Baese-Berk, 2019; Baese-Berk & Samuel, 2016). However, as in other areas of perception-production research, training studies vary widely in methodological choices, choices that have a direct impact on perception-production findings (cf. Sakai & Moorman, 2018).

It seems reasonable, therefore, that we use the data of the present study to evaluate the perception production link. Our study does not involve training in one domain and testing in the other, so that no claims can be made with respect to causality. However, we will be able at least to determine whether the accuracy of the perceptual representation of the target sounds is positively correlated with the accuracy of the individual L2 learner's active production of the same sounds.

2.6 Language dominance

It is well known in the literature on language acquisition that 'balanced' bilinguals, i.e., bilinguals who are equally proficient in their two languages, are rare (e.g., Grosjean, 1982, 2010). This holds not only for the end state which bilinguals reach, but also for the developmental stages which they pass through on their way. Typically, bilingual children, even if exposed to both languages from birth, are more proficient, or dominant, in one of their two languages (e.g., Paradis, 2007). Children's relative proficiency in their two languages will, in some sense, be a function of the amount of language to which they are exposed in those languages. In the context of bilingualism, dominance refers to observed asymmetries of skill in, or use of, one language over the other. Thus, a Spanish-English bilingual who is Spanish-dominant may process Spanish speech more easily than English speech, access lexical items faster in Spanish than in English, and use Spanish more often on a daily basis than English (Birdsong, 2014).

The concept of language dominance captures disparities in rate and complexity of a bilingual's development of two languages in that the language developing faster and with greater complexity is usually denoted as one's dominant language whereas its counterpart is referred to as his/her weaker language (Yip, 2013). Correlated with degree of language use and found to be influential in language choice, language dominance is expected to impact on both frequency and complexity of bilinguals' code-mixing (Genesee, Paradis & Crago, 2004; Montrul, 2013). Unsworth (2015) explores the extent to which children's language experience and their absolute and relative language proficiency are related, with a view to determining

whether measures of language experience can be used as indicators of language dominance in studies of bilingual language acquisition.

Unsworth states that language dominance is understood as bilingual children's relative proficiency in their two languages. In addition, she considers that language dominance can also be conceived of as a much broader concept, involving "a linguistic proficiency component, an external component (input), and a functional component (context and use)." Unsworth then reviews factors affecting bilingual children's language environments, including parental language strategy, the status of the language(s) (minority/majority, high/low prestige), type of education (monolingual/bilingual/immersion, etc.), siblings and birth order, literacy and literacy-related activities, amongst others. These factors can affect both the amount and type of language exposure available, leading to considerable variability in bilingual children's language experiences. Like monolingual children, and probably even more so, bilingual children also vary in the rate at which they acquire their two languages. According to Unsworth, in any accurate assessment of bilingual children's linguistic abilities, it is important that such differential capabilities are taken into account, whether this is for the purposes of assessing bilingual children in comparison with their monolingual peers, comparing and contrasting the linguistic development of different groups of bilingual children, or examining possible bilingual language outcomes. The current study will also compare bilingual adolescents' performance with their monolingual peers to consider whether bilingualism does have any effect on learning and improving a foreign language.

Another important aspect of bilingual children's language experience which has been related to their rate of acquisition, is language use or output, i.e., the extent to which children actively speak the language in question. Unsworth (2015) briefly demonstrates that bilingual children's (rate of) acquisition has been linked to both quantitative and qualitative properties of their language experience, including amount of exposure, children's own language output, and whether input is from native or non-native speakers; whilst these factors have been related to absolute measures of proficiency such as MLU and vocabulary size, their relation to children's relative proficiency, i.e., to their patterns of language dominance, remains largely unexplored.

In the present study I compare the possible advantage of bilingual learners of English as a foreign language (EFL) relative to a matched group monolingual EFL learners. As explained in Chapter 1, the bilinguals speak Azerbaijani as the home language and have been exposed to Persian as the language of instruction and education from age 4 onwards. Given the age difference at which the two languages were acquired, we do not expect perfect or balanced bilinguals. Rather we expect to find the students on a cline between Azerbaijani dominant to

Azerbaijani and Persian balanced. It would be unusual, given the language situation in the northwest of Iran, to find bilingual adolescents with dominance of Persian over Azerbaijani.

The language dominance in our Azerbaijani/Persian EFL learners has been quantified by conducting the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al., 2007). This questionnaire asks the respondents to self-estimate the length and of exposure to the languages they command, how often they use the languages in a range of communicative domains, and how they self-rate their proficiency in each of the languages in terms of listening, speaking reading and writing skills. (See Chapter 3 for details). At the same time, we have detailed experimental data on how our participants carried out a perceptual assimilation task, in which they identified each of the 11 monophthongs of American English as one (and only one) of the six vowels of Persian or one of the nine vowels of Azerbaijani. As a complement to the self-rating, we will examine the consistency with which the students performed the assimilation task as a validation measure of the language dominance as indicated by the questionnaire. If the task consistency is positively and strongly correlated with one or more of the questionnaire-based dominance measures, the self-rating of the respondents would be (much) more credible, and we would have yet another objective proxy (in Unsworth's terminology) of language dominance.

This study has accomplished to investigate how English learners with Azerbaijani and/or Persian as their native language pronounce and perceive the vowels of English, and to compare this to the sound structure of the native languages.

2.7. Research questions and hypotheses

In this final section, I recapitulate and refine the research questions I want to answer in the present thesis, and formulate testable hypotheses for the experiment that will be reported in the following chapters.

1. How do (early) monolingual Persian and (early) bilingual Azerbaijani/Persian listeners categorize the pure vowels of American English as exemplars of the vowels of their native language(s)? Hypothesis: The tense-lax counterparts of English vowels will be assimilated into the same native language categories (tested in Chapter 4).
2. Does the perceptual assimilation found in (1) differ between the two groups of listeners in a way that can be explained by a difference between the vowel systems of Azerbaijani and Persian? Hypothesis: the central(ized) AE vowels /ʌ, ʊ/ may be categorized separately by the bilinguals when instructed to assimilate the AE vowels into the vowels of Azerbaijani (tested in Chapter 4).

3. What differences are there in relative language dominance between Azerbaijani and Persian in the early bilinguals, and are these differences in any way reflected by their task performance in (1)? Hypothesis: the more dominant Azerbaijani over Persian, the larger the difference in consistency with which the early bilinguals perform the perceptual assimilation task in Azerbaijani mode versus Persian mode (tested in Chapter 4).
4. How do English learners with Azerbaijani and/or Persian as their native language perceive the vowels of American English? Hypotheses: the non-native listeners will show a rather poorly defined perceptual categories for the AE vowels in which the contrast between adjacent vowel categories is primarily based on a difference in duration (tested in Chapter 3).
5. How does the perception of the vowels of English found in (4) differ from the way native speakers of American English perceive their vowels? Native AE listeners will have more sharply defined perceptual representations of the vowels, in which contrasts between adjacent vowels in the vowel space are primarily based on spectral differences rather than on duration (tested in Chapter 5).
6. To what extent can the differences found in (4) be predicted/explained by properties of the native language of the native language(s) of the nonnative listeners? Hypothesis: the lax AE vowels /ʊ/ and /ʌ/ will be better distinguished from their tense counterparts in the perceptual identification by the early bilinguals than by the monolinguals due to the existence of central vowels in Azerbaijani (tested in Chapter 5).
7. Same question and hypothesis as in (4) but now for the production of the AE vowels (tested in Chapter 6)
8. Same question and hypothesis as in (5) but now for the production of the AE vowels (tested in Chapter 6)
9. Same question and hypothesis as in (6) but now for the production of the AE vowels (tested in Chapter 6)
10. To what extent can the problems in perceptual identification of the AE vowels by the two groups of learners be predicted from the results of the perceptual assimilation task in (1)? Hypothesis: poorly defined categories in (4) will be vowel AE pairs that are implicated in either the Same Category or Category Goodness scenarios established in (1); tested in Chapter 7.
11. Same question and hypothesis as in (10) but for the production of the AE vowels (tested in Chapter 7).

12. To what extent are perception and production of the AE vowels by the EFL learners correlated? Hypothesis: (non)nativeness in the perceptual representation will correlate with (non)nativeness in the production, either in terms of a rank order among the 11 vowels (across all listeners), or in the rank order of listeners (across all vowels); tested in Chapter 7.

Chapter 3

Language dominance in Azerbaijani/ Persian EFL learners. Analysis of LEAP-Q data

3.1. Introduction

Iran is a multi-ethnic and multi-lingual society with about 85 million inhabitants. A recent census shows that 61% of the population is Persian and speaks Modern Persian as its first language. Persian, an Indo-European language, is the national language which has been positioned as the official language of government and education throughout the state of Iran. The second-most frequently spoken mother tongue in Iran is Azerbaijani (also called Azeri or Azari). Azerbaijani belongs to the Turkic language family. Although it differs markedly from Standard Turkish, it is often considered a Turkish dialect.[7] Azerbaijani is the native language of 16% of the Iranian population, i.e., about 13.5 million people.[8] It is spoken in the north of Iran, in the area bordering on Azerbaijan. In the first four years of their life, children speak Azerbaijani as the home language. At the age of 4, children go to primary school, where Persian is the language of instruction. In some cases, the children spend a pre-school year preparing for the all-Persian immersion one year later. As a result of this diglossic language situation, Azerbaijani in Iran are early bilinguals who have learned to speak two languages from childhood onwards, with roughly equal command of the two languages, and whose linguistic performance as adolescents or adults should be on a par with that of monolingual speakers of either Persian or Azerbaijani.

In our project we are interested in the potential advantage of knowing two different (and unrelated) languages from childhood onwards for learning a foreign language, specifically English, at a later stage in life. We investigate the advantage of early bilingualism

[7] The cross-lingual intelligibility of Standard Turkish for Azerbaijani listeners has been estimated at 56% (Salehi & Neysani, 2017). Cross-lingual intelligibility of Azerbaijani for Turkish listeners (using a different method) was estimated at 42% (Sağın-Şimşek & König, 2012).

[8] These are conservative estimates. According to Lazerte (2021) the official demographic statistics in Iran do not generally include self-declared ethnic identity. For this reason, estimates of Iran's Azerbaijani Turk population range from 18 million to 40 million, depending on the sources consulted. A more realistic estimate would be that Azerbaijani constitute well around 23 percent of the entire population of Iran, concentrated in the six northwestern provinces (Shaffer, 2021).

in the field of phonetics, i.e., the acquisition of the sound system of the foreign language. As a first step in our project, we targeted two groups of adolescent EFL learners in Iran. At the time of testing, they were roughly 16 years of age, and had taken English lessons in secondary school for some 4 years. We have made recordings of these learners when they produced the vowels, consonants and a range of consonant clusters (in onset and in coda position in syllables) in English target words in fixed carrier phrases, and read aloud a short piece of connected speech. We also tested the learners' perceptual representation of the vowels of English, and – as a preliminary step – asked them to map the vowels of (American) English to the vowel inventories of their native languages (using a perceptual assimilation task, see § 3). The materials used in the perceptual assimilation task were presented twice to each participant so that we could estimate the participants' consistency in performing the assimilation task. It is our working hypothesis that this consistency will increase as participants have a better and more sharply defined perceptual representation of the vowel categories in their native language(s).

This hypothesis breaks down into several sub-hypotheses. First, when the EFL learner is an early monolingual in just a single language, i.e., Persian, the perceptual representation of the six native vowels of Persian will be excellent, and the assimilation task will be done with a high level of consistency. However, when the learner is an early bilingual speaker of Azerbaijani and of Persian, the mental representation of each of the two vowel systems may be less sharply defined than in the case of a monolingual speaker. Moreover, within the group of early bilinguals there will be differences in language dominance (e.g., Luk & Bialystok, 2013).

Depending on the amount of language input and age of acquisition in each of the two languages, one language may be more fully acquired and therefore better represented in the learner's mind than the other. It is reasonable, therefore, to expect our early bilinguals to differ in the relative strength of the two languages they command. For some, Azerbaijani will be strongly dominant, while the two languages will be more equally balanced for others. Sebastián-Gallés & Soto-Faraco (1999) studied the perceptual sensitivity of early bilingual speakers of Catalan and Spanish. They found that, even as adults, the bilinguals who were Catalan-dominant were more sensitive to the Catalan contrasts than their Spanish-dominant counterparts, while the reverse was true for Spanish contrasts.

In the present study, we will test the hypothesis that language dominance in our Azerbaijani/Persian early bilinguals will be reflected by the difference in consistency with which they perform the perceptual assimilation task in each of their two languages. Moreover,

we test the second hypothesis that the bilinguals will be less consistent overall in the perceptual assimilation task than their monolingual Persian peers.

In § 3.2, we will describe how we established the language dominance in Azerbaijani and Persian in our early bilinguals, and compare the results obtained for our bilinguals and monolinguals. Then we will explain the Perceptual Assimilation task our EFL learners performed (§ 4) and examine to what extent differences in language dominance are reflected in the consistency found in the assimilation task (§ 3.3).

3.2. Using the LEAP-Q to establish language dominance

The *Language Experience And Proficiency Questionnaire* (LEAP-Q) was developed by Marian, et al. (2007). The LEAP-Q is a validated questionnaire tool for collecting self-reported proficiency and experience data from bilingual and multilingual speakers aged 14 to 80. It is available in over 20 languages, and can be administered in a digital, paper-and-pencil, and oral interview format (Kaushanskaya, 2020: 954). We used the paper-and-pencil version of the LEAP-Q that is available for use in Iran.[9] Our participants filled in the questionnaire on sheets of paper. We obtained responses from 23 Azerbaijani-Persian bilinguals (12 female) and 21 monolingual Persian participants (11 female). All participants filled in the same questionnaire using Persian-Arabic script. The responses were the copied into digital LEAP-Q forms (in English) by the author, and collected automatically in an (SPSS-readable) Excel sheet.

In the LEAP-Q, respondents are asked nine general questions (Q1.1-9) which have to be answered by providing free-format information such as the name of one or more languages, percentages of the time devoted to specific language activities, etc. These questions (in English) are listed in Appendix 1 of Marian et al. (2007: 966). In the second part of the LEAP-Q, questions Q2.1-7 are in fixed format, requiring the respondent to tick a number on a scale from 0 to 10, and answer these questions for each of the maximally five languages listed in response to Q1.1. We will first discuss the free-format answers given to Q1.1 through 1.9, and then examine the numerical responses to the questions in part 2.

In **Q1.1**, students were asked to list maximally five languages in order of dominance (as they perceived it). All monolinguals listed Persian as their most dominant language and English as their second-most dominant language. All bilinguals listed Azerbaijani as their most dominant language with Persian in second place, with just one exception who mentioned

[9] https://bilingualism.soc.northwestern.edu/wp-content/uploads/2013/06/LEAPQ-Farsi.doc

English as his second-most dominant language (and Persian fourth). All other bilinguals mentioned English as their third-most dominant language.

Q1.2 asked students to specify the order in which they had acquired the languages they mentioned in Q1.1. All monolinguals acquired Persian as their first language. All bilinguals indicated that they acquired Azerbaijani before Persian. More in general, the order of acquisition of the spoken language perfectly reflects the self-estimated dominance.

In response to **Q1.3**, the students listed what percentage of the time they were *currently and on average* exposed to each language (percentages should add up to 100). It is not clear over what period of time the "current average" has to be taken. Be this as it may, the answer could be useful to assess the dominance-ratio between Azerbaijani and Persian for the bilinguals. For the monolinguals, the current exposure should be high, say 70% at least – it would be strange if more than 30% of a student's time would be taken up by exposure to foreign languages at school. The results (Table 3.1, Q1.3) show that the monolinguals think they have exposure to Persian about 72% of the time, and some 19% to English (the remaining 9% is not specified in the table; it is divided over a range of other languages mentioned, e.g., Arabic, Chinese, German, Hungarian, Turkish, Urdu). Only one monolingual mentioned exposure to Azerbaijani (10% of the total exposure time). The bilinguals, even as adolescents, stated that they had exposure to Azerbaijani in 60% of the time against only 20% in Persian (and 14% in English).[10]

[10] We may define a P/A dominance ratio for current exposure by dividing the percentage of the time spent on exposure to Persian by the total percentage of the time the participant claims to be exposed to either Persian or Azerbaijani, %P / (%P + %A). The mean ratio then turns out to be .99 for the monolinguals against .27 for the bilinguals, $t(42) = 16.0$ ($p << .001$).

Table 3.1. Selected LEAP-Q results for bilingual Azerbaijani-Persian and monolingual Persian learners of EFL. For each question the mean (Mn), standard deviation (SD) and range (Rg) of the responses is given.[a]

	Bilingual Azerbaijani/Persian (*N* = 23)									**Monolingual Persian (*N* = 21)**					
	Azerbaijani			Persian			English			Persian			English		
	Mn	SD	Rg	Mn	SD	Rg	Mn	SD	Rg	Mn	SD	Rg	Mn	SD	Rg
Q1.3. Estimated current average exposure to language (adds up to 100%)															
Exposure	60.00	21.9	5-90	19.52	12.2	5-50	13.61	8.5	4-30	71.57	17.8	40-98	18.76	12.0	2-48
Q1.4-5. Relative preference per language for … (adds up to 100%)															
Reading	24.78	24.2	0-90	38.57	26.9	7-80	27.65	19.5	2-80	55.05	25.1	0-98	35.57	25.1	2-90
Speaking	55.22	29.2	5-90	20.78	16.0	3-50	20.61	19.8	2-85	67.38	26.5	10-100	26.90	22.1	0-80
Q2.1. Age milestones (years)															
Started learning	1.30	.47	1- 2	4.04	2.08	1- 7	11.30	1.78	7-15	1.05	.22	1- 2	9.33	2.27	4-12
Fluent talker	4.57	2.59	2-12	7.83	2.89	4-15	14.50	1.50	12-17	4.48	1.29	3- 8	14.14	2.43	10-18
Started reading	7.67	2.09	5-13	6.65	.88	5- 8	11.40	3.03	1-15	6.62	.59	5- 7	10.24	2.12	6-13
Fluent reader	10.38	2.36	7-16	9.17	1.30	7-12	14.05	3.53	1-17	8.65	.59	7- 9	14.14	2.29	10-18
Q2.2. Immersion duration (years)															
Country	16.78	.74	15-18	16.78	.74	15-18	.00	.00	0- 0	16.43	1.50	14-18	.05	.22	0- 1
Family	16.65	1.07	13-18	7.17	8.24	0-18	.10	.45	0- 2	16.43	1.50	14-18	.00	.00	0- 0
School	10.70	.88	8-11	10.74	.92	8-12	6.50	1.40	3- 9	10.76	2.07	8-17	6.57	2.13	3-10
Q2.3. Self-reported proficiency in…[b]															
Speaking	9.61	.94	7-10	8.39	2.08	2-10	5.80	2.46	2-10	9.52	.75	8-10	6.90	2.28	3-10
Understanding	9.26	1.25	5-10	9.48	.85	7-10	6.00	2.03	1-10	9.67	.66	8-10	6.62	2.40	3-10
Reading	6.43	2.97	0-10	9.57	1.08	5-10	7.10	1.92	4-10	9.48	.93	7-10	7.67	1.74	5-10
Q2.4. Contribution to language learning from…[c]															
Family	8.74	2.56	1-10	4.48	3.82	0-10	2.35	2.80	0-10	6.76	3.10	1-10	4.29	2.69	0-10
Friends	.57	1.44	0- 5	2.22	3.29	0-10	4.30	3.60	0-10	3.19	3.96	0-10	2.52	2.73	0-10
Reading	10.00	.00	10-10	3.83	3.94	0-10	2.35	3.63	0-10	10.00	.00	10-10	.38	1.12	0- 5
TV	2.52	2.25	0- 5	7.35	3.92	0-10	2.00	2.81	0-10	4.81	3.50	0-10	5.67	4.21	0-10
Radio	3.30	3.05	0-10	8.30	2.74	1-10	6.35	3.00	1-10	6.76	3.10	1-10	7.71	3.18	1-10
Self-instruction	.96	1.66	0- 5	2.61	3.53	0-10	.75	2.45	0-10	.52	1.50	0- 5	1.57	3.17	0-10
Q2.5. Extent of current language Exposure…[d]															
To family	9.17	2.29	1-10	4.43	3.87	0-10	.95	1.47	0- 5	7.86	2.54	5-10	4.10	3.19	0-10
To friends	2.22	2.11	0- 5	6.26	3.71	0-10	4.05	3.27	0-10	7.24	3.55	0-10	5.67	3.90	0-10
To reading	9.78	1.04	5-10	3.35	3.50	0-10	.35	1.14	0- 5	10.00	.00	10-10	.24	.44	0- 1
To TV	3.52	3.34	0-10	8.91	2.11	5-10	3.80	2.69	0-10	7.14	2.54	5-10	7.10	2.49	5-10
To radio	1.83	1.97	0- 5	7.87	3.55	0-10	1.85	2.16	0- 5	6.90	2.49	5-10	5.19	3.63	0-10
Self-instruction	.09	.29	0- 1	1.04	2.42	0-10	5.00	4.59	0-10	2.62	3.94	0-10	5.29	3.64	0-10
Q2.6-7. Self-reported foreign accent as perceived by…[e]															
Self	2.78	2.11	0- 6	4.17	2.13	0- 9	5.70	1.95	0- 9	2.52	3.54	0-10	5.62	2.44	1-10
Others	4.43	2.86	1-10	4.39	1.62	0- 5	5.55	3.52	0-10	1.43	2.09	0- 5	6.05	2.82	1-10

Notes:
a. Question numbers refer to their listing in Marian et al. (2007: 966−967).
b. 0 'none' to 10 'perfect'.
c. 0 'not a contributor' to 10 'most important contributor'.
d. 0 'never' to 10 'always'.
e. 0 'none' to 10 'pervasive'.

Q1.4 asks the respondents in what percentage of the cases they would prefer to *read*, in each of the languages they command, a translation of a text that was originally written in a language they would be unable to understand. **Q1.5** asks the respondents what percentage of the time they would prefer to *speak* in each of the languages they listed (assuming the

interactant is equally fluent in each language chosen). The percentages should add up to 100. The results indicate that the bilinguals preferred to read the translations in Persian rather than in Azerbaijani (39% vs 25%, and 28% for English), which probably means that the Western-style Azerbaijani orthography was a problem for the participants. The monolinguals preferred translations into Persian (55%, against 36% into English). This contrasts rather sharply with the responses to Q1.5, which revealed a clear preference on the part of the bilinguals to speak in Azerbaijani rather than in Persian (55% vs. 21%), while the monolinguals preferred to speak Persian (67% of the time) rather than English (27%). Generally, then, students preferred to use the language they listed as first acquired and most dominant.[11]

In **Q1.6**, the students were asked to name the cultures with which they identified and to express the strength of their identification on a scale from 0 to 10. With just one exception, all respondents identified with Iranian culture first (one calls it 'Persian culture'). The exception mentioned Azerbaijanian culture first, and Iranian second. All bilinguals mention Turkish (or sometimes Azerbaijanian) as their second-most favorite culture. Only three monolinguals list Turkish/ Azerbaijanian as their second choice; they tend to list American culture second. The cultural identification is at odds with the linguistic preferences. We will not use the responses to Q1.6 in our attempts to quantify Persian/Azerbaijani language dominance.

Q1.7 inquired about the length of the respondent's education and highest level attained. Since all our respondents went through the same curriculum, the responses are predictable from the student's age, which is the topic of one of the later questions. **Q1.8-9** are either not applicable to our respondents (when did you emigrate to the USA?) or were uniformly filled in with negative answers (vision/hearing impairment, language or learning disability).

In the second part of the LEAP-Q, **Q2.1** asks the respondents to specify, for each of the languages they listed in Q1.1, their age (in years) when (a) they were first exposed to them, (b) when they considered themselves fluent speakers, (c) when they started reading, and (d) when they considered themselves fluent readers. The responses indicate that the monolinguals started learning their mother tongue (Persian) at the age of 1.05 years, while the bilinguals said they started learning Azerbaijani at age 1.30; their acquisition of Persian started at age 4.04. Participants considered themselves fluent in their first language about 3 years later (age 4.48 for monolinguals, 4.57 for bilinguals). The bilinguals stated they were fluent in the second language (Persian) at the age of 7.83. In the LEAP-Q responses, some of

[11] Two bilingual students who expressed a strong desire to spend most of their speaking time with an English interactant (85 and 60%), may have been strongly motivated learners of English who believed they might improve their English skills by interacting with a fluent (native) speaker of English.

the bilinguals mention age 3 as the starting point of learning Persian, probably because they attended pre-school or kindergarten at the age of 3. Reading is a different matter altogether. All respondents started reading in Persian at the same age, i.e., 6.62 and 6.54 for monolinguals and bilinguals, respectively. Reading in Azerbaijani, for the bilinguals only, started a year later, at age 7.67. Monolinguals considered themselves fluent readers in Persian at age 8.65, followed a little later by the bilinguals at age 9.17. Possibly, having to learn a second writing system (Western instead of Arabic script) caused some delay here.

In **Q2.2** the respondents had to specify, for each of their languages, how many years and months they had spent in (a) a country (b) a family, and (c) a school where the language was spoken as the primary vehicle of communication. In the large majority of the responses, our respondents specified years only. In our data analysis, the occasional specification of months was converted to an extra year if the number of months was larger than six. Since all respondents hailed from Iran, the answer to question (a) was roughly the same for monolinguals and bilinguals (16.43 vs 16.78 years, with no difference between Persian and Azerbaijani). These numbers also correspond to the respondents' age. The bilinguals indicate that the time they spent in the homes of Persian-speaking families was considerably shorter (7.17 years) than for the monolinguals. The length of (self-reported) immersion in the school context in Persian (and Azerbaijani for the bilinguals) was the same for all respondents, i.e., roughly 10.7 years.

Q2.3 in the LEAP-Q is probably the most relevant question for our purpose. Students specified how proficient they considered themselves in each of the languages they command, in terms of (a) speaking, (b) listening and (c) reading. The latter two questions relate to receptive language skills (the questionnaire does not ask the respondents to say anything about writing proficiency). On a scale from 0 to 10, the students considered themselves equally proficient in speaking their first language, i.e., 9.52 and 9.61 for the monolinguals and bilinguals, respectively. The bilinguals consider themselves slightly less proficient in Persian, 8.39 than in Azerbaijani. This difference is significant, $t(22) = 2.4$ ($p = .024$). We take this as an indication that the bilinguals have a realistic view of their language proficiency, and that Azerbaijani is indeed the stronger of the two languages they acquired at a young age. No significant differences can be observed in the self-rated proficiency between monolinguals and bilinguals when it comes to listening skills. As for reading Persian, there is no significant difference between monolinguals (9.48) and bilinguals (9.57). However, the bilinguals rate their reading proficiency in Azerbaijani (6.43) significantly lower than in Persian, $t(22) = 4.9$ ($p < .001$). In our research we are primarily concerned with the proficiency and dominance in

the spoken language modality, so that we will ignore the reading proficiency score in our attempts to find a measure of relative language dominance of Azerbaijani over Persian.

Questions **Q2.4-5** are concerned with the settings in which the participants acquired and currently use their various languages. The self-estimations are included in Table 4.1 under Q2.4 and 2.5, respectively. We will not comment on the results here, as these background data are not of immediate use in our attempts to define a quantitative measure of language dominance.

Finally, we need to consider **Q2.6-7**. Here the respondents had to specify, for each of their languages, the strength of a non-native accent (a) as perceived by themselves and (b) as indicated to them by others. The self-perception of non-native accent in their respective first language is low for monolinguals (2.52) and bilinguals (2.78) alike, $t(42) = .3$ ($p = .768$, ins.). The bilinguals, however, rate their non-native accent in Persian as significantly stronger (4.22), $t(22) = 2.2$ ($p = .040$, 2-tailed). Moreover, the monolinguals report virtually no comments by others (1.43 on a scale from 0 to 10) on their non-nativeness in Persian – which means that they speak Persian without any accent. The bilinguals, however, report more comments on their non-native accent, both when they speak Azerbaijani (4.43) and when they speak Persian (4.39). The difference between monolinguals and bilinguals in Persian is significant, $t(42) = 3.0$ ($p < .001$); the difference between speaking Azerbaijani and Persian by the bilinguals is not, $t(22) = .1$ ($p = .955$, ins.). This might indicate that the bilinguals' pronunciation of the two languages they learned at a young age, is somewhat compromised in both languages, i.e., that bilingualism comes at a price. Probably, some (weighted) mean of these non-nativeness ratings could be a powerful index of relative language dominance of Azerbaijani over Persian.

It would seem reasonable to hypothesize that bilingual students who are (or consider themselves to be) highly native and proficient in Azerbaijani, will be less native and proficient in Persian, and *vice versa*. The more dominant one language, the less dominant the other language will be. This would predict a negative correlation between the two languages. In Table 3.2 we show a non-redundant correlation matrix (i.e., lower triangle) for the relevant variables.

Some correlations stand out immediately. Speaking and listening proficiency in Azerbaijani are strongly correlated ($r = .863$); the same correlation is still significant but weaker ($r = .585$) in Persian (so apparently the skills diverge more in the less dominant language). Also, self-assessed non-native accent correlates strongly with reported identification as a non-native, both for Azerbaijani and for Persian. These four positive correlations are highlighted in green cells in the matrix. Crucially, there is a (moderate but significant) negative correlation between the student's identifiability as a non-native speaker of Persian and his/her identifiability as a non-native of Azerbaijani – as predicted. Note that

the proficiency in Azerbaijani (whether speaking or listening) correlates negatively with the proficiency measures in Persian, and with the self-rated strength and identifiability of a non-native accent in Azerbaijani – but positively with the same accent-ratings for Persian.

Table 3.2. Correlation matrix of eight self-rated performance measures (scales from 0 to 10) for 23 bilingual Iranian participants with Azerbaijani (AZ) as L1 and Persian (PE) as L2. Pearson's *r* in upper part of cells, *p*-value in bottom part.

	SpeakAZ	ListenAZ	SpeakPE	ListenPE	AccentAZ	IdentifAZ	AccentPE
Proficiency Listening in AZ	**.863**** < .001						
Proficiency Speaking in PE	−.150 .247	−.058 .396					
Proficiency Listening in PE	−.097 .330	−.037 .433	**.585**** .002				
Non-native accent in AZ	**−.388*** .034	**−.408*** .027	**.455*** .015	.265 .111			
Identified as non-native in AZ	**−.509**** .007	**−.440*** .018	.268 .108	.230 .146	**.620**** .001		
Non-native accent in PE	**.392*** .032	**.389*** .033	−.335 .059	−.276 .101	−.233 .143	−.210 .169	
Identified as non-native in PE	**.454*** .015	**.502**** .007	**−.430*** .020	**−.356*** .048	**−.487**** .009	**−.513**** .006	**.781**** < .001

** Correlation significant at the 0.01 level (1-tailed).
* Correlation significant at the 0.05 level (1-tailed).

3.3. Consistency in perceptual assimilation

In the perceptual assimilation task, which I will present in chapter 4, listeners were asked to decide with which of the sounds in their native language they identified each of a number of unfamiliar target sounds, and then rated the goodness of the target sound as a token of the native category chosen. In our experiment we asked our monolingual and bilingual participants to identify tokens of the monophthongal vowels of American English (AE) as instances of the six vowel categories of Persian, /i, e, æ, ɑ, o, u/, and (in the case of the bilinguals) also as instances of the nine vowels of Azerbaijani, which has roughly the same six vowels as are used in Persian plus three central vowels /y, ɯ, œ/ (for more details on the vowel systems of AE, Persian and Azerbaijani, see Van Heuven et al., 2020 and references therein). Following established practice in, e.g., Peterson and Barney (1952) and Hillenbrand et al. (1995), stimuli were the words *heed* /hid/, *hid* /hɪd/, *hayed* /hed/, *head* /hɛd/, *had* /hæd/, *hud* /hʌd/, *hod* /hɑd/, *hawed* /hɔd/, *hoed* /hod/, *hood* /hʊd/, and *who'd* /hud/. These words had been spoken by two male native speakers of AE in a fixed carrier phrase *Now say ... again*. The 2 × 11 target words had been digitally excised from their spoken context and were presented to each listener over headphones twice in different random orders per listener, who heard 44 target words in all.

Listeners were instructed to categorize the vowel in the target word as a token in either Persian (with forced choice from a set of six alternatives) or in Azerbaijani (with forced choice from nine alternatives). The alternatives were shown as six or nine response buttons on a computer screen, one of which the listener had to click on using a mouse pointer.

Immediately after clicking a response button the listener had to click one of five activated buttons at the bottom of the screen to indicate the goodness of the token as an exemplar of the category just chosen. Response latency was measured in milliseconds from the moment the onset of the target word was made audible until the goodness button was pressed. The listeners in the experiment were the same 44 adolescents whom we asked to fill in the LEAP-Q in § 3.2. Note that the monolingual Persians performed the perceptual assimilation task only once, assimilating the AE vowels to the Persian response set. The bilinguals performed the task twice, once in Persian, the second time in Azerbaijani. For details of the procedure, we refer to Afshar and Van Heuven (2021) as well as chapter 4.

In the present paper we will not be concerned with the distribution of the choices the two groups of listeners made for the target vowels. We will only analyze the consistency with which the respondents performed their perceptual assimilation task. Earlier research has shown that listeners are more consistent in their perceptual labeling choices as they are more familiar with the phonological system of the language they respond in (Van Zanten & Van Heuven, 1984; Van Heuven & Van Houten, 1989; Van Heuven, 2017). When hearing a target sound that is a good or excellent representative of a sound category in the listener's native language, the choice is easy: the listener will make the same decision quickly on both occasions, with high typicality ratings. However, when hearing a sound that is a poor token of one native category (or even several adjacent native categories), the decision is difficult: the typicality rating will be low, different choices may be made on first and second presentation, and the response will only be given after some delay.

We define our listeners' response consistency as the percentage of repeated target pairs with identical choices divided by the total set of pairs presented (N = 22). Table 3.3 presents the overall results we obtained.

Table 3.3. Overall Response consistency, Goodness rating (on a scale from 1 to 5 = best) and Response latency (ms) on first and second presentation for monolingual Persian and bilingual Azerbaijani/Persian listeners, when assimilating American English vowels to the vowels of Persian (PE) or Azerbaijani (AZ).

	Monolinguals		Bilinguals				Comparison			
	(a) PE		(b) PE		(c) AZ		(a) vs. (b)		(b) vs. (c)	
	Mean	SD	Mean	SD	Mean	SD	t(42)	p	t(22)	p
Consistency (%)	78.35	13.63	77.67	11.38	58.30	16.65	.18	.856	6.13	< .001**
Goodness1 (1..5)	4.06	.59	3.77	.50	3.61	.69	1.76	.086	1.37	.184
Goodness2 (1..5)	4.05	.62	3.83	.56	3.60	.71	1.23	.227	1.82	.083
Latency1 (s)	3.04	.75	3.18	.77	3.67	.77	.61	.546	5.61	< .001**
Latency2 (s)	2.68	.69	2.94	.62	3.36	.74	1.30	.201	3.27	.004*

Notes: * very significant ($p \leq .01$); ** highly significant ($p \leq .001$)

The bilinguals, when responding in the Persian mode, are only marginally less consistent than the monolinguals; the difference of (less than) 1 percent is insignificant by a t-test for independent groups. Goodness ratings are somewhat more favorable when given by the monolinguals than by the bilinguals, both on first and on second presentation of the stimuli, but again the differences between the two groups are not significant.

Before analyzing the response latencies, we first applied data trimming in order to exclude responses that were excessively long. The trimming was done such that the 10 percent longest responses were removed from the data. The cutoff for the 90th percentile was found at 7.02 seconds. The results then show, first of all, that the reaction times were shorter in the second half of the experiment than in the first, with a mean difference for the monolinguals of 359 ms and for the bilinguals of 242 ms. A repeated measures ANOVA with Repetition as a within-subjects factor and Language background (monolingual, bilingual) as a between-subjects factor bears out that the main effect of Repetition is highly significant, $F(1, 42) = 21.20$ ($p < .001$, $p\eta^2 = .335$) but that the interaction between Repetition and Language background is not, $F(1, 42) = .81$ ($p = .374$, $p\eta^2 = .019$). The most likely reason why participants are faster in the second half of the experiment is that they got more familiar with the location of the response buttons on the computer screen. The more relevant factor of Language background turns out to have no effect. To be true, the monolinguals are faster than the bilinguals overall but the mean difference of 140 ms after the first presentation and even of 257 ms after the second presentation is not significant, $F(1, 42) = .944$, $p = .337$, $p\eta^2 = .022$).

The results are rather different when we compare the responses given in the Persian mode with those in the Azerbaijani mode. Since these differences occur within participants, the t-test for correlated samples can be used. When responses are given in terms of the Azerbaijani vowel system, consistency is about 20 points poorer than when the bilinguals respond in the Persian mode. The difference is highly significant. This indicates that the

Azerbaijani response mode (the dominant L1 for all bilinguals) is more difficult than the Persian mode, even though Persian is the less dominant language for these participants. Similarly, the response latency is longer in the Azerbaijani mode than in the Persian mode, both on first and on second presentation of the stimuli. The difference is (highly) significant on both occasions. There are no significant differences in goodness ratings, neither on first nor on second presentation, although there is a slight tendency for the ratings to be lower when the response mode is Azerbaijani. Normally, we would expect higher consistency and faster responses in the dominant language mode. In the present case, however, the vowel system of the dominant language, i.e., Azerbaijani, is more complex than of the less dominant language, Persian. For Azerbaijani, the participants have to make a choice from nine possibilities, against only six for Persian. This would seem the primary reason why the choice in the Azerbaijani response mode is intrinsically more complex, and therefore yields more inconsistencies and longer latencies. The problem may have been aggravated further by the relative unfamiliarity of the bilinguals with the writing system of Azerbaijani, which we used to identify the response alternatives on the computer screen.

3.4. Language dominance and PAM consistency

In order to test the hypothesis that the difference in consistency with which our bilinguals perceptually assimilate the vowels of AE to Azerbaijani vs. Persian, reflects a difference in language dominance at the level of the individual participant, we computed a number of simple difference measures, by subtracting the scores found for Persian (consistency and LEAP-Q measures) from the scores found for Azerbaijani. Since Azerbaijani is the dominant language (given the LEAP-Q measures), we expect positive differences; negative differences would indicate that Persian is the more dominant language. A caveat is in order here where it concerns the difference in consistency. We have seen that consistency is poorer overall for Azerbaijani than for Persian due to the smaller number of response categories and greater familiarity with the writing system for Persian. For this reason, most of the consistency difference scores will be negative. Nevertheless, as the dominance of Azerbaijani over Persian is stronger, smaller negative differences will be obtained and in some extreme cases of Azerbaijani dominance we may even find a positive difference. In all cases we expect to find positive correlations between the LEAP-Q-based dominance measures and the difference in consistency as defined here.

Table 3.4 summarizes the difference scores we examined. Here the difference in consistency is the dependent variable (criterion), which we want to predict from one or more independent variables based on differences in LEAP-Q scores (predictors).

Table 3.4. Difference scores defined for Azerbaijani (AZ) - Persian (PE) early bilinguals (*N* = 23).

	Name	Definition
1.	Δ Consist	Consistency in AZ minus Consistency in PE (= dependent variable)
2.	Δ Exposure	Current exposure (% of time) to AZ minus PE
3.	Δ Speak	Proficiency (0..10) speaking in AZ minus PE
4.	Δ Listen	Proficiency (0..10) understanding spoken AZ minus PE
5.	Δ Accent Self	Strength of self-perceived nonnative accent (0..10) in AZ minus PE
6.	Δ Accent Others	Frequency of comments by others on nonnative accent (0..10) in in AZ minus PE

Table 3.5 is a correlation matrix of these six variables (non-redundant lower triangle only).

Table 3.5. Non-redundant lower half of correlation matrix for difference measures defined in Table 3.4. Significant *r*-values in bold face ($p < .010$).

		Δ Consist	Δ Expo	Δ Speak	Δ Listen	Δ Acc. S.
Δ Exposure	*r*	−.195				
	p	.372				
Δ Speak	*r*	.205	.283			
	p	.347	.191			
Δ Listen	*r*	.042	.398	**.615**		
	p	.850	.060	.002		
Δ Accent Self	*r*	−.115	−.330	**−.626**	**−.603**	
	p	.602	.124	.001	.002	
Δ Accent Others	*r*	−.305	**−.547**	**−.563**	**−.627**	**.768**
	p	.157	.007	.005	.001	< .001

The two self-ratings for non-native accentedness are positively correlated. A stronger non-native accent in Azerbaijani than in Persian (as judged by the participants themselves) corresponds rather well with the frequency of comments received on their way of speaking. The strength of non-native accent (i.e., sounding Persian when speaking Azerbaijani and *vice versa*) is inversely correlated with the difference in self-reported exposure to, and preference for speaking and listening to Azerbaijani versus Persian. It should be feasible to derive an index of language dominance of Azerbaijani over Persian from these LEAP-Q based measures that makes at least a reasonable prediction of the consistency index we computed for the participant's PAM decisions. In order to find such an index, we performed a multiple linear regression analysis with the five LEAP-Q dominance measures as predictors and the difference in PAM consistency as the criterion. The full model yields an *R* of .623 (which accounts for 39% of the variance in Δ Consistency). An optimal model, obtained by backward elimination of predictors, yields $R = .592$, with Current exposure and the two Accentedness ratings as remaining predictors. The index derived from this analysis, i.e.,

$$.417 \times Z(\Delta \text{ Accent Self}) - .938 \times Z(\Delta \text{ Accent Others}) - .571 \times Z(\Delta \text{ Exposure})$$

explains 35% of the variance in the z-normalized difference in consistency between Azerbaijani and Persian.

3.5. Discussion and conclusions

In this study we tested the hypothesis that a difference in language dominance in early bilinguals would be reflected by the relative consistency with which such bilingual respondents perform a perceptual assimilation task in the two languages they learned during childhood. The experience and proficiency in Azerbaijani (a Turkic language, acquired from birth onwards) and of Persian (acquired from age 3 onwards) was established by administering the LEAP-Q questionnaire to 23 Azerbaijani/Persian early bilingual adolescents, as well as by a matched group of Persian monolinguals.

The results of the questionnaire reveal that the early bilinguals acquired Azerbaijani before they acquired Persian, and considered themselves more proficient in Azerbaijani than in Persian. Their current exposure to Azerbaijani generally exceeded that of Persian, and the large majority of the early bilinguals considered their Azerbaijani accent in Persian stronger than their Persian accent in Azerbaijani, which impression corresponded quite well with the frequency of comments received on their pronunciation of the two languages.

Both groups of adolescents performed a perceptual assimilation task in which they had to identify tokens of American English (AE) monophthongs as instances of the vowels in their native language(s), i.e., to one of the six vowels of Persian and, for the bilinguals, also as one of the nine vowels of Azerbaijani. No significant differences were found in the consistency and speed with which the monolinguals and early bilinguals performed the perceptual assimilation task for Persian. For the bilinguals, assimilation of the AE vowels to the nine vowels of Azerbaijani was more difficult, in terms of consistency and speed, than to the six vowels of Persian, which effect is most likely caused by the greater uncertainty yielded by nine response alternatives for Azerbaijani versus six in Persian.

Although no single dominance measure derived directly from the LEAP-Q questionnaire correlated significantly with the difference in consistency with which the bilinguals assimilated the AE monophthongs to the nine vowels of Azerbaijani or to the six vowels of Persian, a combination of three LEAP-Q dominance measures accounted for 35% of the variance in the consistency differences observed in our group of 23 early bilinguals. This does support our hypothesis that language dominance is reflected in the consistency with which a perceptual assimilation task can be performed in the two languages acquired by an early bilingual. However, we consider the correlation insufficiently strong and too complicated to advance it as a straightforward, reliable and valid indicator of language dominance.

Chapter 4

Perceptual assimilation Study

4.1. Introduction [12]

Studies on second language (L2) acquisition have long established that certain L2 sounds are more difficult to acquire than others. The ease or difficulty of acquiring certain L2 sounds is often attributed to the influence of first language (L1) phonological knowledge. The assumption is that learning an L2 sound is easier when the L2 sound is similar to an L1 sound and is more difficult when the L2 sound is different from an L1 sound. There are many psychological and linguistic factors that need to be taken into account. Thus, the question of how we actually perceive the similarities and differences between native and non-native sounds still intrigues researchers and has been the impetus for the formulation of theories on L2 phonological learning (Pilus, 2010).

Learning how to carve up reality into categories is one of the most important tasks of the infant. It is an essential part of language learning. The native language magnet (NLM) theory (e.g., Kuhl, 1991; Kuhl & Iverson, 1995) argues that human infants in the first 6 to 9 months of their lives set up prototypes of the speech sounds they hear in their environment. The prototype would be the ideal realization of a sound in its category, located at the largest possible distance away from the prototypes of competing categories in the same sound space. Human infants are born with pre-wired auditory categories, which are subdivisions of the auditory space represented somewhere on the cortex, where each category is an area within which differences between sounds are (relatively) difficult to perceive and which is bounded by corridors of high sensitivity to auditory differences, so-called natural boundaries. As a first approximation to category formation, the infant sorts the incoming sounds into the pre-wired categories. Within the categories the infant then notices that the distribution of the category members is not uniform but gravitates towards one specific spot somewhere central within the category. This gravitational point corresponds to the category prototype. Prototypical exemplars of a category are easier to categorize, easier to remember and are preferred over other instances of the same category. The basic tenet of the NLM model is that sounds are

[12] This chapter is an extended version of Afshar, N. & V. J. van Heuven (2022). Perceptual assimilation of English vowels by monolingual and bilingual learners in Iran. *Argumentum*, 18, 172 –191.
DOI: 10.34103/ARGUMENTUM/2022/9

more difficult to discriminate from each other as they are closer to the prototype. It is as if the prototype draws similar sounds towards it, and that the magnetic pull gets stronger as the sounds are closer to the prototype. It follows from this account that two sounds that find themselves halfway between two competing prototypes (i.e., adjacent categories), are relatively easy to discriminate (Van Heuven, 2022).

Given this state of affairs, it is important that we establish how the sounds of a foreign language we aim to learn, are mapped onto the prototypes in our native language. The Perceptual Assimilation Model was designed to provide a typology of assimilation patterns that may occur when a listener with native language L1 is confronted with the sounds of an unknown language L2. The PA model recognizes three types of non-native sounds:

C or *Categorized*. The foreign sound is accepted as an instance (whether good or poor) of one (and only one) of the sound categories in the native system. For instance, English /i/ is perceived by Dutch listeners as a token of the vowel sound /i/ in their native language in spite of a difference in duration (Collins & Mees, 1984).

U or *Uncategorized*. These sounds fall in between two or more categories of the listener's L1. They are poor tokens of multiple adjacent categories in the L1. As just one example, British English /ʌ/ was uncategorized for Italian listeners, with responses shared between the Italian /a/ and /ɔ/ vowels (Sisinni & Grimaldi, 2009).

N or *Non-assimilable*. Sounds are non-assimilable when they are unlike any category in the listener's L1, i.e., are outside the phonological space of the learner's L1 (e.g., African click sounds are so unlike any consonant type in English that the English listener thinks the speaker is clapping his hands or flicking his fingers while talking, Best, McRoberts & Sithole, 1988).

This basic tripartite division of sounds is then used to describe a number of assimilation scenarios that may apply when a pair of foreign sounds has to be discriminated by a native listener of L1. Here we will summarize four scenarios, the ones most often found:

SC: *Single Category scenario*. Two different foreign sounds (i.e., contrastive phonemes in L2) are classified as equally good instances of one single category in the listener's L1. The prediction is that such a contrast in the L2 will escape the learner's attention and will constitute a persistent learning problem.

TC: *Two Categories scenario*. Two contrastive sounds {x, y} in the L2 are assimilated in a one-to-one fashion to two contrastive sounds {a, b} in the learner's L1. The prediction is that the learner will easily discriminate between the two foreign sounds, and that the contrast will not present a learning problem.

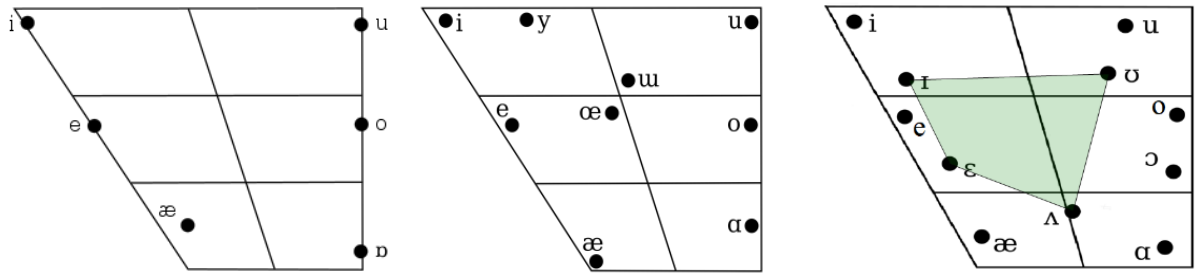
CG: *Category Goodness scenario*. Two different (contrastive) sounds in the L2 both map onto a single category in the learner's L1 but one matches the L1 category clearly better than the other. The listener will quickly notice the difference in goodness between the two foreign sounds, which will guide him to set up a split in his native category to accommodate the contrast (learn a new category boundary).

UC: *Uncategorized/Categorized* scenario. One of two contrastive sounds in the L2 is assimilated to a category in the L1 while the other sound remains Uncategorized. This is like the CG scenario but the category of the less typical member is now undecided. Discrimination will be reasonable to good but the formation of the new category will be more difficult because it contains parts of multiple adjacent categories in the L1.

The primary aim of the present study is to determine how the monophthongal vowels of American English are perceptually assimilated by EFL learners in Iran. More specifically, I study two groups of Iranian learners of EFL. One group is monolingual (at the onset of EFL learning) and speaks Persian as the first and only native language. A second group was tested in the North-West of Iran near the border with Azerbaijan. At the onset of EFL learning, these learners were early bilinguals with two native languages, i.e., Azerbaijani and Persian. These learners have typically acquired Azerbaijani as their first language at home, and then learned Persian from the age of four onwards at school, where Persian is the language of instruction. When these early bilinguals participated in the present perceptual assimilation study, they were around 16 years of age (mean age = 16.9 years, against 16.5 for the monolinguals), and estimated their oral skills about equal (9 or better on a scale from 0 to 10) in both their native languages. Although quite a number of studies have been done on the perceptual assimilation of English sounds, including vowels, to the native languages of groups of monolingual learners of EFL, the perceptual assimilation by early bilingual and multilingual learners is understudied. The secondary aim of the present study is to determine whether the monolingual Persian EFL learners have different assimilation patterns than the bilingual EFL learners when the latter are instructed to map the foreign vowels onto the vowels of Persian. A related question is whether the English vowels assimilate in the same or in a different way to the vowels that are shared between Persian and Azerbaijani. Comparing the results for the bilinguals with those obtained for their monolingual Persian peers may tell us if the extra vowels in the Azerbaijani set affect the task performance in the PAM test, whether positively or negatively.

4.2. Characterization of the vowel systems involved

The monophthongal vowel system of Persian distinguishes three degrees of height (high, mid, low) and two degrees of backness (front, back). Lip rounding is unmarked, i.e., typologically normal, such that front vowels are pronounced with spread lips and back vowels with rounded lips. Persian has no contrast based on vowel duration (short, long) or tenseness (lax, tense). The approximate positions of the six vowels in the IPA vowel chart is shown in panel A of Figure 2.1, which I repeat here for the reader's convenience as Figure 4.1.



A. Persian B. Azerbaijani C. American English

Figure 4.1. IPA vowel diagrams for the vowel inventories of Modern Persian (A, Majidi & Ternes, 1999), Azerbaijani (B, Ghaffarvand Mokari & Werner, 2016, p. 509) and American English (C, modified from Manell, Cox & Harrington, 2009; Ladefoged & Johnson, 2011, p. 197). The shaded quadrilateral connects the four short (sometimes called 'lax') vowels.

The vowel system of Azerbaijani is almost the same as that of Persian as far as the peripheral vowels (also called edge vowels) is concerned but it is augmented with three vowels in the central region of the vowel space, yielding a total of nine, as shown in Figure 4.1B. The coupling of backness and lip rounding is more complex in Azerbaijani in that the three central vowels have a-typical lip rounding. Phonologically, Azerbaijani /y/ and /œ/ are front vowels (as they are in Turkish) but with (marked) lip rounding. The phonologically high /ɯ/ is a back vowel with marked spread lips. Like Persian and Turkish (closely related to Azerbaijani with a fair degree of mutual intelligibility), Azerbaijani has no length or tenseness contrast in the vowels.

The pure (monophthongal) vowel system of American English is more complex than that of either Persian or Azerbaijani. Although considerable regional variation exists, most varieties distinguish eleven vowels that are normally analyzed as monophthongs, as illustrated in Figure 4.1C. I adopt here the analysis of the American English vowel system given by Celce-Murcia et al. (2010), which was also followed by Yavaş (2011). This system has five unrounded front vowels and four rounded back vowels with four degrees of height: high, high-mid, low-mid and low. The lowest back vowel /ɑ/ is described as unrounded. The monophthongs can be split into a group of seven long vowels, and a smaller group of four short vowels, which not only have shorter durations, but also assume a rather more centralized vowel quality, and have no diphthongization. The long vowels are articulated closer to the outer perimeter of the vowel

space, and are often referred to as ‘tense’, while the more centralized short vowels are also called ‘lax’ (e.g., House, 1961; Strange et al. 2004; Wang & Van Heuven, 2006; Celce-Murcia et al., 2010; Yavaş, 2011). The tense vs. lax properties distinguish between the members of the high-mid vowel pairs /e, ɪ/ and /o, ʊ/. There is one central monophthong: mid-low /ʌ/. The (mid-)low back vowels are best analyzed as long and tense vowels as in *law* /lɔ/ and *father* /fɑðɚ/. The high-mid tense vowels /e/ and /o/ are semi-diphthongs in most varieties of English, including American English. They are grouped here with the monophthongs because the slight diphthongization is not essential for their identification, and when pronounced as monophthongs (as they are in some varieties, e.g., Scots English) they remain distinct from each other and from all other vowels – which is not the case for the full diphthongs /ai, ɑu, ɔi/. Here, too, I follow the analysis adopted by Celce-Murcia et al. (2010: 114–116) and Yavaş (2011, p. 77–79). Also, in line with these authors, I exclude all vowels that only occur as positional allophones before coda /r/, such as [ɚ], which is listed among the monophthongs by, e.g., Ladefoged and Johnson (2011).

The auditory analyses of the vowel systems of Azerbaijani and of Persian in Figure 4.1A-B, have been complemented with acoustic measurements of vowel formants and duration in order to map out the phonetic details of the respective vowel spaces. Ansarin (2004) measured F1 and F2, but not the duration, in /hVd/ words spoken by 12 Persian women. Ghaffarvand Mokari, Werner and Talebi (2017) measured F1, F2, F3 and duration of the six monophthongs of Tehrani Persian (28 male, 25 female speakers) in /bVd/, /dVd/ and /hVd/ monosyllables. Aronov et al. (2017) measured vowel formants and duration in both free conversation and read-aloud word lists for two Tehran speakers of Persian. A subsequent analysis of informal Persian speech by Jones (2019) reveals that there that there is no significant length distinction between any pair of vowels (mean vowel durations between 56 and 78 ms, for /e/ and /u/ respectively), that there is regional variation in the vowel space, and that the low back vowel may be better characterized as a diphthong [ɔɒ] and is higher than often assumed. The latter finding is consistent with the earlier measurements by Ansarin (2004) and Aronow et al. (2017). The general configuration of the vowels in the acoustic space is in agreement with Figure 4.1A.

F1 and F2 (but not the duration) of the nine vowels of Azerbaijani spoken by 30 male and 30 female speakers were measured in /ʃVr/ monosyllables by Peivasti (2012). No breakdown by gender was presented. Ghaffarvand Mokari and Werner (2016) measured F1, F2, F3 and duration of the nine vowels of Azerbaijani produced in /bVd/ words by 20 male and 23 female monolingual speakers. The acoustic vowel plots correspond well with the traditional vowel diagram in Figure 4.1B. The perceptual assimilation of the vowels of Standard Southern

British English (SSBE) by Azerbaijani monolinguals was subsequently studied, and predictions of perceptual and acoustic confusion of SSBE vowels by 20 male and 20 female Azerbaijani EFL learners were tested (Ghaffarvand Mokari & Werner, 2017).

Pillai and Delavari (2012) reported F1, F2 and duration in eight British English pure vowels spoken in monosyllables in a fixed carrier by 13 Persian EFL speakers (7 male, 6 female). Lack of spectral and temporal contrast was observed for the tense-lax pairs /iː~ɪ/ and /uː~ʊ/. The lax pair /ʌ~ɒ/ was conflated but the quality difference between /e~æ/ was upheld.

A contrastive acoustic study of nine monophthongs of American English (excluding the semi-diphthongs /eː, oː/) and the six Persian monophthongs was reported by Sadeghi and Bigdeli (2018). Tokens were sampled from existing databases of read-out continuous speech, equally divided over (an unspecified number of) male and female speakers. Formants (but not durations) of individual vowel tokens were visualized in scatterplots but no centroids or dispersion measures were provided. Subsequently, Bigdeli and Sadeghi (2020) asked 15 listeners to identify each of nine English monophthongs (three different tokens per vowel) as a vowel of Persian, with two or three response alternatives. The assimilation pattern (English > Persian) was as follows: /iː/ > /i/, /ɪ, ɛ/ > /e/, /æ/ > /æ/, /uː, ʊ/ > /u , and /ʌ, ɔ, ɑː/ > /ɑ/. No typicality judgments were given.

Mirahadi et al. (2018) measured F1, F2 and F3 (but not duration) in the six vowels of Persian as spoken by 25 male and 25 female Azerbaijani-Persian bilingual adults. No comparison was made between bilinguals and monolingual Persian speakers.

There are no contrastive studies yet of the perceptual assimilation of the vowels of American English by monolingual Persian vs. bilingual Azerbaijani/Persian listeners. The present paper aims to fill this lacuna. There is also a pedagogical motivation for the study. If AZ and Persian children experience different pronunciation difficulties in EFL, then this would complicate the teaching of EFL. So we need to know whether such complications are required or not.

4.3. Methods

4.3.1. Materials

We selected two male speakers from a larger group of 20 native speakers of American English from the recordings collected by Wang and Van Heuven (2006, 2014). These were the only two speakers in the set who observed a proper contrast between the (mid-)low back vowels /ɑ/ and /ɔ/. Speakers had been recorded on digital audio tape (DAT) in a sound-insulated recording booth through a Sennheiser MKH-416 microphone. Materials were later

downsampled (16 KHz, 16 bits. For each speaker the following set of 11 monosyllabic words or phrases was excerpted from the fixed carrier *Now say … again*, using the digital waveform editor in the Praat (version 6.1.05) speech processing software (Boersma & Weenink, 2019; Boersma & Van Heuven, 2001): *heed* /hid/, *hid* /hɪd/, *hayed* /hed/, *head* /hɛd/, *had* /hæd/, *hud* /hʌd/, *hod* /hɑd/, *hawed* /hɔd/, *hoed* /hod/, *hood* /hʊd/, and *who'd* /hud/, following the established practice in, e.g., Peterson & Barney (1952) and Hillenbrand et al. (1995). Acoustic details of the 22 tokens used in the PAM test can be found in Appendix 4.1. The Praat MFC scripts are included as Appendix 4.2.

4.3.2 Participants

Two groups of listeners participated in the experiment. The first group comprised 22 native speakers (11 male, 11 female) of Modern Persian. They were secondary school pupils in Tehran with a mean exposure to (American) English of roughly 6 years in a school setting. The second group consisted of 27 early bilingual listeners (11 male, 16 female) with Azerbaijani and Persian as their first two languages (see above for more explanation on their language background). The bilinguals were comparable to the monolinguals in all relevant aspects (age, exposure to English, level of education). They were tested in secondary schools in the city of Marand in the East Azerbaijan Province in the North West of Iran.

All listeners filled in the Language Experience and Proficiency questionnaire (LEAP-Q) developed by Marian, Blumenfeld and Kaushanskaya (2007). This questionnaire asks the participant to estimate their experience with and exposure to the languages they command, and to self-rate their proficiency and (non-)nativeness in each of these languages.

4.3.3. Procedure

Participants sat at a table in a small-size quiet office. They listened over a good quality headset (Sennheiser PC3) to the stimuli being played to them from a notebook computer. The monolingual participants saw a screen as shown in Figure 4.2A. On the screen, six common Persian words were printed in Arabic script (as is standard in the Iranian school system), which were identical in all respects except for the vowel. Each response button therefore corresponded to one of the six vowels of Persian, arranged in two columns (left: front vowels, right: back vowels) by three rows (top: high vowels, middle: mid vowels, bottom: low vowels). Participants were instructed to listen to each English word being played just once, and to decide which of the six vowels contained in the words written in the response buttons resembled the vowel they had just heard most. Participants responded by clicking the button of their choice with a mouse

pointer. Immediately after the mouse click the goodness scale at the bottom of the screen turned from grey to yellow. Participants then decided whether the vowel they had selected was a poor or a good token of its category on a 5-point scale (where 5 signified 'good'). Then the screen was reset, and after 2 seconds the next stimulus word was made audible.

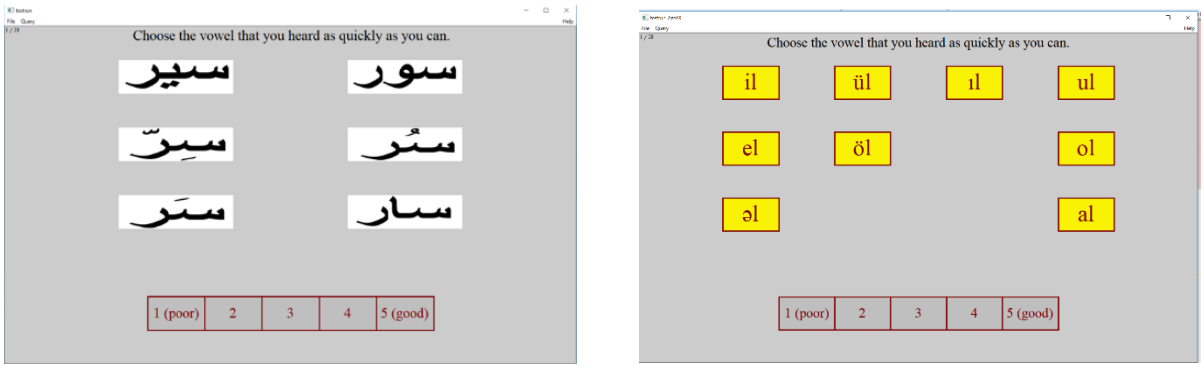


Figure 4.2. Screens showing the six response categories (in Arabic script) for the Persian version of the perceptual assimilation test (panel A, left). Panel B (right) shows the screen used for the Azerbaijani version of the test, with nine response categories in Azerbaijani orthography. The goodness scale at the bottom of the screen was activated only after the participant clicked a response vowel.

The 22 stimulus types were presented to each listener twice, in different quasi-random orders excluding immediate repetition of the same stimulus. A counter in the top-left corner of the screen kept the participants informed of their progress. No other feedback was given. Stimulus presentation and response collection was done through the Praat MFC script.

The bilingual participants took the assimilation test twice in immediate succession on the same day, the first time the Persian version (Figure 4.2A) and the second time in Azerbaijani. In the latter version, they saw the screen shown in Figure 4.2B, which contained nine rather than six vowel response buttons, with the keywords printed in standard Azerbaijani spelling. All instructions were in the target language, i.e., English (see Appendix 3.2 for details). Again, the keywords were minimally different short words which differed in the vowel only. The words were arranged in four columns, corresponding, respectively, to front spread, front rounded, back spread and back rounded vowels. As for Persian, high, mid and low vowel buttons were listed on the top, middle and bottom row, respectively. The Azerbaijani spelling resembles that of Turkish, with the addition of the symbol 'ə', which in Azerbaijani spelling stands for open front [æ].

4.4. Statistical considerations

The present study is not an experiment. It is a data-oriented, exploratory study in which no hypotheses are being tested. The stimulus variable is the vowel type that is presented for identification to the participant. There are 11 different vowel types. There are

three response variables, i.e., the vowel category (forced choice from 6 possibilities in Persian response mode and 9 categories in Azerbaijani mode), the typicality (or 'goodness' judgment (with forced choice from 5 scale positions), and the response time (in milliseconds) that elapses between the moment the stimulus sound is made audible and when the respondent clicks the goodness button. This latter variable will not be analyzed because, in retrospect, it will be unclear whether the response time is determined by the difficulty to choose the vowel category *per se* or to decide on the goodness. The primary response is categorical, with a forced choice from 6 or 9 discrete categories. The number of responses in each category can be counted, and expressed as a percentage relative to the maximum number of responses given to the stimulus. The category with the largest number of responses for a particular stimulus vowel is the *modal* category. The typicality judgment will be treated as a continuous variable on an interval scale, as is customarily done for typicality judgments (e.g., Kuhl & Iverson, 1995). Since Guion et al. (2000), perceptual assimilation studies often use the so-called Fit-index, which is a compound measure of how well a non-native L2 sound fits a native L1 sound category. The Fit-index is computed by taking the goodness judgment and multiplying it by the proportion of responses given to the category. For instance, if the American-English vowel /i/ (as in *heed*) is perceived as a token of Persian /i/ in 87 percent of the cases (proportion is .87) with a mean goodness judgement of 4.3 (5 = perfect token), the Fit-index equals 4.3 × .87 = 3.7. The stimulus vowel is also perceived as a token of Persian /e/ but in only the remaining 13 percent of the cases with a mean goodness of 3.5. Persian /e/ then has a Fit-index of 3.5 × .13 = .5. The Fit-index is treated as a continuous variable at the interval level of measurement. The index is used to classify a vowel in the foreign language as either Uncategorizable or Categorizable in the listener's native language, and, in the latter case, whether it should be considered a Good, Fair, or Poor token of the L1 category. This four-way split requires three boundaries or cut-off values. To compute these cut-off values, the Fit-indexes are collected for only the modal response categories found for the stimulus (= L2) sounds, i.e., 11 modal response categories, which should have the highest Fit-index found for each L2 sound. The cut-off between Good and Fair is then the mean of the modal Fit-indexes plus 1 standard deviation. The cut-off between Fair and Poor is at the mean of the modal Fit-indexes, while any foreign sound at less than 1 SD below the mean is classified as uncategorized. This four-way split has recently been proposed (e.g., Wang & Chen, 2019, 2020) to replace the more traditional split based on rather arbitrary but simple decision rules based on response prevalence only, such as modal response ≥ 50% for Good, 25-50% for Poor and > 25% Uncategorized. I will use both approaches in the analysis in the next section.

4.5. Results

Table 4.1A presents the perceptual assimilation results for the early monolingual Persian listeners. The 11 American English stimulus vowels are in the rows, while the six vowels of Persian they could be matched with are in the columns. Given 22 Persian listeners and 4 tokens of the same vowel (produced by two different male speakers), the maximum number of votes for one particular response would be 88. The counts in the cells of the table have been converted to percentages in order to facilitate comparison with the results obtained by the 27 early bilingual listeners (yielding a maximum of 108 responses) in Tables 4.1B-C. Three columns in Table 4.1A-B have been greyed out; these are response vowels that occur in Azerbaijani only. Green (dark-shaded) cells in the table contain responses with ≥50% prevalence among the listeners. Yellow (light-shaded) cells have responses on which between 25 and 50% of the listeners converged. When <25% prevalence was obtained for a response category, the cell has been left white. The left-hand smaller-sized number in each cell is the mean of the goodness ratings given to the particular stimulus-response pair. Next to it we specify the Fit-index, which is defined as the prevalence of the response category (expressed as a proportion) multiplied by the mean goodness rating (see § 4.4).

Any non-native vowel that is assimilated to a native category in ≥50% of the responses, is considered *Categorized* (C). When there are two (or more) response categories with ≥25% of the votes, thereby adding up to ≥50% of the responses, the non-native vowel is considered *Uncategorized* (U). In all other cases, the non-native vowel is considered *Non-assimilable* (N). An alternative method categorizes the stimulus vowel as a Good, Fair, or Poor token of an L1 vowel depending on the distribution of the Fit-index over the modal responses per AE vowel (see § 4.4). When ≥1 SD above the mean of the modal Fit-indexes, the AE vowel is a Good (G) token of the L1 vowel (green), between the mean and +1 SD it is a Fair (F) token (yellow), between the mean and –1 SD it is a Poor (P) token (orange), and any Fit-index <1 SD leaves the stimulus vowel unclassified (grey). Table 1 presents the results in terms of both methods.

Table 4.1. Perceptual assimilation of eleven vowels of American English to the six vowels of Persian by early monolingual Persian listeners (A), and by early bilingual Azerbaijani/Persian listeners (B). Panel C shows the results of the bilinguals when instructed to assimilate the English vowels to the nine vowels of Azerbaijani. The three added vowel response categories have been greyed out in the Persian response mode. The prevalence of a response vowel is expressed as a percentage (large print). Green cells contain responses with ≥50% agreement. Yellow cells contain responses with 25 to 50% agreement. The numbers in small print are the mean goodness ratings for the particular stimulus-response pair (left) and the Fit-index (right, bolded). Categorization based on Fit-index is indicated in the right-hand margin (for details see text).

A. Monolingual Persian (N = 22)	Stimulus vowel			Response vowel: i	Y	e	œ	æ	ɑ	o	ɯ	u	Fit-index/ Category	
	heed	i	C	**87** 4.3 **3.7**		13 3.5 .5							3.7	G
	hid	ɪ	C	**39** 3.9 **1.5**		**60** 3.9 **2.3**						1 3.0 .0	2.3	P
	hayed	e	C	**39** 3.6 **1.4**		**61** 3.5 **2.1**							2.1	P
	head	ɛ	C	2 4.0 .1		**90** 4.0 **3.6**		6 3.8 .2	1 4.0 .0				3.6	F
	had	æ	C	2 4.5 .1		**37** 4.1 **1.5**		**61** 4.4 **2.7**					2.7	P
	hud	ʌ	U			7 2.8 .2		1 4.0 .0	**37** 4.0 **1.5**	**39** 4.2 **1.6**		15 4.1 .6	1.6	U
	hod	ɑ	C					4 3.3 .1	**96** 4.5 **4.3**				4.3	G
	hawed	ɔ	C						**95** 4.4 **4.2**	5 4.5 **0.2**			4.2	G
	hoed	o	C						1 3.0 .0	**52** 3.8 **2.0**		**46** 3.9 **1.8**	2.0	P
	hood	ʊ	C			1 3.0 .0		2 4.5 .1	1 4.0 .0	**56** 4.2 **2.4**		**39** 4.1 **1.6**	2.4	P
	who'd	u	C							19 3.8 **0.7**		**81** 4.3 **3.5**	3.5	F

Mean = 3.0; SD = 1.01

B. Bilingual Persian (N = 27)	Stimulus vowel			Response vowel: i	y	e	œ	æ	ɑ	o	ɯ	u	Fit-index/ category	
	heed	i	C	**98** 4.0 **3.9**		2 2.5 **0.1**							3.9	G
	hid	ɪ	C	**58** 3.7 **2.1**		**38** 3.5 **1.3**				3 3.0		1 4.0 .0	2.1	P
	hayed	e	C	**25** 3.7 .9		**74** 3.6 **2.7**			1 1.0 .0				2.7	P
	head	ɛ	C	5 4.0 .2		**93** 3.7 **3.4**		2 3.5 .1	1 3.0 .0				3.4	F
	had	æ	C	2 2.5 .1		**41** 3.8 **1.6**		**56** 4.0 **2.2**		1 4.0 .0			2.2	P
	hud	ʌ	N	2 4.0 .1		9 2.9 .3		1 2.0 .0	20 3.8 .8	**48** 3.8 **1.8**		19 3.3 .6	1.8	U
	hod	ɑ	C	1 4.0 .0				2 3.5 .1	**95** 4.1 **3.9**	2 4.0 .1			3.9	G
	hawed	ɔ	C			2 2.5 .1		1 3.0 .0	**84** 4.2 **3.5**	11 3.5 .4		2 3.5 .1	3.5	F
	hoed	o	C	2 4.0 .1		1 4.0 .0				**61** 4.0 **2.4**		**36** 3.8 **1.4**	2.4	P
	hood	ʊ	C	2 4.0 .1		3 4.0 .1			3 3.3 .1	**41** 3.9 **1.6**		**52** 3.9 **2.0**	2.0	U
	who'd	u	C	1 4.0 .0					1 1.0 .0	7 3.8 .3		**91** 4.1 **3.7**	3.7	G

Mean = 2.9; SD = 0.82

C. Bilingual Azerbaijani (N = 27)	Stimulus vowel			Response vowel: i	y	e	œ	æ	ɑ	o	ɯ	u	Fit-index/ category	
	heed	i	C	**75** 3.8 **2.9**		11 3.6 .4					14 3.8 .5		2.9	F
	hid	ɪ	C	**38** 3.7 **1.4**	2 3.5 .1	**36** 3.5 **1.3**	3 3.0 .1	2 2.0 .0			18 3.8 .7	2 4.0 .1	1.4	P
	hayed	e	C	7 2.9 .2		**77** 3.8 **2.9**	2 2.5 .1	7 3.4 .2	1 4.0 .0		5 3.6 .2	1 5.0 .1	2.9	F
	head	ɛ	C	6 3.1 .2	1 3.0 .0	**77** 3.8 **2.9**	1 4.0 .0	8 3.8 .3			6 3.2 .2	1 4.0 .0	2.9	F
	had	æ	C	2 3.0 .1		22 3.4 .7	1 3.0 .0	**65** 3.8 **2.5**	6 4.1 .2	1 4.0 .0	3 2.7 .1		2.5	F
	hud	ʌ	N	2 2.0 .0	17 3.7 .6	6 3.0 .2	16 3.6 .6	4 3.0 .1	17 3.6 .6	18 3.5 .6	9 3.3 .3	12 3.7 .4	0.6	U
	hod	ɑ	C	1 3.0 .0			1 1.0 .0	9 3.0 .3	**83** 4.0 **3.3**	1 4.0 .0	3 4.3 .1	2 2.5 .1	3.3	G
	hawed	ɔ	C		1 4.0 .0		1 4.0 .0	6 3.2 .2	**83** 3.9 **3.2**	1 4.0 .0	5 4.4 .2	4 2.8 .1	3.2	G
	hoed	o	C		15 3.7 .6		**48** 3.8 **1.8**			**28** 3.8 **1.1**	2 2.0 .0	7 3.8 .3	1.8	P
	hood	ʊ	C	1 4.0 .0	24 3.7 .9	6 3.0 .2	**30** 3.4 **1.0**	2 3.5 .1	1 3.0 .0	15 3.6 .5	4 3.0 .1	19 3.8 .7	1.0	U
	who'd	u	C		**38** 3.8 **1.4**	1 5.0 .1	13 3.4 .4			24 3.8 .9	1 4.0 .0	23 3.8 .9	1.4	P

Mean = 2.2; SD = 0.96

When we apply the simple decision rules to the results in Table 4.1, we observe that ten out of eleven American English vowels are either C; the exception is the vowel [ʌ] (as in *but*), which is not assimilated to any of the vowels of either Persian or Azerbaijani by the bilingual listeners (i.e., N), whether responding in the Persian mode or in the Azerbaijani mode.

Interestingly, [ʌ] is U in the perception of the monolingual Persian listeners, where the responses are equally divided between Persian /ɑ/ and /o/. Given that [ʌ] is a central vowel, it is not surprising that the three central vowels in Azerbaijani compete with /ɑ/ and /o/ and together draw 42% of the responses, thereby depleting other response categories. When the bilinguals respond in the Persian mode, only /o/ is chosen in >25% of the responses, without reaching the majority criterion for C. Here the bilinguals differ from the monolinguals even when they respond in the same mode, i.e., Persian.

Given an inventory of eleven there are (11 × 10) / 2 = 55 possible contrasts between American English monophthongs. The majority of these pairwise contrasts are between vowels that are non-adjacent in the English vowel space, and which assimilate to different vowel categories in either Persian or in Azerbaijani. For instance, in Table 4.1A, English /i/ is assimilated to Persian /i/ in 87% of the judgments with a goodness rating of 4.3 and a Fit-index of 3.7 (Good). English /u/ assimilates to Persian /u/ with 81% with a goodness of 4.3 and Fit-index of 3.5 (Fair). This is an example of a Two Category (TC) assimilation scenario, for which the prediction is that the members of the contrast are easily discriminated by Persian learners of English.

In all, there are 40 TC pairs in Table 4.1A. Of the remaining 15 contrasts, ten involve the U vowel /ʌ/, which is paired with one of the other ten vowels, all of which are C. Since one vowel in the UC pair maps onto a known category in the L1 while the other vowel remains a difficult choice, the members of the pair should be rather easy to discriminate. The prediction, therefore, is that the Persian learners will quickly realize that the vowel /ʌ/ is different from any vowel in their own language – so that they are prompted to set up a new category for the non-native sound.

Five more contrasts are between vowels that are adjacent in the English vowel space. Three of these are in a Single Category scenario: both /ɪ, e/ assimilate to Persian /e/ (Fit-indexes: 2.3, 2.1 = Poor-Poor), both /ɑ, ɔ/ map onto Persian /ɑ/ (Fit-indexes: 4.3, 4.2 = Good-Good), and both /o, ʊ/ assimilate to Persian /o/ (Fit-indexes: 2.0, 2.4 = Poor-Poor). Learners will find the SC members difficult to distinguish, and will need a lot of time and effort to learn how to split their single native category into two new, smaller categories. Without explicit instruction, the learner will never be aware that the contrast exists in the foreign language and will conflate the categories forever (Flege, 1995).

Finally, two contrasts in Table 4.1A are between adjacent English vowels, one of which is a convincing token of a Persian vowel, while the other, although it assimilates to the same category, is a clearly poorer exemplar of it. This Category Goodness (CG) scenario applies to

the contrasts /ɪ, ɛ/ and /e, ɛ/. All three vowels map onto Persian /e/, but /ɛ/ is perceived as a much better exemplar (Fit-index: 3.6 = Fair) of Persian /e/ than either /ɪ/ (too high, Fit-index: 2.3 = Poor) or /e/ (too long and diphthongal, Fit-index: 2.1 = Poor) is. Fairly good discrimination is predicted for CG pairs, and it will be easier for the learner to discover that and how his/her single native category has to be subdivided into one part that matches with exemplars rather close to the Persian prototype, while the other part is more remote from the prototype and contains non-typical exemplars only.

When the bilinguals respond in the Persian mode, 41 of the 55 pairwise contrasts are of the TC type. The mid-low central vowel /ʌ/ is N, i.e., is outside the Persian vowel space – but only just. As a result, all ten contrasts involving /ʌ/ are of the NC type – for which scenario the same prediction holds as for the UC type with the monolingual EFL learners. There are no CG contrasts in Table 3.1B, but four contrasts are of the SC type: /i, ɪ/, which both assimilate to Persian /i/ (Fit-indexes: 3.9, 2.1 = Good-Poor), /e, ɛ/ assimilate to /e/ (Fit-indexes: 2.7, 3.4 = Poor-Fair), /ɑ, ɔ/ assimilate to /ɑ/ (Fit-indexes: 3.9, 3.5 = Good-Fair), and /u, ʊ/ assimilate to /u/ (Fit indexes: 3.7, 2.0 = Good-Uncategorized).

In Table 4.1C, where the bilinguals respond in the Azerbaijani mode, the number of native vowel categories is nine rather than six. Since there are more response categories, it is harder for a non-native vowel to meet the 25% or 50% lower-bound criteria. This shows up as a smaller number of non-native vowels that are C: the vowels /ɪ/ and /o/ are U, while /ʌ/, /u/ and /ʊ/ are N. This reduces the number of easy TC contrasts to 13. Only two contrasts are of the difficult SC type, i.e., /e, ɛ/ both assimilate to Azerbaijani /e/ (Fit-indexes: 2.9, 2.9 = Fair-Fair), and /ɑ, ɔ/ to Azerbaijani /ɑ/ (Fit-indexes: 3.3, 3.2 = Good-Good). The remaining 40 contrasts are classified as UC (14), NC (20), NU (5) or NN (1). Interestingly, there are no instances of the CG scenario. It should be noted that the classification of several of these contrasts into scenarios may be unduly rigid. For instance, /ʊ/ would have been counted as U (between /œ/ and /y/) with 1 point more /y/-responses; /u/ would have been U (between /y/ and /o/) with 1 point more /o/-responses. Similarly, /o/ would have been C instead of U with 2 points more /œ/-responses.

Overall, the Fit-indexes for the modal responses per stimulus vowel) are lower when the assimilation task involves a choice from nine response alternatives (Azerbaijani) than when the same (early bilingual) listeners respond in the Persian mode with six response categories (means: 2.87 vs. 2.27). Monolingual listeners have slightly higher Fit-indexes than the bilinguals when both respond in Persian mode: 3.00 vs. 2.87). The overall effect is significant by a one-way Repeated-Measures Analysis of Variance, $F(2, 20) = 10.4$ ($p = .001$, $p\eta^2 = .510$).

Bonferroni post-hoc tests (α = .050) confirm that the Azerbaijani Fit-indexes are significantly lower than those for the two Persian response sets, which do not differ significantly from each other. Moreover, the Fit-indexes for the monolinguals and bilinguals (responding in Persian mode are very strongly correlated (r = .917, p < .001). The correlations are appreciably lower when the Azerbaijani response set is compared (r = .700, p = .017 with monolinguals, r = .693, p = .018 with bilinguals).

Table 4.2 summarizes the most important differences in the perceptual assimilation patterns observed between monolinguals and bilinguals in Persian response mode, as well as between the bilinguals in Persian versus Azerbaijani mode. The table exemplifies that the monolinguals have different assimilation patterns than the bilinguals, even when the latter respond in Persian mode. It suggests that the major difficulties will be in the contrast between the short and long high-mid vowels (both front and back) for the monolinguals, while keeping the high-mid vowels separate from the high vowels as well as from the (mid-)low vowels. The bilinguals, however, are predicted to experience difficulties in distinguishing the high vowels from the mid-high vowel on the one hand, and the mid-high from the mid-low vowels on the other hand. This also suggests that the bilinguals attend to differences in vowel quality and not so much to differences in vowel duration. The bilinguals' assimilation patterns for AE front vowels are the same, whether they respond in Persian mode or in Azerbaijani mode. For the back vowels, the Azerbaijani response mode generates a different pattern due to the attractiveness of two of the central vowels (see above).

Table 4.2. Summary of Same Category (SC) contrasts in American English (AE) vowels as perceived by monolingual Persian EFL learners (left) and by bilingual EFL learners in either Persian (L2) or Azerbaijani (L1) mode. Two AE vowels joined by a brace denote a SC contrast (reddish cells); a yellow cell denotes a Category Goodness (CG) contrast or similar.

	Monolinguals			Bilinguals			
	Persian	AE		Persian	Azerbaijani	AE	
Front vowels		*heed*	} /i/		/i/ {	*heed*	i
	/e/ {	*hid*				*hid*	ɪ
		hayed	} /e/		/e/ {	*hayed*	e
		head				*head*	ɛ
Back vowels		*who'd*	} /u/			*who'd*	u
	/o/ {	*hood*			/œ/ {	*hood*	ʊ
		hoed				*hoed*	o
	/ɑ/ {	*hawed*	} /ɑ/		/ɑ/ {	*hawed*	ɔ
		hod				*hod*	ɑ

The low-back AE vowels /ɔ, ɑ/, both of which are pronounced long (see Appendix, Figure A2), project onto the same vowel in Persian as well as in Azerbaijani for both groups of EFL learners. Although this predicts a learning problem for the EFL learners, the general advice is not to make this a priority in the teaching of American English pronunciation. The /ɔ, ɑ/ contrast has a low functional load, and is subject to the low-back vowel merger in the pronunciation of most native speakers of American English, who then pronounce all low back vowels as low /ɑ/ (e.g., Ladefoged & Johnson, 2011: 212–213; Carley & Mees, 2020).

4.6. Conclusion and discussion

The first question we asked was how the two groups of EFL learners assimilate the vowels of English to those of Persian, and – in the case of the early bilinguals – to those of Azerbaijani. The results bear out that of the eleven vowels of American English the great majority (i.e., ten) are Categorized in Persian, whether the listeners are monolingual or bilingual when they began learning English as a foreign language. The only Uncategorized vowel is /ʌ/, which makes sense because Persian has no central vowels. The critical SC contrast scenario is predicted for Persian EFL learners in only three vowel pairs, all of which involve members that are adjacent in the vowel space. These are the American English vowel pairs /ɪ, e/, /ɑ, ɔ/ and /o, ʊ/. One SC contrast, between the half-open back vowels /ɑ, ɔ/, can be ignored as a learning problem, because this contrast is also absent in many regional varieties of American English, including Californian English (Ladefoged & Johnson, 2011, p. 212–213). Persian has no tense-lax vowel contrast (nor does Azerbaijani). This is reflected by the SC assimilation pattern observed for /e, ɪ/ and /o, ʊ/ for the monolinguals, and in /i, ɪ/ and /u, ʊ/ for the bilinguals. The bilinguals have another SC pair in /e, ɛ/, which would suggest that their Persian category /e/ is larger than its counterpart is for the monolinguals – while the bilinguals' category for /i/ would be smaller, possibly due to competition from Azerbaijani /y/. The assimilation of the tense-lax pairs in the high/mid-section of the vowel space would /be the most important difference between the monolingual and bilingual participants. This may be due to some form of interaction on the part of the bilinguals with the competing vowel system of Azerbaijani. Azerbaijani has three central vowels in the high-mid part of the vowel space, so that the dispersion area for the peripheral vowels /i, e, o, u/ is more limited in the bilingual Persian vowel system than for monolingual Persians.

The predictions derived here from the perceptual assimilation results can be provisionally checked against literature data on the acoustic and perceptual discrimination of the British English (SSBE) vowels by monolingual Azerbaijani and Persian EFL learners. The

data show that both groups of EFL learners experience the same problems in their production and perceptual discrimination of the English monophthongs. Azerbaijani EFL learners showed poor perceptual discrimination in four SSBE vowel pair: /ʌ~ɒ/, /ɑ:~ʌ/, /u:~ʊ/, /ɑ:~ɒ/ (.45 < A' <.58) but not for the other seven pairs tested: /ɔ:~ɒ/, /æ~ʌ/, /i:~ɪ/, e~æ/, /ɔ:~ʊ/, /i:~e/, /ɑ:~ɔ:/ (.75 < A' < .87). No comparison is available with SSBE native listeners. The differences in perceptual discrimination were echoed in the formant structure and/or duration of the vowel production by the same learners (Ghaffarvand Mokari & Werner, 2017). No data on perceptual discrimination of SSBE vowel pairs by monolingual Persian EFL learners are available at this time. However, insufficient acoustic contrast was observed for the pairs /ʌ~ɒ/, /i:~ɪ/, /u:~ʊ/ but not for /e~æ/ (Pillai & Delavari, 2012). This suggest that the three extra central vowels in the inventory of Azerbaijani do not provide an advantage for Azerbaijani over Persian EFL learners, which parallels the predictions derived from our perceptual assimilation results for American English vowels.

More specific tests of the predictions made on the basis of our PAM test will be carried out in the next stage of our project, by mapping out the perceptual representation of the American English vowels by our learners and comparing it to that of American native listeners (see Van Heuven et al., 2020 for preliminary results, see also Chapter 5), and by studying the acoustic characteristics of the AE vowels produced by our learners and by American native speakers (see Chapter 6). Testing the predictions derived from the present PAM study will also allow us to check whether the categorization of the AE vowel contrasts based on simple decisions rules should be refined by including the differences in Fit-indexes of the members of contrasts, on the basis of which some of the SC contrasts may have to be regarded as CG contrasts – with different degrees of fit between the members.

Chapter 5

Mapping perceptual vowel spaces in native and foreign language

5.1. Introduction[13]

In studies on the phonetics of vowel systems, the usual procedure is to record a number of speakers of the language variety of interest and then measure the lowest two to four resonances in the speaker's vocal tract as an indication of how each vowel is pronounced in terms of tongue height and backness. Here, the center frequency of the lowest resonance, called first formant or F1, corresponds to the openness of the vowel (openness is inversely related to vowel height), while the second lowest resonance (F2) is an indication of the vowel backness (Ladefoged & Johnson, 2010; Ladefoged & Disner, 2012). More precisely, the F2 reflects the length of the oral tract, which is determined not only by the constriction point (where the body of the tongue most closely approximates the palate or backwall of the throat) but also by lip protrusion (or rounding). Measuring the F1 and F2, and plotting the coordinates in a two-dimensional map, then gives a good impression of the general organization of the vowel system. Since individual speakers have different shapes and sizes of their vocal tracts and of the cavities therein, there is considerable variability in the exact location of the vowels on such maps. In practice, the mean (also called centroid) of the dispersion cloud of each vowel is taken as the most representative or typical realization of the particular vowel type. Vowel duration is often added as a third parameter to define the vowel space of the language (variety).

Such representations of the vowel system in a universal vowel space are an important tool in the teaching of the pronunciation of a language to non-native learners. By comparing the system of the target language with the learner's native language, differences and similarities in the organization of the respective vowel systems can be illustrated, potential learning problems can be identified, and specific instructions can be formulated to explain to

[13] This chapter is an extended version of Van Heuven, V. J., Afshar, N. & Disner, S. F. (2020). Mapping perceptual vowel spaces in native and foreign language: Persian learners of English compared with American native speakers. In: Sz. Bátyi & Zs. Lengyel (Eds.), *Kétnyelvüség: Magyar és nem Magyar kontextus. Tanulmányok Navracsics Judit köszöntésére/Bilingualism: Hungarian and non-Hungarian context. Studies in honor of Judit Navracsics*. Veszprém: Pannon Egyetem, 113–130.
https://www.researchgate.net/publication/347252430

the learner how s/he should modify the native vowel category so as to articulate a more authentic vowel in the target language.

It is insufficiently realized in the teaching of the phonetics of foreign languages that studying the acoustics of the vowel systems per se does not reveal the full organization of a vowel system, and – more importantly – does not reveal the (often incorrect) perceptual representation of the vowel system of the target language. What is needed to appreciate the representation of the vowel system in the mind of the learner (and of the native speaker) is a perceptual mapping. Using perceptual techniques allows the re- searcher to establish so-called trading relationships between the parameters that define the individual listener's vowel space. As a case in point, Van Heuven (1986) studied the mental representation of the vowel system of Dutch with native Dutch listeners and with Turkish immigrants who had lived in the Netherlands for eight years or longer. Dutch contrasts tense and lax vowels in pairs, the members of which are rather close to one another in the spectral space but differ in duration by a 2-to-1 ratio. In one vowel pair, /a/ is articulated as a long front vowel, which is contrasted with a short back vowel /ɑ/.

In the present study we used a universal reference set of synthesized vowel tokens in a C_C context that can be used to map out the (oral) monophthongs of any language (see Van Heuven, Afshar & Disner, 2020 for details and background). The vowel space is defined by three parameters, i.e., vowel height (F1), vowel backness/rounding (F2) and length (vowel duration). In the reference set we devised, the vowels occur between an initial consonant /m/ and a final consonant /f/. These are labial consonants, which are easy to synthesize and which are pronounced the same in any language. Using these materials, a vowel identification experiment was conducted to answer the following question:

> What is the mental conception monolingual Persian and bilingual Azerbaijani/Persian learners of English have of the American English vowel system in terms of the vowel quality (color) and vowel duration compared with native speakers of American English?

5.2. Methods

5.2.1. Participants

The participants were monolingual Persian EFL leaners (N = 21), and bilingual Azerbaijani/Persian EFL leaners (N = 27), who were adolescents and had taken about 6 years of English lessons. They were high-school students who had not specialized in English and had not spent time in an English-speaking environment, i.e., the type of speaker that is the

typical ELF user in international settings. These participants were the same individuals who were described in Chapters 3 and 4, in which the data were collected on their language background and language dominance (LEAP-Q) and their perceptual assimilation of the vowels of American English was determined.

The same vowel identification task, using the same materials and procedure, was performed by a groups of American native control listeners (N = 20), all of whom spoke a form of General American English. These native speakers hailed from many different states in the USA, although half of them were born and bred in California.[14] Background information on these control listeners is provided in Appendix 5.1.

5.2.2. Materials

The stimuli for the vowel identification experiment were sampled from a two-dimensional spectral grid defined by the center frequencies of the two lowest resonances in the human vocal tract, the formants F1 (representing the vowel height dimension) and F2 (as a correlate of backness and lip rounding). F1 values ran from 2.5 to 8.5 Bark, in 7 steps of 1 Bark. The F2 values were varied in 9 steps of 1 Bark between 6 and 14 Bark. This yields 9 × 7 = 63 different vowel spectra in a rectangular matrix, as is illustrated in Figure 5.1. The use of equal steps (of 1 Bark) ensures that each step yields a change of vowel quality that is perceptually the same (Traunmüller 1991). A number of F1-by-F2 combinations cannot be produced by the human vocal organs, and when synthesized, sound inhuman. These impossible/inhuman F1-by-F2 combinations were eliminated from the set of 63, so that a subset of 43 legitimate combinations remained, which together constitute a uniformly sampled universal acoustic vowel triangle.[15] For details of the stimulus synthesis and motivation of choices made, I refer to Van Heuven et al. (2020).

The vowels were synthesized, using the LPC vocoder implemented in Praat, in an isolated a C_iVC_f monosyllable, where the initial C was [m] and the final C [f]. The formant transitions were copied from natural human tokens of the corner vowels in /mif/, /maf/ and /muf/, and adjusted by linear interpolation to reach the steady-state formant values defined in Figure 5.1. Each token was synthesized once with a vowel duration of 200 ms and a second

[14] The data collection for the American controls was done by Prof. dr. Sandra F. Disner at the University of Southern California in Los Angeles. A preliminary report of the data, based on the results of 16 American listeners, was given in Van Heuven et al. (2020). The present chapter includes data for four more listeners.

[15] In the synthesis system we used, the number of formants was five, within a frequency band from 0 to 5 KHz. Only the formants F1 and F2 were varied independently. We did not independently vary any of the higher formants. The center frequency of F3 was set at 650 Hz above the F2, with a lower limit of 2450 Hz. The F4 and F5 frequencies were kept constant throughout at 3500 and 4500 Hz, respectively.

time of 300 ms, yielding a total set of 86 stimulus types. Oscillograms, and spectrograms with formant tracks overlaid of selected sample stimuli can be found in Appendix 5.2.

F2 (across) F1(down)			F2 (step, Bk, Hz)								
			1.	2.	3.	4.	5.	6.	7.	8.	9.
			14.0	13.0	12.0	11.0	10.0	9.0	8.0	7.0	6.0
Step	Bark	Hertz	2357	2031	1746	1497	1278	1086	915	764	628
1.	2.5	237									
2.	3.5	339									
3.	4.5	447									
4.	5.5	565									
5.	6.5	694									
6.	7.5	838									
7.	8.5	998									

Figure 5.1. Steady-state F1 and F2 values for reference vowels. F1 is varied in 7 steps of 1 Bark (with equivalent hertz values shown) while F2 is varied in 9 steps. Twenty impossible/non-human F1-F2 combinations (grey cells) are excluded, leaving a vowel triangle of 43 perceptually equidistant points.

5.2.3. Procedure

The participants listened to the 86 synthesized /mVf/ vowel stimuli over good-quality headphones in individual sessions. The participant saw eleven response buttons on a computer screen arranged as shown in Figure 5.2.

The eleven response vowels were exemplified by a single keyword each. The keywords were supposed to be well-known to the students. In an immediately preceding experiment, the non-native participants had been exposed to two tokens of the eleven vowels in /hVd/ context spoken by two male native speakers of American English (four tokens per vowel, see chapter 2 for details). The /hVd/ words could not be used as response alternatives in the present experiment because many of the words/phrases (or their spelling-to-sound correspondence) would be unfamiliar to the Iranian learners (e.g., *heed*, *hayed*, *hud*, *hod*, *hawed*, *hoed*, *who'd*). The keywords that were used instead end in (dark) /l/ but the onset consonant could vary.[16] The participants were told to imagine what the vowel in each keyword sounds like, and then to decide which of the eleven response vowels came closest to the vowel sound in the /mVf/ token they had just heard. They were instructed to indicate the vowel of their choice by clicking on the corresponding response button, which would then be greyed out, while at the same the row of goodness buttons on the bottom row of the screen was activated (turned red) to invite the

[16] It seems impossible to find a minimal set of well-known words containing the 11 monophthongs of American English.

participant to judge whether the vowel sound s/he had just heard was a good, average or poor token of the vowel category selected. Response time until the typicality judgment was made, was measured with a precision of 1 ms. Then the screen would be restored to the initial setting, and after 1000 ms the next stimulus was made audible. Stimulus presentation and data collection were controlled by a computer script written for the ExperimentMFC module in the Praat software (Boersma & Weenink, 2019).

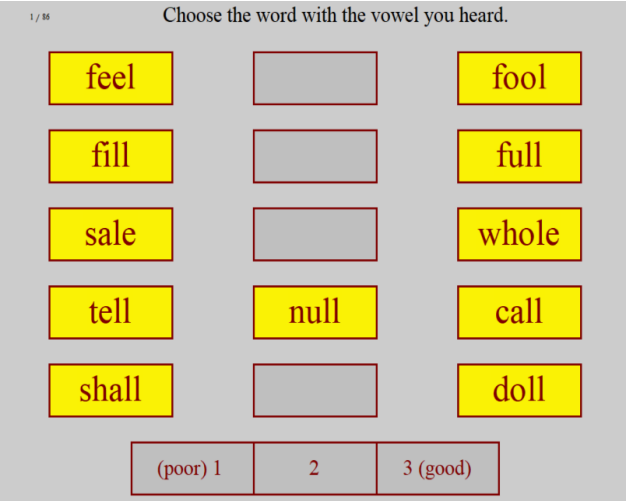


Figure 5.2. User interface for vowel identification experiment. The eleven pure vowels of American English are represented by keywords. The location of the buttons mirrors the position of the vowels in a traditional vowel diagram. At the bottom of the screen three response buttons are available for typicality (goodness) judgments; these became active only after the participant identified the vowel.

5.2.4. Data analysis

The factors in this experiment are first of all the three stimulus variables, i.e., the frequency (step) of the F1, F2 and the duration of the vowel sound. These are continuous variables at the ratio scale level of measurement. F1 and F2 were sampled with 7 and 9 (perceptually equidistant) steps, respectively. The spectral space was sampled symmetrically but not orthogonally, since 20 combinations of F1 and F2 were not used. Full orthogonality exists between the spectral types and the temporal factor: every spectral sampling point occurs with short and long duration. Further independent variables are the type of listener, i.e., native controls versus nonnative learner, with a division of the latter group in monolingual and early bilingual EFL learners. Individual listeners are nested under gender (male vs female). As before, there were three dependent variables: the choice of vowel (forced choice from 11 response categories), the goodness judgement (three degrees), and the response time. Goodness and response time were recorded but will not be analyzed in this dissertation.

I will analyze the results in three steps. The first stage of the data analysis computes descriptive statistics for each of the 11 vowel types, as a way to define the perceptual representation of the vowel space in the minds of the three listener groups. The perceptual representation will be expressed in terms of the centroids of each vowel category in the F1-by-F2 space, i.e., the intersection of the mean F1 and mean F2 value computed for a vowel category. The centroids represent the preferred ('prototypical') location of each vowel in the spectral space. Since we also want to know where the boundaries are between adjacent vowel categories, I will compute the two-dimensional dispersion as ellipses in the spectral space. The ellipses are drawn at ± 1 standard deviation from the centroid along the first two principal components of the scatter cloud of measurement points around the centroid. These ellipses theoretically include 46% of the central-most datapoints around the centroid. The more the ellipses of two vowel types overlap, the poorer the separation of the vowel types in the mind of the listener.

In the second step, I will simply count the distribution of the responses over the 11 response categories for each of the 86 synthesized vowel stimuli. The crucial statistic here is the mode, i.e., the most frequent vowel response per stimulus. The modal response will be mapped for the three listener types separately for the 43 short and 43 long stimulus types. Differences due to listener type and stimulus duration will be counted, but not analyzed by inferential tests.

In the last stage of the data analysis I will examine to what extent the perceptual representation of the American English vowels on the part of the EFL learners resembles that of the native control listeners. I will assume that the modal response given by the native listeners to each of the 86 stimulus vowels is the correct identification. Then any deviant response from the modal 'norm' can be considered an error. The results will be presented in terms of confusion matrices, with the 'norm' (L1 modal) vowel category in the rows and the actually responded vowels in the columns. 'Correct' (i.e., identical to the modal L1 response) identifications appear in the cells along the main diagonal of the matrix. Finally, confusion graphs will be produced in which the more frequent incorrect identifications ('confusions') are represented as arrows pointing away from the intended vowel to the error vowel. The differences in number of confusions between native AE listeners and the two groups of EFL learners will be evaluated by chi-square testing. I will report the value of the contingency coefficient phi (φ) as a measure of the strength of the association, and determine the significance of the association by chi-square (χ^2). In order to obtain a clearer view of the effect of stimulus duration, chi-square tests will also be done on the differences in number of

deviations from the L1 norm for short and long synthesized stimulus vowels separately, for which I will again report φ and χ^2.

5.3. Results

5.3.1. Perceptual representation: centroids and dispersion ellipses

Figure 5.3 presents the centroids and dispersion ellipses in the F1 by F2 plane (both axes in Barks) for the responses of the 21 Persian learners of English. All responses were included, and no weighting for typicality was applied. The ellipses were drawn at ± 1 standard deviation away from the centroid along the first two principal components of the scatter clouds around each centroid (thereby including the most typical 46% of the vowel qualities associated with the category).[17] Panel A shows the results obtained for the short vowel tokens, panel B does the same for the long vowel stimuli.

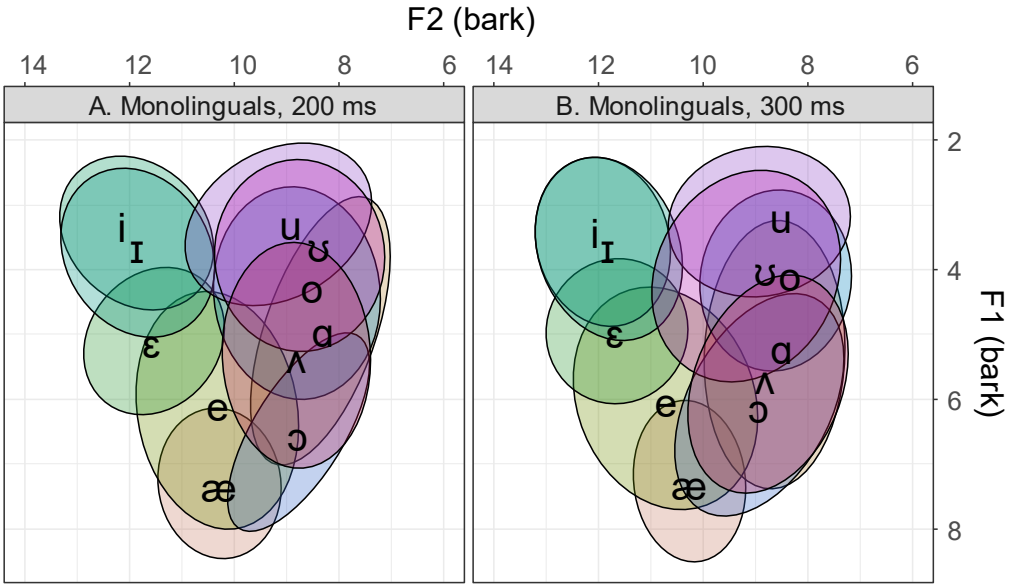


Figure 5.3. Centroids and dispersion ellipses (± 1 SD) in an F1-by-F2 plane (axes in Barks) for short (panel A) and long (panel B) stimulus vowels, as perceptually labelled by 21 monolingual Persian learners of English.

Figure 5.3 shows that the qualities of the synthesized stimuli were interpretable by the monolingual Persian learners of English. Stimuli with /i/-, /u/- and /æ/-like quality vowels end up in the left-top, right-top and mid-bottom part of F1-by-F2-plane. We also observe that the learners make a clear distinction, with a large distance between the centroids and no overlap between the spreading ellipses, between English /ɛ/ and /æ/, which is what one would expect

[17] The plots were produced with Visible Vowels (Heeringa & Van de Velde, 2018) – with some post-editing.

given the presence of near-equivalents in the learners' L1 – whereas this particular contrast is a well-known problem for many other EFL speakers. What immediately strikes the eye is the absence of a contrast between the tense and lax (mid-) high front vowels /i/~/ɪ/ and between the high back vowels /u/~/ʊ/. The centroids are virtually in the same locations and the spreading ellipses show almost complete overlap. We also see very little difference in the configuration of the vowel responses between the two panels – no effects of the difference in vowel duration are apparent.

Figure 5.4 presents the responses of the 27 early bilingual Azerbaijani/Persian EFL learners. The organization of the figure is the same as for Figure 5.3.

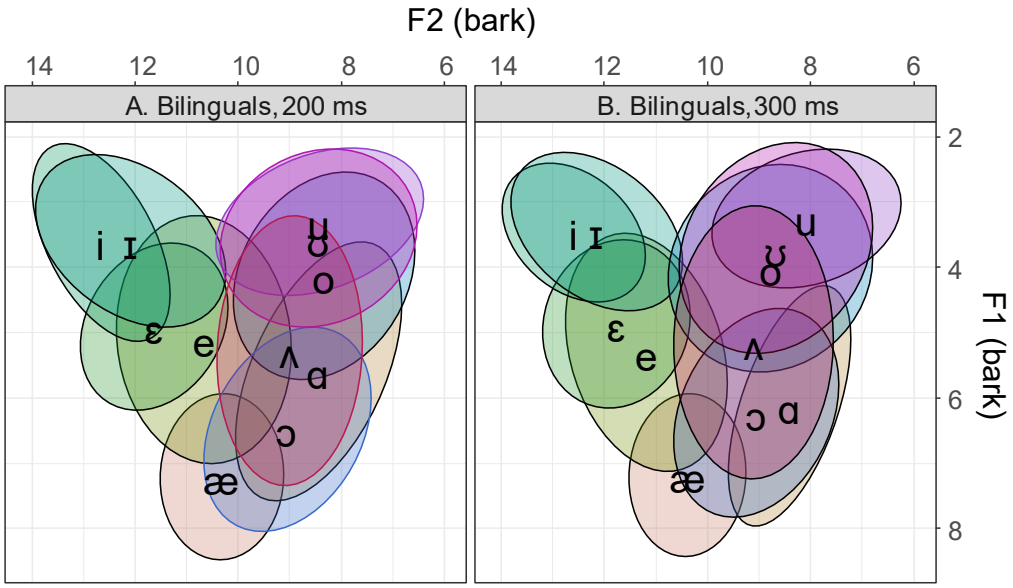


Figure 5.4. Centroids and dispersion ellipses (± 1 SD) in an F1-by-F2 plane (axes in Barks) for short (panel A) and long (panel B) stimulus vowels, as perceptually labelled by 27 bilingual Azerbaijani/Persian learners of English.

The observations that were made for the monolingual EFL learners equally apply to the early bilingual learners. There is almost complete overlap between the location of the centroids and the dispersion ellipses for the short/lax and long/tense counterparts in the pairs /ɪ~i/ and /ʊ~u/ but a clear distinction between mid and low front vowels /ɛ~æ/ is maintained. The (mid)- high back vowels are poorly separated, as are the (mid-)low back vowels. The centralized mid-low back vowel /ʌ/, however, seems to be better separated from the other (mid-)low back vowels (/ɔ, ɑ/) than was seen in the monolingual results.

The results for the American native listeners are shown in Figure 5.5. The most striking difference between the foreign and the native representation of the American vowel system is

that the American listeners maintain a clear difference between the long/tense versus short/lax counterparts, /i/~/ɪ/ and /u/~/ʊ/. There is a large distance between the centroids associated with the members of these pairs, and there is no, or only relatively little, overlap between the associated spreading ellipses. Notice that the centroids of the lax vowels, especially those of /ɪ/, /ʊ/ and /ʌ/ are rather centralized when computed for the stimulus vowels with long duration, but assume more peripheral locations (closer to the tense counterparts) when heard with short duration. This would be a first indication that vowel quality and duration are in a trading relationship in the native speakers' mental representation of the tense-lax vowels. For a phonetically long vowel to be perceived as a lax member of a contrast, it has to be very clearly centralized. When a less centralized vowel is short (enough), it will still be perceived as lax. It is also apparent that the native listeners do not maintain distinct vowel categories for pair of tense vowels /ɑ/~/ɔ/ and for the lax vowels /ʊ/~/ʌ/.[18]

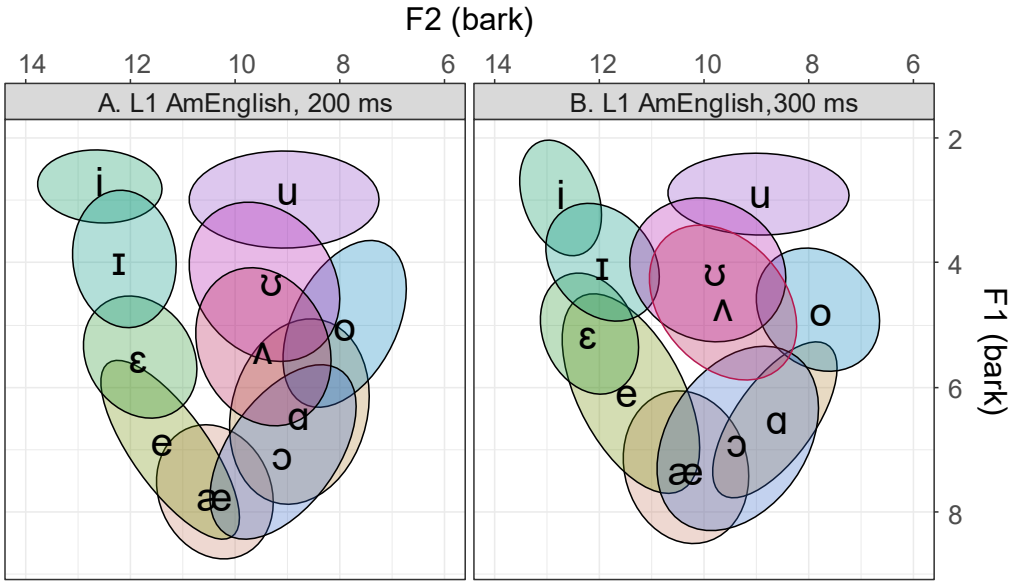


Figure 5.5. Centroids and dispersion ellipses (± 1 SD) in an F1-by-F2 plane (axes in Barks) for short (panel A) and long (panel B) stimulus vowels, as perceptually labelled by 20 American native listeners.

Figure 5.6 plots the mean duration that could be computed for the synthesized vowel stimuli that were identified with each of the 11 response categories, with separate lines for the three groups of participants. The vowel types are ordered in ascending duration based on the results of the American native listener group. It can be seen that the two groups of EFL learners associate approximately the same durations with the vowel types, while these differ

[18] Most Californian speakers, and speakers in the western half of the USA, do not distinguish between /ɔ/ and /ɑ/ (the vowel sounds in *caught* versus *cot*), characteristic of the *cot–caught* merger (Labov et al., 2006; see also Ladefoged & Johnson (2011: 212-213). Also, /ʊ/ is moving towards [ʌ] (so that, for example, *book* and *could* in the California dialect start to sound, to a GA speaker, more like *buck* and *cud*), /ʌ/ is moving beyond [ɜ].

substantially from the durations selected by the native listeners. The correlation between the vowel durations assigned to the 11 vowel types by the two learner groups is r = .904 (p < .001). The correlation in assigned vowel durations is much poorer but just significant for the bilingual learners and the American native listener, r = .542 (p = .043, one-tailed), while the correlation between the monolinguals and the controls fails to reach significance, r = .308 (p = .178, ins.).

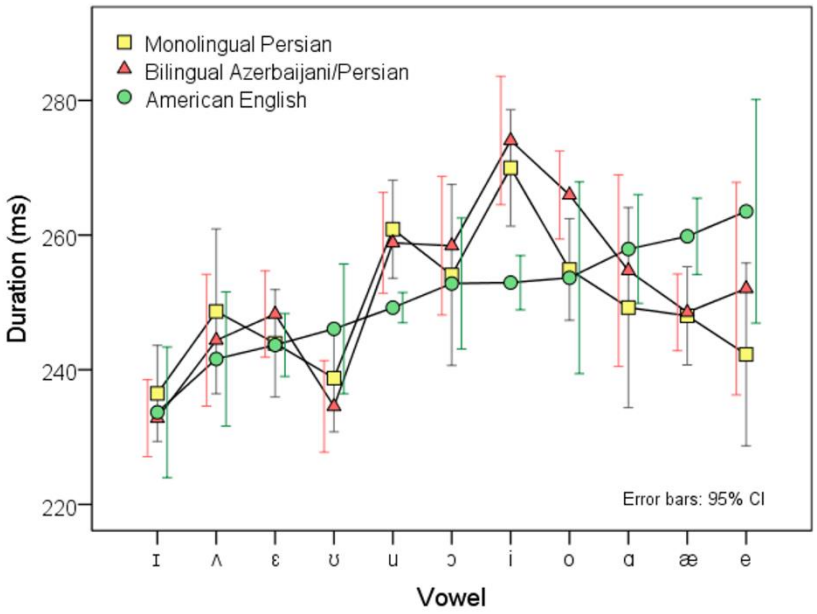


Figure 5.6. Mean duration (ms) of 11 American English vowel types identified in synthesized vowel stimuli, with separate lines for the three participant groups (N = 21 for monolingual EFL learners, 27 for bilingual EFL learners and 20 for native control listeners). Error bars represent the 95% confidence of the mean.

All three groups associate short durations with the lax vowels. Divergence between the EFL learners and the native controls is seen prominently for a number of tense vowels. Typically, the duration of /i/ and /u/ is overestimated in the perceptual representation of the EFL learners, possibly because they are intent on signaling a difference between lax /ɪ, ʊ/ (which indeed have the shortest durations selected) and tense vowels /i, u/, so that the duration contrast is exaggerated, to compensate for a lack of contrast in vowel quality. The EFL learners underestimate the duration that is appropriate for the phonetically tense vowels /æ/ and /e/.

5.3.2. Dividing up the vowel space

In this section we will consider how the vowel space is divided up by the native and non-native listeners. We will do this separately for the short vowel set and for the long vowel set. Each set comprises 43 vowel points that only differ in their quality (vowel color). The next two figures show, for each sampling point in the vowel space, which of the 11 response categories the stimulus was assigned to. If 50% (or more, i.e., the absolute majority) of the

responses converged on a particular category, a large-print bolded phonetic symbol is entered in the figure. Agreement between 25 and 50% of the responses is indicated by smaller print (unbolded). When less than 25% agreement was obtained for a particular vowel category, the cell was left blank (this occurred rarely). Figure 5.7A shows the modal mapping of the English vowel space as entertained by native American listeners responding to the short vowel stimuli.

A. Short stimulus vowels

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1.	**i**	**i**	u	**u**	**u**	**u**	**u**	**u**	**u**
2.	**i**	**i**	ɪ	ɪ	**u**	**u**	**u**	**u**	**u**
3.		ɪ	ɪ	ɪ	ʊ	**ʊ**	o	**o**	
4.		**ɛ**	**ɛ**	**ɛ**	ʌ	**ʌ**	ʌ	ɑ	
5.			**ɛ**	ɛ	ʌ	ʌ	**ɔ**		
6.				**æ**	**æ**	ɔ			
7.				**æ**	**æ**	ɔ			

B. Long stimulus vowels

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1.	**i**	**i**	u	**u**	**u**	**u**	**u**	**u**	**u**
2.	**i**	**i**	ɪ	ɪ	**u**	**u**	**u**	**u**	**u**
3.		**ɛ**	**ɛ**	ɪ	ʊ	ʊ	o	**o**	
4.		**ɛ**	**ɛ**	ɛ	ʌ	ʌ	ɑ	ɑ	
5.			æ	**æ**	æ	**ɔ**	**ɑ**		
6.				**æ**	**æ**	**æ**			
7.				**æ**	æ	æ			

X: >50% agreement; x: 25-50% agreement

Figure 5.7. Modal responses by 20 American native listeners for 43 vowel stimuli differing in F1 (vertically) and in F2 (horizontally) center frequencies. Vowel duration is either 200 ms (panel A) or 300 ms (panel B). For specifications of F1 and F2 steps see Figure 5.1. Large bolded symbols denote a majority decision with 50% or more agreement. Small symbols indicate a modal response with agreement between 25 and 50%. Cells with a modal response < 25% are left blank.

The native listeners have a rather straightforward division of their vowel space. The top left of the space is taken up by tense /i/, while the top center and top right area is almost exclusively occupied by tense /u/. Then, going down along the front edge of the diagram, there is a well-demarcated area for /ɛ/ and a smaller area for maximally open /æ/. The central portion of the vowel space is taken up by the other lax vowels /ɪ, ʊ, ʌ/. As can be expected from the results shown in Figure 5.5, the open and half open back vowels /ɔ, ɑ/ are not delineated from each other, and, if anything, they seem reversed in the listeners' perceptual representation. Note that all tense vowel categories – with the exception of the semi-diphthong /e/ and merging /ɑ, ɔ/ – are the modal response for at least two sample points in spite of their short duration. This is an indication that duration is not likely to be the primary cue in the tense-lax contrast in (American) English.

Figure 5.7B shows the results in terms of preferred responses of the American listeners for the 43 vowel types with long duration (300 ms). It would appear that category boundaries between pairs of vowels that differ in height are very regular, and are almost exclusively based

on F1 frequency. The tense vowel categories have expanded their size somewhat (from 26 to 31 sample points) since the vowel duration of 300 ms fits the specification of this type of category. By the same token, the area taken up by lax vowels has shrunk (from 17 to 12 sample points). The effect of longer duration is surprisingly large for /æ/, which vowel expands its area from 4 sample points to 9. This would be yet another indication that /æ/ in American English is phonetically tense and long. The semi-diphthong /e/ is never perceived with at least 25% agreement, not even when the duration matches its internal (long) specification. This would indicate that the closing gesture (diphthongal trajectory defined by gradual lowering of F1 frequency) is indispensable for /e/ − but not necessarily so for /o/.

The vowel labeling by the monolingual Persian learners of English is shown in Figure 5.8A for short stimuli and 5.8B for long stimuli.

A. Short stimulus vowels

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1.	**ɪ**	**ɪ**	ɪ	ɪ	ʊ	u	u	ʊ	**ʊ**
2.	**ɪ**	**ɪ**	**ɪ**	ɪ	**ʊ**	ʊ	ʊ	**ʊ**	**ʊ**
3.		**ɪ**	ɛ	ɛ	ɛ		o	o	
4.		ɛ	**ɛ**	ɛ	æ			ʌ	
5.			æ	æ	æ		ʌ		
6.				**æ**	æ	æ			
7.				**æ**	**æ**	**æ**			

B. Long stimulus vowels

1.	2.	3.	4.	5.	6.	7.	8.	9.
i	**i**	ɪ	u		ʊ	u	**u**	u
i	ɪ	**i**		u	u	ʊ	o	u
	ɛ	ɛ	ɛ		o	o	o	
	ɛ	ɛ	ɛ		æ	ɔ	ɔ	
		æ	**æ**	**æ**	ʌ	ʌ		
			æ	**æ**	æ			
			æ	**æ**	**æ**			

X: >50% agreement; x: 25-50% agreement

Figure 5.8. Majority responses by 21 Persian learners of English for 43 vowel stimuli. See Figure 5.7 for details.

It would appear, from Figure 5.8A, that the identification of English lax high vowels by the Persian L2 listeners is largely determined by their short duration. Nine sample points are taken up by /ɪ/ (against only 5 in the case of the native listeners). Similarly, lax /ʊ/ is identified in 8 sample points (against a mere 2 for the native listeners). Moreover, /ɪ/ and /ʊ/ are preferred responses even when the vowels are articulated at high rather than mid-high values (i.e., at an F1 value of 2.5 Bark) – which is never the case with American listeners. The area devoted to open /æ/ is far too large (10 sample points while the native listeners identify /æ/ in only 4 sample points), and partly occupies space where native listeners perceive /ɛ/. The central vowel /ʌ/ is too far back and infringes on the area where the native listeners perceive /ɔ, ɑ/. Remarkably, the learners identify two sample points as tense (diphthongal) /o/, even when the vowel duration is short. The conclusion must be that the mental representation of the English vowels as entertained by Persian learners of English is distorted and incorrect. Now

turning to the modal vowel responses by the Persian learners for the long vowel stimuli (Figure 5.8B), we see an unsystematic scattering of tense and lax vowel responses, with /i, ɪ/ for front vowels and /u, ʊ/ for back vowels. This would seem to indicate that the Persian learners rely on duration as the primary cue distinguishing the lax and tense members of the (mid-)high vowel pairs. At short durations the dominant percept is the lax vowel (9 vs 0 sample points for /ɪ/ vs /i/, 8 vs 2 sample points for /ʊ/ vs /u/). When the duration is long, there is a preponderance of tense vowel responses, with 2 vs 4 sample points for /ɪ/ vs /i/, and 2 vs 7 sample points for /ʊ/ vs /u/. Again, /ʌ/ is identified as a back vowel. The /ɔ, ɑ/ category is the modal response for 2 sample points, which suggests that these vowels are conceived of as long/tense vowels by the Persian EFL students. All low vowel samples are identified as tokens of /æ/.

In the last part of this section, I will present the vowel categorization data collected for the early bilingual Azerbaijani/Persian learner of English as a foreign language. Figure 5.9 shows the modal vowel response for each of the 43 synthesized vowels with short duration (panel A) and for the long duration (panel B).

The general pattern of vowel responses given by the early bilinguals is the same as that of the monolinguals. The modal responses given to the short stimuli in the four top rows (i.e., with F1-values suggesting high and mid vowels), are nearly always lax vowels /ɪ, ɛ, ʌ, ʊ/, while the three bottom rows (suggestive of open vowels) are almost exclusively assigned to /æ/.

A. Short stimulus vowels

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1.	**ɪ**	**ɪ**	**ɪ**	ɪ	ʊ	ʊ	ʊ	u	**ʊ**
2.	**ɪ**	**ɪ**	**ɪ**	ɪ	ʊ	ʊ	**ʊ**	**ʊ**	ʊ
3.		**ɪ**	ɛ	ɪ	ɪ	ʊ	ʊ	o	
4.		**ɛ**	**ɛ**	ɛ	ɪ	ʌ	ʌ	ʌ	
5.			æ	**æ**	æ	ʌ	ʌ		
6.				**æ**	**æ**	æ			
7.				**æ**	**æ**	æ			

B. Long stimulus vowels

1.	2.	3.	4.	5.	6.	7.	8.	9.
i	i	ɪ		ʊ	ʊ	u	**u**	**u**
i	ɪ	**i**	ɪ		u	u	u	u
	i	ɛ	**i**	ɪ	o	o	o	
	ɛ	**ɛ**	ɛ		ɔ	ʌ	ɔ	
		æ	**æ**	**æ**	ʌ	ɔ		
			æ	**æ**	æ			
			æ	**æ**	æ			

X: >50% agreement; x: 25-50% agreement

Figure 5.9. Majority responses by 27 early bilingual Azerbaijani/Persian learners of English for 43 vowel stimuli. See Figure 5.7 for details.

In order to appreciate the similarities and differences in the perceptual representation of our three groups of participants more easily, Table 5.1 summarizes the number of sample points that are assigned to each of the 10 response categories (with /ɔ, ɑ/ merged) broken

down by participant group. The bottom half of the table shows the absolute difference in number of sample points that were assigned to a particular vowel with a modal choice ≥ 25%.

Table 5.1. Modal vowel response category (with /ɔ, ɑ/ merged) broken down by duration of synthesized vowel for three groups of participants (L1 native listeners, monolingual Persian EFL learners, early bilingual Azerbaijani/Persian EFL learners). In cells marked '??' no modal response ≥25% could be found. |Δ| is the absolute difference between a pair of participant groups in number of stimulus types assigned to a modal response category.

L1	Duration	Modal response vowel											
		i	ɪ	e	ɛ	æ	ɑ/ɔ	ʌ	o	ʊ	u	??	Total
Am. English native	100 ms	4	5	-	5	4	4	5	2	2	12	-	43
	200 ms	4	3	-	5	9	4	2	2	2	12	-	43
Monolingual EFL learner	100 ms	-	9	-	6	10	-	2	2	8	2	4	43
	200 ms	4	2	-	-	9	2	2	4	2	7	4	43
Bilingual EFL learner	100 ms	-	12	-	4	9	-	5	1	11	1	-	43
	200 ms	6	4	-	4	9	3	2	3	2	7	3	43
|Δ| native-mono	100 ms	4	4	-	1	5	4	3	-	6	10	4	41
	200 ms	-	1	-	5	-	2	-	2	-	5	4	19
|Δ| native-biling	100 ms	4	7	-	1	5	4	-	1	9	11	-	42
	200 ms	1	1	-	5	-	3	-	1	-	5	2	18
|Δ| Mono-biling	100 ms	-	3	-	2	1	-	3	1	3	1	4	18
	200 ms	2	2	-	4	-	1	-	1	-	-	1	11

This table shows that the number of stimulus types that are assigned to each of the 10 categories with a modal response of at least 25% (so excluding the column marked '??') tends to be the same for the two EFL learner groups. The bottom rows specify the absolute number of stimulus types that differ in their modal response between the two learner groups. The greatest discrepancies between the native respondents and the EFL learners are in the vowel pairs /ɪ, i/ and ʊ, u/. For the native participants, the division of responses over the lax and tense members of these pairs is virtually the same for short and long stimuli. This demonstrates that duration is not important for the identification of the lax and tense members of the pairs, and therefore that the contrast is predominantly cued by the difference in vowel quality). For the two Iranian respondent groups, the duration of the stimulus vowel makes all the difference between the members of these pairs. If there is a difference between the two EFL learner groups it would be in the fact that the early bilinguals rely (even) more on the duration cue than the monolinguals do. Also, the bilinguals have a tendency to respond the central /ʌ/ more than the monolinguals, while the monolinguals use the front counterpart /ɛ/ more often. This may have its cause in the presence of central vowels in the inventory of Azerbaijani.

5.3.3. Native and nonnative vowel identification compared in detail

Figures 5.8-9 show the majority ('modal') decision for each of the 86 synthetic vowels in the stimulus material, as they were made by the two groups of EFL learners. Quantifying the differences between native AE listeners and the nonnative students in detail can be done by using the majority decision of the native AE listeners (in Figure 5.7) as the absolute criterion for correct vowel identification. Using this majority criterion quite a few responses given by AE listeners will deviate from the majority decision and can be considered "wrong" or "sub-optimal" responses or "confusions". The same criterion can be used to analyze the nonnative responses given to the same stimuli. In that case, the number of confusions will be considerably larger, and the discrepancies between the error responses by the native and nonnative groups of listeners can be quantified by subtraction.

The first step in the analysis was to count the number of responses given to each of the eleven response categories (in the columns) for each of the 86 stimulus types (in the rows). Appendices A5.3-5 contain the results of the counting, for 20 American native listeners, for the 21 monolingual Persian and for the 27 early bilingual EFL students, respectively. On the left side of the table, the responses are given for short vowel stimuli (vowel duration = 200 ms), the right-hand side shows the responses to the long vowel type (300 ms). The rows are ordered by ascending F1 frequency, and then by descending F2 frequency. The modal response for each stimulus type was then determined and indicated in the table in bold face in green-shaded cells. In the native AE results, five stimulus vowels yielded a multimodal distribution of responses. In those cases, one of the multiple modes was given preference such that contiguity in the dispersion areas in Figures 4.2-3 was optimized.

If we now take the modal response category of the group of native listeners as the correct (or at least preferred) response for a stimulus vowel, a percentage of correct scores can be computed for the American native speakers. Table 5.2 is a confusion matrix with the modal (preferred) category in the rows and all observed response categories in the columns. The cells along the main diagonal contain the 'correct' responses for the native listeners. The off-diagonal cells contain alternative non-modal choices made, which can be considered 'wrong' or a-typical.

The proportion of 'correct' responses (in cells along the main diagonal) is 1027 / 1720 = 59.7%. The vowels /e/ and /o/ are used quite infrequently as response categories. This will be due to the circumstance that the mid-high long/tense vowels in American English should be diphthongized. Moreover, /e/ never ends up as a modal response for any of the 86 synthesized vowels, so that the row for /e/ remains empty.

Table 5.2. Confusion matrix of all observed responses against modal ('correct') response category for 20 American native listeners. Cells contain row percentages. Correct responses (agreeing with the modal response) are in bold face in green-shaded cells. Confusions ≥ 10% are in red-shaded cells. Marginals specify number of observations in row or column.

		All observed responses											Total
		i	ɪ	e	ɛ	æ	ɑ	ʌ	ɔ	o	ʊ	u	
Modal AE response category	i	**83.1**	11.3	1.9	2.5						.6	.6	160
	ɪ	6.7	**42.5**		14.2	3.3		4.2		.8	11.7	16.7	120
	e												0
	ɛ	1.3	11.3	10.0	**67.5**	1.3	0.6	2.5	1.3		4.4		160
	æ	.3	.7	9.0	12.7	**48.3**	8.7	2.0	18.0	.3			300
	ɑ					5.0	**41.3**	6.3	20.0	25.0	1.3	1.3	80
	ʌ		.7		5.0	3.6	12.9	**39.3**	5.0	7.9	25.0	.7	140
	ɔ			2.5		20.0	23.8	2.5	**46.3**	3.8	1.3		80
	o						3.8	13.8	7.5	**52.5**	20.0	2.5	80
	ʊ		10.8	3.3	7.5	.8	.8	21.7	.8	.8	**41.7**	11.7	120
	u	2.5	2.1		.8		.4	2.7	.2	1.5	12.1	**77.7**	480
Total		156	113	52	187	177	103	127	124	86	183	412	1720

Table 5.3 repeats this procedure for the monolingual Persian students' responses to the same stimuli. The total number of responses that concur with the modal AE response, i.e., the sum of the numbers in the green-shaded cells along the main diagonal, is 616, which is 34.1% correct responses. Since the semi-diphthong /e/ never reached a modal response in the perceptual results for the Americans, the /e/ row remains empty in Table 5.3.

Table 5.3. Confusion matrix of all observed responses against AE modal response category for 21 monolingual Persian learners of English listeners as a foreign language. For more information see Table 5.2.

		All observed responses											Total
		i	ɪ	e	ɛ	æ	ɑ	ʌ	ɔ	o	ʊ	u	
Modal AE response category	i	**43.5**	47.6	1.2	1.8			.6		.6	3.0	1.8	168
	ɪ	19.8	**42.9**	3.2	15.1		.8	1.6		.8	7.9	7.9	126
	e												0
	ɛ	9.5	14.3	15.5	**45.8**	4.8	1.2	1.2	.6	1.8	3.6	1.8	168
	æ	1.6	2.2	14.9	6.0	**54.6**	2.9	3.8	8.6	1.9	2.9	.6	315
	ɑ			6.0		4.8	**15.5**	25.0	28.6	10.7	7.1	2.4	84
	ʌ	2.0	4.8	10.2	7.5	21.1	7.5	**13.6**	8.8	10.9	10.2	3.4	147
	ɔ			16.7	1.2	27.4	8.3	16.7	**16.7**	6.0	7.1		84
	o	1.2	2.4	1.2	1.2		9.5	7.1	4.8	**38.1**	21.4	13.1	84
	ʊ	6.3	15.1	5.6	20.6		1.6	7.1	.8	18.3	**11.9**	12.7	126
	u	6.0	8.9	2.0	2.2	.2	4.8	2.6	1.0	14.1	29.4	**29.0**	504
Total		161	238	131	168	239	77	100	89	167	238	198	1806

Clearly, then, the monolingual Persian EFL learners' perceptual representation of the American vowel systems departs strongly from the native norm. Only for /æ/ do the learners concur with the native listener norm in more than 50%. To illustrate graphically how the monolinguals' conception of the AE vowel space differs from that of the native AE listeners,

the deviations from the AE norm in the PA matrix are captured in Figure 5.10B. These deviations can be compared with the deviations from the AE norm (defined by the modal response by native listeners) in the native AE matrix, which are illustrated in Figure 5.10A.

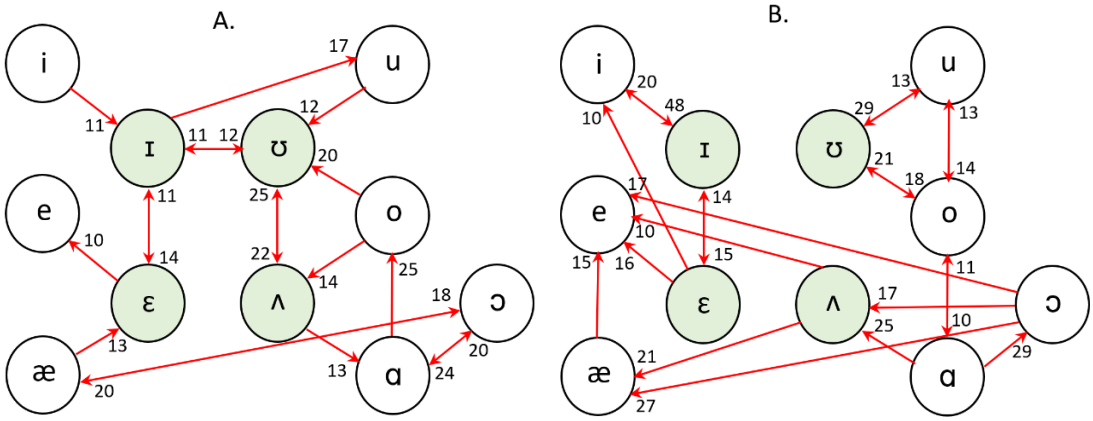


Figure 5.10. Vowel confusion structure of eleven American English monophthongs as identified for 86 synthesized vowel sounds by 20 American native listeners (panel A) and by 21 monolingual Persian EFL learners (panel B). Confusions < 10% have been omitted. Short ('lax') vowels in shaded circles. Arrows point away from the 'correct' modal vowel (according to the AE norm) to the incorrectly identified vowel. The percentage of confusion is indicated at the arrow heads.

The network in Figure 5.10 contains eleven nodes representing the monophthongs of American English. They are arranged in stylized fashion according to their position in a two-dimensional vowel space with vowel height (or F1) vertically and constriction place (backness, F2) horizontally. The seven long ('tense') vowels are on the outer perimeter, while the four short ('lax') vowels form an inner tetragon. The vowel /ɔ/ is positioned somewhat higher and to the right of low-back /ɑ/. This vowel seems to upset the symmetry in the AE vowel system, which may be one reason why /ɔ/ and /ɑ/ often merge into a single low back /ɑ/ vowel in American English. Single-headed arrows point away from the 'correct' (modal) vowel response to a non-modal ('incorrect') response category. Only confusions larger or equal than 10% are indicated. The percentage of confusion ('error percentage') is indicated at the arrow head. Double-headed arrows indicate two-way confusion between two nodes.

Since the native listeners deviate from the modal response per vowel category in 40% of the cases, there is a substantial confusion (or disagreement) even for the native listeners in Figure 5.10A. However, the disagreement is generally small, and never in excess of 25%. Disagreement is largest for the vowel pairs /ɔ, ɑ/ and /ʌ, ʊ/, both of which are often implicated in current vowel mergers in American English. Most Californian speakers do not distinguish between /ɔ/ and /ɑ/ (the vowel sounds in *caught* versus *cot*), characteristic of the *cot–caught*

merger.[19] Also, /ʊ/ is moving towards [ʌ] so that, for example, *book* and *could* in the California dialect start to sound, to a Standard American English speaker, more like *buck* and *cud*.[20]

Figure 5.10B shows that there a lot more confusion in the Persian EFL conception of the AE vowel system. There is very substantial confusion between /i/ and /ɪ/, especially from tense to lax. The same asymmetrical confusion is seen for the high-back pair /u/ ~ /ʊ/. The (mid) low back vowels /ɔ, ɑ/ are often identified as central lax /ʌ/. Moreover, the low back and central vowel qualities are often mistakenly identified as front vowels, where vowels that native listeners associate with /ɔ/ or /ʌ/ are identified as mid front /e/ (in 17 and 10% of the cases), or as low front /æ/ (27 and 21% of the cases).

The analysis is repeated for the early bilingual Azerbaijani/Persian EFL learners in Table 5.4 and the summary confusion graphs in Figure 5.11AB. For ease of comparison, Figure 5.11A, which shows the AE native listener identifications, is a copy of Figure 5.10A.

Table 5.4. Confusion matrix of all observed responses against AE modal response category for 27 early bilingual Azerbaijani/Persian learners of English listeners as a foreign language. For more information see Table 5.2.

		All observed responses											Total
		i	ɪ	e	ɛ	æ	ɑ	ʌ	ɔ	o	ʊ	u	
Modal AE response category	i	**34.7**	56.9	1.9	3.7				.5	1.4	.9		216
	ɪ	16.7	**35.2**	5.6	19.1		.6	5.6	.6	3.1	8.6	4.9	162
	e												0
	ɛ	13.9	13.0	8.3	**51.4**	5.6	.9	.5	.5	2.3	1.9	1.9	216
	æ	.7	1.7	6.4	6.7	**57.3**	4.9	6.9	11.1	2.2	1.7	.2	405
	ɑ	.9		.9		8.3	**16.7**	24.1	29.6	7.4	10.2	1.9	108
	ʌ	1.1	7.4	4.8	6.3	17.5	6.9	**21.7**	13.8	11.1	5.8	3.7	189
	ɔ			1.9	.9	24.1	13.9	28.7	**20.4**	4.6	5.6		108
	o		1.9	1.9	1.9		10.2	10.2	5.6	**36.1**	15.7	16.7	108
	ʊ	6.2	21.6	4.3	13.0	1.2	3.7	7.4	1.9	13.6	**16.0**	11.1	162
	u	2.8	10.3	1.7	3.9	.6	1.5	8.6	.9	14.2	28.4	**27.0**	648
Total		166	333	89	238	318	96	215	143	209	282	233	2322

Overall, the differences between the two groups of EFL learners is small so that, with minor exceptions, the same structure is seen in Tables 5.3 and 5.4, as well as in the Figures 5.10B and 5.1B. The proportion of correctly identified stimuli is 796 / 2322 = 34.3%. The difference in relative number of correct, i.e., native-like, identifications between the two groups of EFL learners is negligible and statistically insignificant (see Table 5.5).

[19] See also Ladefoged & Johnson (2010: 212–213) and Ladefoged & Disner (2012: 45).
[20] https://web.stanford.edu/~eckert/vowels.html.

Table 5.5 summarizes the number of incorrect vowel identification found for the native AE listeners, and for the two groups of EFL learners. The differences in number of 'confusions', i.e., deviations from the L1 modal response, between groups of listeners are evaluated by chi-square analysis.

Table 5.5. Correct (according to L1 AE modal response) vs. confused responses (count plus row percentages) in vowel identification task by three groups of listeners. Chi-square and phi are specified for full matrix and by submatrices for pairwise comparisons of groups. Significant p-values are bolded.

Group	Response				
	Correct		Confused		Σ
1. L1 AE	1027	59.7%	693	40.3%	1720
2. L2 monolingual	616	34.1%	1190	65.9%	1806
3. L2 bilingual	796	34.3%	1526	65.7%	2322
Σ	2439		3409		5848

Comparison	φ	χ^2	df	p
All	.236	324.8	2	< **.001**
1 vs 2	.257	232.0	1	< **.001**
1 vs 3	.253	258.0	1	< **.001**
2 vs 3	–.002	0.0	1	.908

The American native listeners deviate significantly less from the norm response than either group of nonnatives do. These latter do not differ from each other in number of deviations from the L1 norm response.

Although the above shows that the two groups of EFL learners produce almost the same responses to the 86 synthesized stimulus vowels, some differences can nevertheless be observed, which I will highlight in the next paragraph.

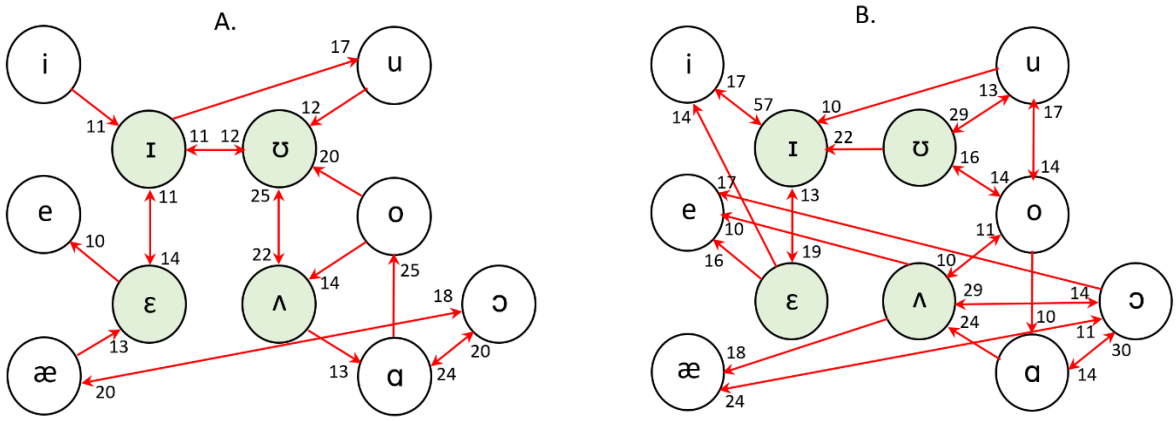


Figure 5.11. Vowel confusion structure of eleven American English monophthongs as identified for 86 synthesized vowel sounds by 20 American native listeners (panel A) and by 27 bilingual Azerbaijani/Persian EFL learners (panel B). For more information see Figure 4.10.

I subtracted the confusion percentages in Table 5.4 from the percentages in the corresponding cells in Table 5.3. Out of the 100 off-diagonal cells that contain confusions only six contain a difference between the two EFL groups greater than 5 percentage points (whether positive or negative). This bears out once more that the perceptual representations of the AE vowel systems as conceived by the two learner groups are virtually the same. The six discrepant confusions can be divided into two groups of three. The monolinguals show more frequent confusions than the bilinguals in the pairs /ɔ > e/ (15 points more), /æ > ɛ/ (9 points more), and /ʊ > ɛ/ (8 points more). These confusions all involve the mid front vowels as the deviant member. Conversely, in the second triplet, the bilinguals show more confusions than the monolinguals. These are the pairs /ɔ > ʌ/ (12 points more), /i > ɪ/ (9 points more) and /ʊ > ɪ/ (7 points more). These confusions involve pairs of which the deviant member is a central(ized) vowel. This would make sense, as Azerbaijani has central vowels in its inventory which could prompt the bilinguals to substitute these as targets for the AE lax /ɪ/ and /ʊ/. A similar, but weaker, tendency can be seen for the centralized vowel /ʌ/, which the bilinguals identify more often than their monolingual peers for vowels which native listeners perceive as exemplars of back /ɔ/ or /o/. Overall, the bilinguals identify stimuli as tokens of lax vowels in 46.0% of the responses (1068 / 2322), against 39.2% (744 / 1806). The larger percentage of lax vowel identifications by the bilinguals is almost significant, $\chi^2(1) = 3.7$ ($p = .054$, ins).

As a last illustration of the difference in weight attached to vowel duration in the mental representation of native listeners and the EFL learners, Table 5.6 specifies how often each of the 11 vowels was identified by native and nonnative listeners across all 43 vowel qualities generated in the stimulus material, separately for short vowels and long vowels (for the complete breakdown of the responses by vowel quality and duration see Appendix A5.3-5 for the native, monolingual and bilingual nonnative listeners, respectively).

Table 5.6. Number of responses in each of 11 vowel categories to short vs long vowel duration in synthesized stimuli accumulated across all 43 vowel quality differences, broken down by language background of the listener. The absolute and relative difference in number of responses is listed in de columns under Δ and %, respectively. Summary statistics are Chi-square and Phi. For more information see text.

			L1 Am. English				Monolingual Persian				Bilingual Azerbaijani/Persian			
			200 ms	300 ms	Δ	%	200 ms	300 ms	Δ	%	200 ms	300 ms	Δ	%
Response vowel	Short/ lax	ɪ	70	43	27	1.6	150	88	62	3.4	217	116	101	4.3
		ɛ	105	82	23	1.3	88	80	8	.4	122	116	6	.3
		ʌ	69	58	11	.6	52	48	4	.2	120	95	25	1.1
		ʊ	100	83	17	1.0	143	95	48	2.7	177	105	72	3.1
	Long/ tense	i	74	82	−8	−.5	54	107	−53	−2.9	48	118	−70	−3.0
		e	19	33	−14	−.8	69	62	7	.4	44	45	−1	−.0
		æ	73	104	−31	−1.8	117	122	−5	−.3	160	158	2	.1
		ɑ	44	59	−15	−.9	39	38	1	.1	45	51	−6	−.3
		ɔ	55	69	−14	−.8	37	52	−15	−.8	59	84	−25	−1.1
		o	42	44	−2	−.1	78	89	−11	−.6	75	134	−59	−2.5
		u	209	203	6	.3	76	122	−46	−2.5	94	139	−45	−1.9
	Total		860	860			903	903			1161	1161		
	Σ short/lax		344	266			433	311			636	432		
	Σ long/tense		516	594			470	592			525	729		
	$\chi^2(1)$		15.5 ($p < .001$)				34.0 ($p < .001$)				72.2 ($p < .001$)			
	φ		.095				.137				.176			

The Δ columns specify the difference in identification of the response vowel between the short and the long vowel duration. The %-columns express this difference as a percentage relative to the total number of responses given by the group of listeners, i.e., 1720 by the native listeners, 1806 by the monolingual Persian EFL learners, and 2322 by the early bilinguals. It is then easily seen that the effect of vowel duration is much larger, also relatively, for the nonnatives than for the native listeners. In the table, the rows containing lax vowels are marked in yellow. The ratio of all lax vs. all tense vowel responses by native listeners is 40:60 for short vowels, against a ratio of 31:69 for long vowels. For the monolingual Persian EFL learners these ratios are, respectively, 48:52 and 34:66; for the early bilingual Azerbaijani/Persian group the ratios are 55:45 and 37:63. For both language groups there is a significant association between tenseness of the response vowel category and the duration of the stimulus vowels (long = tense, short = lax) but the association (expressed by the Phi coefficient) is stronger for the EFL groups (φ = .137 for the monolingual Persians; φ = .176 for the bilinguals) than for the native group (φ = .095). In comparison with the American native listeners, the association of tenseness and vowel duration is significantly stronger for the bilinguals EFL learners, $\chi^2(1) = 51.8$ ($p < .001$), but not for the monolingual Persian EFL learners, $\chi^2(1) = 1.8$ ($p = .185$, ins.). This shows, once more, that duration is a stronger cue in the tense-lax contrast in English for the Iranian EFL learners but only for the early bilinguals.

5.4. Conclusions and discussion

In terms of substance, the results of the present study have shown that American native speakers have a rather straightforward perceptual representation of their monophthongal vowel system. The centroids of the 11 vowel categories and the overall topology of the vowel space closely match with what is traditionally reported for American English based on acoustic measurements of vowel production (e.g., Peterson & Barney, 1952; Hillenbrand et al., 1995; Wang & Van Heuven, 2006). The results also show that the tense-lax contrasts in the vowel pairs /i/~/ɪ/ and /u/~/ʊ/ are cued primarily by spectral properties (i.e., vowel color) rather than by a difference in duration. This latter finding strengthens a similar conclusion by Hillenbrand et al. (2000) based on the perception of natural (i.e., human) vowel tokens with manipulated duration. The results also bear out that the contrast between the (mid-)low back vowels /ɑ, ɔ/ does not exist for our American listeners, and that the contrast between the centralized back vowels /ʌ, ʊ/ is weak, especially when their duration is a-typically long.

The Persian learners of English clearly have an incorrect representation of the target vowel system. They have a flawed cue weighting in the tense-lax contrasts in the high and mid-high vowel region. In their mental representation, the tense vowels are long and the lax vowels short but there is no difference between them in terms of vowel quality, i.e., formant structure. The Persian learners also fail to distinguish between the back vowels /ɑ, ɔ/ but in a way that departs from the native listeners: the part of the vowel space that in the native representation is occupied by a merged category /ɑ, ɔ/ is filled with /ʌ/ and /æ/, which are not back vowels in the native listeners' representation. Given this incorrect mental representation of the English monophthongs on the part of the Persian learners, we predict similar errors and confusions in their articulation.

The early bilingual Azerbaijani/Persian EFL learners deviate in practically the same way from the American native controls. One important difference between the two EFL learner groups is that the bilinguals' mental representation of the American vowel system is that the bilinguals are more prone to identify stimuli as tokens of one of the central(ized) vowels of American English. Possibly, the availability of central vowels in the inventory of the bilinguals' dominant native language, i.e., Azerbaijani, makes it easier for these EFL learners to detect a quality difference between back and more central(ized) vowels.

Although both groups of EFL learners seem unaware of the quality difference between the lax and tense members of the contrastive pairs /ɪ~i/ and /ʊ~u/, they are clearly aware of the fact that the lax vowels have shorter durations than the tense counterparts. More in general, the EFL learners have a quite reasonable perceptual representation of the temporal

differences in the American English vowel system. However, they exaggerate the temporal contrast between the lax and tense members of the contrastive pairs. On the one hand, the vowel duration of /ɪ/, and especially /ʊ/ is conceived of as too short, while the selected duration of /u/, and especially /i/ is too long. This exaggeration of the temporal difference between lax and tense vowels will probably be reflected in the learners' pronunciation.

In terms of methodology, the perceptual identification of synthesized vowel tokens covering the complete vowel space for monophthongs has yielded credible effects, and has uncovered substantial differences between the nonnative learners and the native listeners of American English. As far as we are aware, this is the first time the perceptual labeling technique has been used to map out the mental conception of the complete inventory of English monophthongs by native listeners and nonnative learners. Earlier studies were inconclusive or incomplete because these targeted only a subset of the English vowels (e.g., Van Heuven, 2017 targeting only the short monophthongs) or because the vowel space was not systematically sampled (e.g., Schouten, 1975).

The artificial vowel set we generated can be used to map out the monophthongs of any language. However, the results show that American listeners are reluctant to accept the monophthongal vowel sounds as acceptable exemplars of semi-diphthongs, especially in the case of /e/ (somewhat less so for /o/). These were the least favored response categories in our experiment, and especially the behavior of /e/ was unexpected. The centroid of /e/ was at a more open location than for /ɛ/, the spreading ellipse was unusually large and stretched out along the front side of the vowel diagram. When diphthongs are to be mapped in future experiments, the stimulus material should be more complex in order to contain realistic exemplars of diphthongal categories. One way to do this is to generate diphthongal glides in the F1-F2 space but the danger of combinatorial explosion looms large.

Chapter 6

Contrastive acoustic vowel analysis

6.1. Introduction

When a foreign language is learned after the age of puberty, it is usually the case that the learner's native language interferes with the perception and production of the foreign (or: target) sounds (e.g., Escudero, 2005 and references therein). Typically, the sounds of the foreign language are perceived as exemplars of the sound categories of the learner's native language, and sounds of the learner's native language are used as substitutes in the foreign language. The pronunciation of the foreign language is therefore reminiscent of the sounds (and melodies) of the learner's native language, so that the learner's native language can be determined from subtle but systematic deviations in the learner's pronunciation from the norms that apply to the target language. In the present study we aim to study the pronunciation of English by bilingual Azerbaijani/Persian, and monolingual Persian speakers of English, and compare this with the pronunciation of American native speakers of English.

In recent years, several studies have been done on L2 speech production and perception of speakers with different L1s. When subjects with different L1s speak English as a foreign language, their pronunciation will be different from native speakers of English (e.g., Ghaffarvand Mokari et al., 2013). One of the major difficulties in the pronunciation of English lies in different realization of vowels. There are factors that cause a foreign accent which have received too little attention in the technical literature. Piske, MacKay and Flege (2001) provide a list of variables which partially determine the degree of foreign accent in an L2, i.e., gender, age of L2 learning (AOL), length of residence in an L2 speaking country (LOR), formal instruction, motivation, language learning aptitude, and amount of L2 use.

In recent years, there has been increased interest in cross-language comparisons of phonetic categories, growing out of the well-documented problems that adult second language (L2) learners have in acquiring a new phonological system (Strange et al., 2014). In his Speech Learning Model (SLM), Flege (1995) claims that continuing problems with "accented" production of phonetic segments can be attributed in large part to L2 learners' representation of the L2 segments as equivalent to "similar" segments in the native language (L1). That is, if the L2 phones are sufficiently similar to L1 phones, they will be perceptually assimilated to those

native categories, with the result that both L1 and L2 segments are produced differently from native monolingual speakers' utterances. If, however, L2 phones are sufficiently dissimilar from any L1 category (i.e., "new"), the L2 learner will (eventually) establish distinct L1 and L2 phonetic categories, and production of the L2 segments will become more native-like. In her Perceptual Assimilation Model (PAM), Best (1994, 1995) also invokes the concept of cross-language phonetic similarity to predict the relative difficulties that listeners will have in perceptual differentiation of non-native segmental contrasts. She describes several patterns of perceptual assimilation of L2 segments to L1 phonological categories, which are determined by the perceived phonetic similarity of L1 and L2 segments. Two L2 segments which are judged as equally "good" instances of a single L1 category (Single Category scenario, SC) will be most difficult to differentiate, while two L2 segments that are assimilated to two different L1 categories (Two Category scenario, TC) will be very easy to discriminate. In addition, contrasting L2 segments that differ in their judged goodness as instances of a single L1 category (Category Goodness scenario, CG) will yield intermediate levels of perceptual difficulty (see Chapter 4 for more details and examples). Finally, if an L2 segment is sufficiently dissimilar from any L1 category, it may be considered an "uncategorizable" speech sound. When paired with another L2 phone that is phonetically similar enough to be categorized as an instance of an L1 category (i.e., it is categorizable), the two phones will be relatively easily discriminated (Uncategorizable Categorizable scenario, UC). According to both these models, then, the perceived similarity of segments in L1 and L2 is an important determiner of the pattern of initial perceptual problems and persistent learning difficulties adult L2 learners have in mastering the L2 phonological system. It is critical, therefore, that cross-language perceptual similarity be established, independent of identification or discrimination performance, in order to predict L2 learning difficulties more accurately. In the work of Flege and Best, for instance, perceptual similarity has been inferred from a comparison of impressionistic descriptions of the phonetic segments (e.g., Best & Strange, 1992), transcriptions or reports from listeners about similarities between native and non-native segments (e.g., Best, Faber & Levitt, 1996) or cross-language comparisons of the acoustic structure of the non-native segments (e.g., Flege, 1987; Bohn & Flege, 1990). In more recent studies, perceptual similarity has been assessed directly, using a perceptual assimilation task in which listeners are presented non-native segments and asked to categorize them with respect to which native category they are most similar and to rate their "category goodness" as exemplars of the chosen categories (e.g., Bohn & Flege, 1990; Guion, Flege, Akahane-Yamada & Pruitt, 2000; Strange, Akahane-Yamada, Kubo, Trent, Nishi & Jenkins, 1998; Strange, Akahane-Yamada, Kubo, Trent & Nishi, 2001).

6.2. Methods

6.2.1. Participants

Two groups of listeners participated in the experiment. These were the same individuals who participated in the earlier experiments. The first group comprised 22 native speakers (11 male, 11 female) of Modern Persian. They were secondary school pupils in Tehran with a mean exposure to (American) English of roughly 6 years in a school setting. The second group consisted of 27 early bilingual listeners (11 male, 16 female) with Azerbaijani and Persian as their first two languages (see Chapters 3 and 4 for more explanation on their language background). The bilinguals were comparable to the monolinguals in all relevant aspects (age, exposure to English, level of education). They were tested in secondary schools in the city of Marand in the East Azerbaijan Province in the North West of Iran.

All listeners filled in the Language Experience and Proficiency questionnaire (LEAP-Q) developed by Marian, Blumenfeld and Kaushanskaya (2007). This questionnaire asks the participant to estimate their experience with and exposure to the languages they command, and to self-rate their proficiency and (non-)nativeness in each of these languages.

6.2.2. Procedure

After the PAM-test (Chapter 4) and the Identification test (Chapter 5), for this task individual participants were instructed to read silently a list of carrier sentences containing the 19 words/phrases that cover AE monophthongs, diphthongs, and r-colored vowels to familiarize themselves with the test material (see Appendix 6.1). Next, they were asked to read the sentences aloud from paper (without touching the paper, so that the recordings could be as noiseless as possible).

Each vowel was recorded three times; once within a common carrier word (Appendix 6.1, column A) and twice in a monosyllabic /hV(r)d/ meaningful word or phrase (Appendix 6.1, Column B) in a carrier sentence *Now say ___again*. The key words rhymed with the /hV(r)d/ words, so as to cue their correct pronunciation. Participants were instructed to take their time, read at a conversational pace, and breathe in after reading each sentence. One item was repeated at the end of the list to avoid list effects. Instructions were spoken using the participants' first language (either Azerbaijani or Persian) by the author. The recordings were made at a sampling rate of 44,100 Hz (16-bit amplitude resolution) using a PC151 Sennheiser headset with an adjustable close-talking microphone and a pop-filter mounted.

6.2.3. Statistical analysis

Vowel duration (milliseconds) and formants F1 and F2 (Barks), respectively representing vowel height and backness/rounding, were measured for the recordings made, and compared with similar data collected for 20 American L1 speakers (10 males, 10 females). Descriptive statistics (uni-dimensional means, standard deviations, bi-dimensional centroids and spreading ellipses) were analyzed by Repeated Measures Analyses of Variance (RM-ANOVA) with Vowel type as the within-speaker factor and Gender as a between-speaker factor. Differences among vowel types were subsequently analyzed by Bonferroni post hoc tests. In a second stage of the statistical analysis the vowel tokens were automatically classified by Linear Discriminant Analysis (LDA, Klecka, 1980) and Multinomial Logistic Regression Analysis (MLRA, Hosmer & Lemeshow, 1989). The classifiers were first trained and tested on the three groups of speakers separately. The models obtained for the native L1 speakers were additionally tested on the vowel tokens produced by the two groups of nonnative speakers. Confusion matrices obtained from the automatic classification were analyzed further by Hierarchical Cluster Analysis (Everitt, 1993) and converted to confusion graphs. All inferential statistical analyses were carried out with the aid of SPSS (version 22).

6.3. Results

6.3.1. Data analysis

The vowels were segmented to determine their duration (in ms). For each target vowel token the formants were then extracted using the Burg algorithm implemented in the Praat speech processing software (Boersma & Weenink, 2019). I looked for five formants in a bandwidth from 0 to 5000 Hz. Formant tracks were overlaid on a wideband spectrogram so that I could visually check whether or not all formant tracks coincided with a dark band of energy in the spectrogram. I wanted at least the lowest two formants (F1, F2) to match convincingly, preferably the lowest three (F1, F2, F3). If no satisfactory visual match could be obtained, the number of formant and/or the maximum frequency was lowered. In a number of cases I had to ask for more formants than could be expected within the given frequency band, in order to force a visual match (this was often needed for the vowel /i/ in *need*, when a spurious formant was found in between the regular F1 and F2; the spurious formant was later deleted from the measurement).

I marked all tokens which seemed incorrectly pronounced, and deleted them from the dataset. Incorrect pronunciations occurred in the vast majority of the tokens of *sawed* and *hawed*, which were then pronounced with a full diphthong /ɑʊ/ (as in *cloud*). As a result of

this, the total number of /ɔ/ tokens was as little as 10 (which is 11% of the number of attempts). The words *bed* and *head* were repeated by the speakers at the end of the vowel reading to prevent list-final intonation. These repeated tokens were eliminated from the set of measurements; in one instance, however, a repeated token was substituted for a token of *head* which was incorrectly pronounced the first time. The formant values were averaged per token over the entire duration measured for the vowel, i.e., including the (hardly present) onset transition and the offset transition into the final /d/. The final transition will affect the F1 in like fashion for all vowel types (F1 always goes down during a final transition). F2 will be affected differently depending on the vowel: for front vowels F2 /i, ɪ, e/ the F2 will fall somewhat during the last 40 ms of the vowel, while it will rise for back vowels. As a result, the mean F2 values per token may suggest a slight centralization along the F2 dimension. Nevertheless, we decided not to shift the end of the measurement interval to the beginning of the final transition, in order to maintain comparability with our control data (Wang, 2007; Wang & Van Heuven, 2006), where the formants were also averaged over the entire vowel duration. Given 45 speakers and two tokens per speaker, the maximum number of tokens was 90 per vowel type. A breakdown of the numbers per group is seen in Table 6.1.

Table 6.1. Number of vowel tokens suitable for statistical analysis broken down by gender of speaker and by Language background.

Gender			Vowel											Total
			æ	ɑ	e	ɛ	i	ɪ	o	ɔ	u	ʊ	ʌ	
Female	Language	Bilingual	26	26	26	26	26	26	24	1	24	26	26	257
		Monolingual	22	22	21	22	21	22	20	2	21	22	22	217
	Total		48	48	47	48	47	48	44	3	45	48	48	474
Male	Language	Bilingual	22	21	22	22	22	22	22	0	22	22	22	219
		Monolingual	20	20	20	19	20	20	17	7	18	20	20	201
	Total		42	41	42	41	42	42	39	7	40	42	42	420
Total	Language	Bilingual	48	47	48	48	48	48	46	1	46	48	48	476
		Monolingual	42	42	41	41	41	42	37	9	39	42	42	418
	Total		90	89	89	89	89	90	83	10	85	90	90	894
	Missing		0	1	1	1	1	0	7	80	5	0	0	96

I will now first present an informal, visual comparison of the EFL and native AE vowels. In later sections I will present numerical data and inferential statistics.

6.3.2. Location of vowel centroids in F1 by F2 plane

Figure 6.1A-B-D-E plots the vowels, after Bark conversion, in an acoustic F1 by F2 plane. The centroids are represented here as colored dots at the intersection of the F1 and F2 coordinates, and are identified by their phonetic symbol. The results are paneled by Language, i.e., whether the speakers were monolingual Persian (panels A-B) or bilingual Azerbaijani/Persian, (panels D-E), and by Gender (boys in panels A-C, girls in panels B-D). The raw formant (and duration) data (in Hz and in ms, respectively), broken down by Context and Gender, for all groups of speakers, are included in Appendix 6.2A-E.

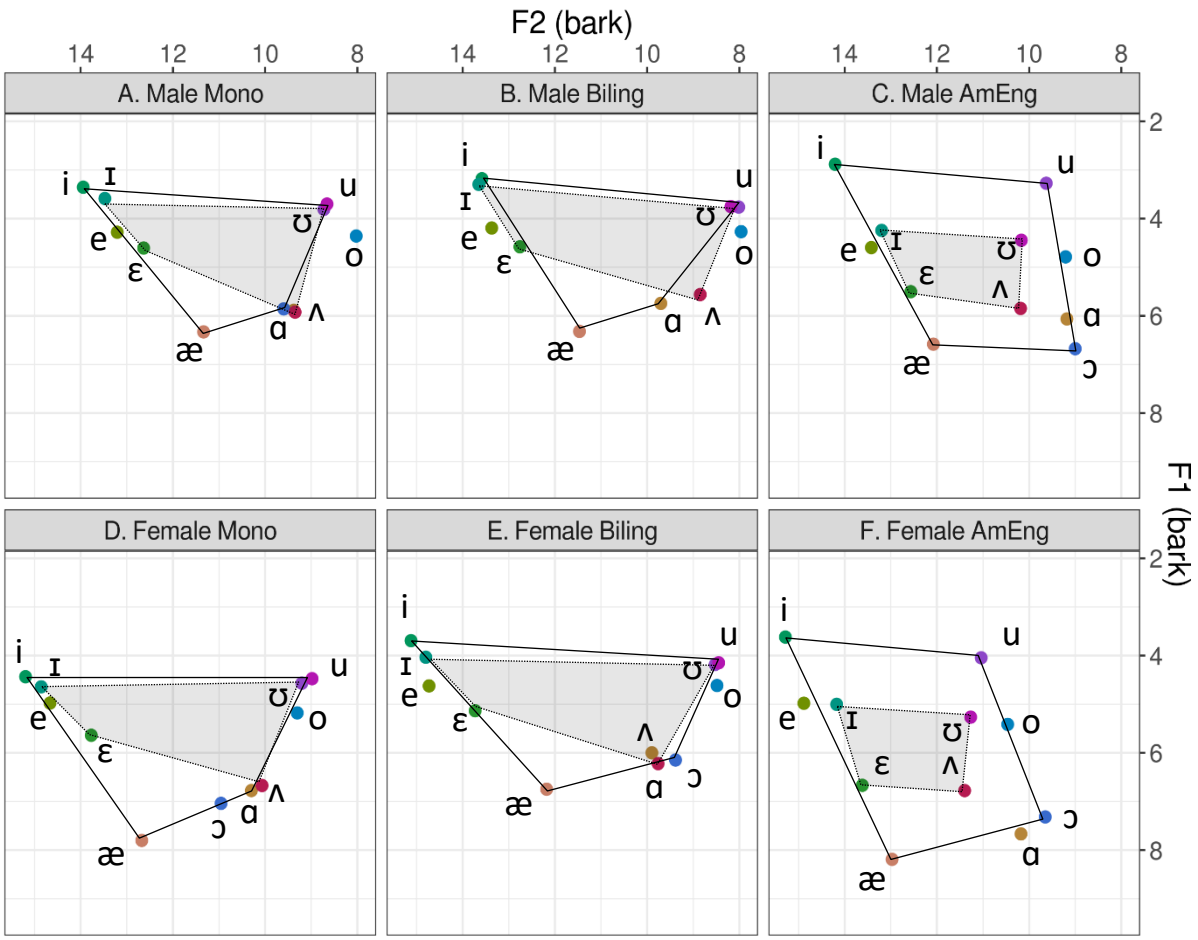


Figure 6.1. Centroids of the eleven American English monophthongs in an F1 by F2 plane (axes in Barks) as produced in /hVd/ items by monolingual Persian (left, panels A, D) and bilingual Azerbaijani/Persian (mid, panels B, E) adolescent learners of English as a foreign language, broken down by gender of the speaker (upper: male, panels A, B; lower: female, panels D, E). No tokens of /ɔ/ were produced for panels A, B. Convex hulls are drawn around the long ('tense') corner vowels. The shaded quadrilaterals in the center of the diagrams join the four short ('lax') vowels. The right-most upper and lower panels represent the same information obtained for ten male (panel C) and ten female (panel D) American native speakers.

The vowels spoken by the boys have lower formant values: all vowels are more towards the top-right of the vowel space than is the case with the girls. This is because boys, during and especially after puberty, have larger resonance cavities than girls. Because of the larger cavities, the resonance frequencies (i.e., formants) are some 15% lower for men than for women. The centroids tend to cluster in small groups. There is a high-back cluster /u, ʊ, o/, a low back cluster /ɔ, ɑ, ʌ/, a front vowel cluster /i, ɪ, e, ɛ/, and a singleton low front /æ/. This patterning is seen for all four groups alike, although there would seem to be some distance between subclusters /i, ɪ/ and /e, ɛ/ for the female speakers. What is especially revealing is that there are no vowels in the center portion of the space. It would appear that monolinguals and bilinguals basically have the same structure in their vowel system of American English. The structure of the EFL vowel system averaged over de four speaker groups is shown in Figure 6.2A. For the sake of visual comparison, panel B shows the location of the 11 centroids for the American speakers (averaged over genders).

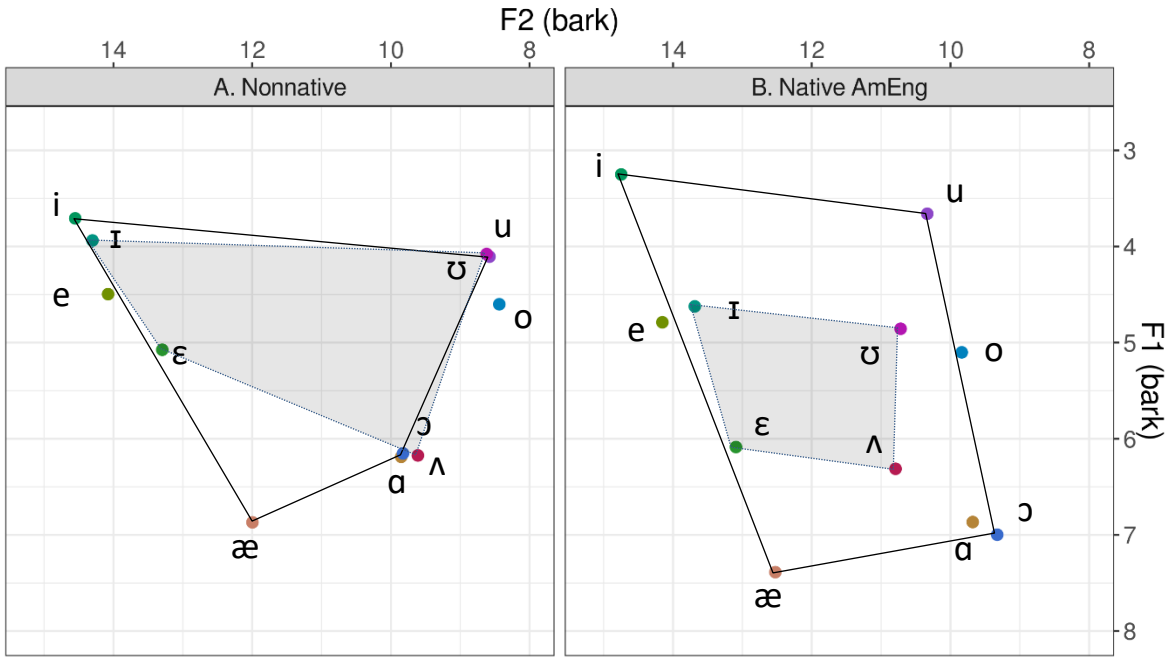


Figure 6.2. Panel A: as Figure 6.1 but averaged over the four Iranian speaker groups. Panel B: as Figure 6.1 but averaged over the male and female native speakers of American English.

In Figures 6.1 and 6.2, the long peripheral vowels (sometimes called tense, see §§ 2.2 and 4.2) at the corner points of the vowel quadrilateral are joined by a polygon. A shaded smaller polygon joins the four short and rather more centralized vowels (also called 'lax'). It can be seen that the lax vowels form an inner polygon in the L1 control data. It is obvious that the Iranian EFL learners do not produce the clear difference between the four short centralized lax vowels and the seven peripheral long vowels. All the vowels in Figure 6.2A lie on the outer edge of the vowel space. Especially /ʌ/ and /ʊ/ should be much more centralized than is done by the EFL speakers.

6.3.3. Dispersion and overlap of vowel categories in EFL and native AE

Plotting the centroids does not tell the whole story. So let us also plot the spread of the vowel tokens and consider the amount of overlap between adjacent vowel categories. Figure 6.3A-F plots spreading ellipses around the centroids shown in Figures 6.1. These ellipses contain (theoretically) the 46% most typical tokens of the vowel category.[21] Individual vowel tokens have been connected with a straight line to their centroids.

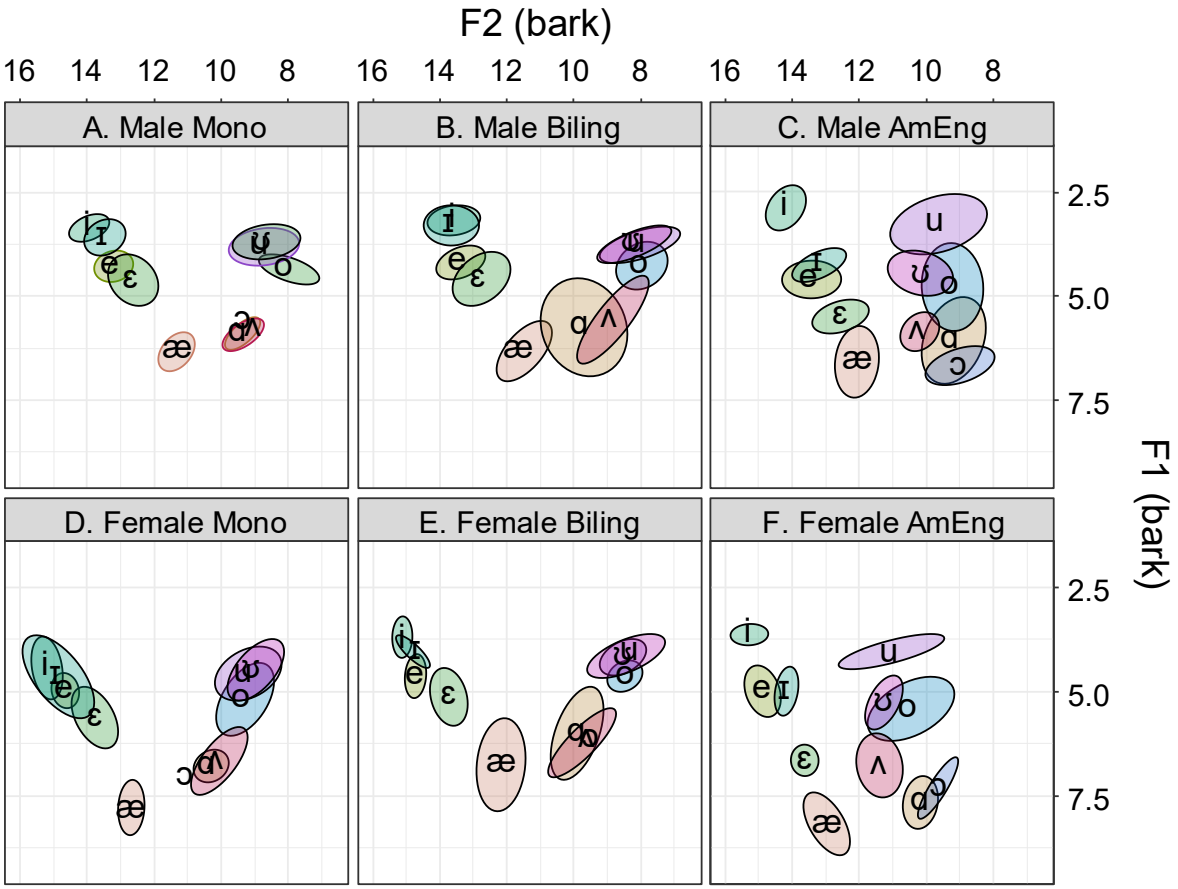


Figure 6.3. Centroids and dispersion ellipses for eleven American English monophthongs produced by monolingual and bilingual groups of EFL learners (/hVd/ items only), broken down by gender. Ellipses are drawn at ± 1 SD along the first two principal components of the scatter clouds. The right-most panels represent the control data produced by 10 male and 10 female native speakers of American English. See Figure 6.1 for details.

The overlap between /i/ and /ɪ/ is substantial in all four panels A-B-D-E. So is the (almost complete) overlap between /u/ and /ʊ/ (possibly, these contrasts will be upheld by a difference in vowel duration). There is a large difference in spectral distribution for the

[21] The ellipses are constructed by computing the first two principal components (PCs) of the scatter cloud around a specific vowel. The F1/F2 coordinates of each vowel token are then projected onto the first and the second PC of the scatter cloud. The ellipse is drawn at plus and minus 1 standard deviation away from the centroid along PC1 and PC2. Within 1 SD up and down from the mean of a distribution lie (theoretically) 67% of the data points. In a 2-dimensional distribution, as in our F1-by-F2 plots, the ellipses include 67% × 67% = 46% of the data points.

members of these long-short vowel pairs in the American control data, which are shown in panels C, F.[22]

In the American native control data, there is no visible spectral overlap between /i/ and /ɪ/, nor between /u/ and /ʊ/. Also, /ɛ/ is clearly separated from /e, ɪ/, and /ʌ/ from the low-back vowel cluster /ɔ, ɑ/. The short/lax vs long/tense mid vowel pairs /e, ɪ/ and /o, ʊ/ will be distinct by a difference in duration.

6.3.4. Vowel duration

The vowel durations are shown for the four groups (male, female; monolingual, bilingual) of Iranian EFL learners in Figure 6.4. Numerical data underlying the graphs can be found in Appendices A6.2A-D for the EFL learners and in A6.2E for the native speakers.

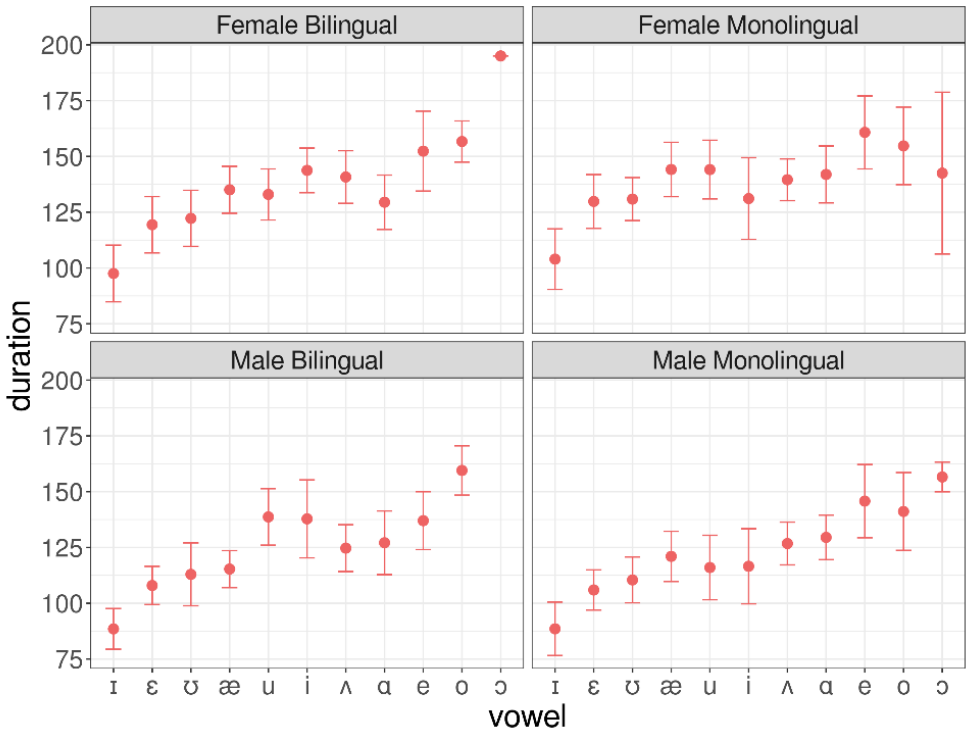


Figure 6.4. Duration (ms) for 11 monophthongs of American English produced by male and female monolingual and bilingual adolescent learners. Error bars are the 95% confidence limits of the mean. Vowels are arranged in ascending order of duration as found for all EFL speakers combined.

Clearly, the vowel durations are strongly correlated for the four learner groups. There does not seem to be a reason why we should keep the vowel durations separate. Figure 6.5, therefore, shows the vowel durations averaged over the four learner groups. For the sake of comparison, Figure 6.5 also shows the same information computed for the 20 native speakers

[22] The American native control plots are based on the original data made available to me by Prof. dr. Wang Hongyan of Shenzhen University. In Wang (2007) and Wang & Van Heuven (2006), the data on the (mid) low back vowel /ɔ/ were omitted. Here the data for all 11 monophthongs are plotted.

of American English. The vowels in Figure 6.5 are arranged in ascending order of duration as produced by the native AE speakers.

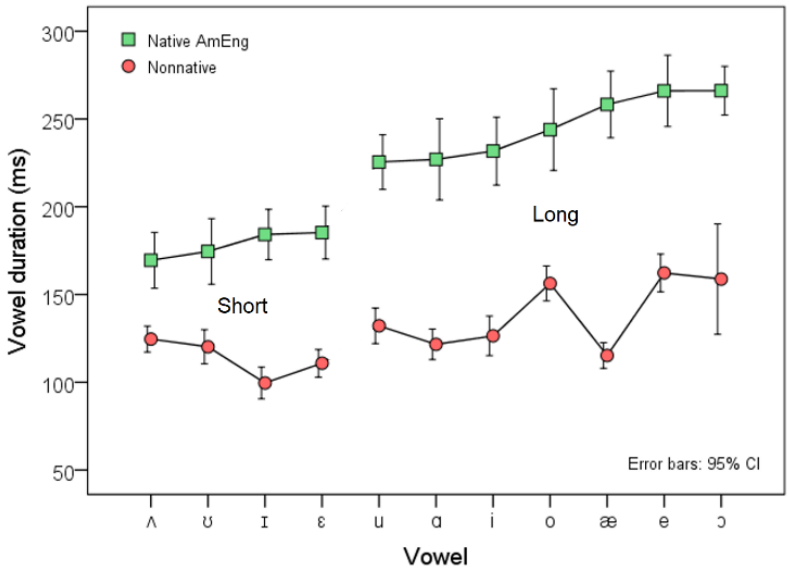


Figure 6.5. Duration (ms) of 11 AE target vowels, averaged over all four groups of EFL learners (red circles, *N* = 45) and over native speakers of American English (green squares, *N* = 20). Vowels are arranged in ascending order of duration as produced by native AE speakers. Error bars are the 95% confidence limits of the mean.

The shortest vowels are the centralized vowels /ɪ, ɛ, ʊ, ʌ/, which mirrors the control data. The longest vowels are peripheral /e, o/. The longest vowel of all is /ɔ/ but this mean is based on 10 tokens only (see Table 6.1). The remaining five vowels have intermediate lengths. This corresponds well with the native control durations, with the exception of just one vowel, i.e., /æ/ which should be in the long vowel category but is even a bit shorter than the longest of the short vowels, i.e., /ʌ/.

In the native AE control data /ʌ/ is the shortest vowel of all. All four short ('lax') vowels are shorter than the long ('tense') vowels. In the control data /æ/, although phonologically lax on distributional grounds, is clearly a long vowel in American English (see §2.2 and §4.2). The EFL learners pronounce /æ/ short, which would be inappropriate for American English. In spite of the exceptions, the relationships among the vowel durations are approximately correct and mirror the duration ratios found for the native control data. The correlation between the vowel durations in the native and nonnative realizations is r = .690 (p = .009, one-tailed). However, absolute vowel durations are about 90 ms shorter for the EFL learners (130 ms) than for the native controls (221 ms). This difference is highly significant by a paired t-test, $t(10) = 11.1$ ($p < .001$). We will come back to this observation in the discussion section.

6.3.5. Inferential statistics for spectral parameters

In this and the following section I present inferential statistics for the effects and interactions observed informally in the previous sections. F1 and F2 values will first be analyzed separately in order to determine whether there are simple, one-dimensional distinctions among the vowel types. The multivariate analysis will be done in § 6.3.6. by Linear Discriminant Analysis (LDA). A Repeated Measures Analysis of Variance (RM-ANOVA) was performed with vowel type (excluding *sawed*/*hawed* with only 10 tokens) and context (CVd ~ hVd) as within-speaker factors. Language (monolingual ~ bilingual) and Gender (male, female) were between-speaker factors. Missing values (for incorrectly pronounced items, see Table 6.1) were restored by non-iterative two-dimensional imputation. I computed the marginal means for rows (= speakers) and columns (= vowel-context combination) in the data matrix, skipping the cells with missing values. I then determined per speaker and per vowel/context combination how much the row and column marginal deviated from the grand mean, and then replaced the empty cell by the grand mean plus the deviation in the two dimensions. The imputed values were thus equal to what would be predicted by linear addition of the speaker effect and the vowel+context effect with no adjustment for a possible two-way interaction. Imputation of missing values was necessary because any speaker with a missing value (even for a single vowel token) would be deleted by the RM-ANOVA. Missing values were found for 10 out of 45 speakers, so that close to 25% of the speakers would have to be discarded from statistical analysis, even though the number of missing values was never more than two for any speaker (total number of missing values was 16 out of 900 cases (= 1.7%).

Table 6.2 summarizes the RM-ANOVAs for F1 frequency (after conversion to Barks), and F2 frequency (in Barks), and Vowel duration. The table specifies all possible main effects and interactions, separately for the within-speaker and between-speaker terms. In none of the three analyses was the condition of sphericity met, so that I used the Greenhouse-Geisser correction of degrees of freedom as a safety precaution. In the table, however, I list the nominal degrees of freedom (for the sake of clarity); the *p*-values listed were computed after GG-correction. Except one, all factors in the RM-ANOVA are dichotomies, which require no post-hoc analyses for multiple contrasts. Vowel type, however, has 10 levels, which were tested pairwise by post-hoc t-tests with Bonferroni correction for multiple comparison. Our criterion for significance is $\alpha = .050$; however, to avoid having to factor in small effects we made the additional requirement that the effect or interaction should have an effect size of

partial eta squared $p\eta^2$ of at least .100. All significant effects/interactions are highlighted in the table; they are additionally bolded if the $p\eta^2$ requirement is met.

Table 6.2. Summary of RM-ANOVA. Dependent variables are F1, F2 and vowel duration. Within-participant factors are Vowel type, and Context (hVd, CVd). Between-participant factors are Gender of speaker, and Language background (monolingual Persian, bilingual Azerbaijani/Persian). All main effects and interactions are listed. Nominal degrees of freedom are reported but *p*-values were computed after Greenhouse-Geiser correction. Significant effects and interactions ($\alpha = .050$) are in highlighted cells. When the effect size $p\eta^2 \geq .100$, the cell entry is also bolded.

Tests of Within-Subjects Effects		**First formant**			**Second formant**			**Duration**		
Effect/Interaction	$df_{1,2}$	F	p	$p\eta^2$	F	p	$p\eta^2$	F	p	$p\eta^2$
Vowel	9, 369	256.0	**<.001**	**.862**	887.7	**<.001**	**.956**	36.3	**<.001**	**.470**
Vowel * Gender	9, 369	1.4	.234	.033	4.8	**.001**	**.105**	0.6	.691	.015
Vowel * Language	9, 369	0.6	.662	.015	2.7	.036	.061	2.2	.057	.050
Vowel * Gender * Language	9, 369	0.6	.709	.014	0.6	.627	.015	0.7	.646	.016
Context	1, 41	0.2	.699	.004	18.4	**<.001**	**.310**	15.2	**<.001**	**.270**
Context * Gender	1, 41	0.3	.564	.008	0.1	.740	.003	0.6	.435	.015
Context * Language	1, 41	0.4	.552	.009	0.7	.411	.017	6.0	**.018**	**.128**
Context * Gender * Language	1, 41	2.5	.124	.057	0.2	.632	.006	1.8	.192	.041
Vowel * Context	9, 369	12.3	**<.001**	**.230**	12.3	**<.001**	**.231**	23.9	**<.001**	**.369**
Vowel * Context * Gender	9, 369	4.5	**<.001**	**.100**	1.3	.246	.032	0.7	.628	.018
Vowel * Context * Language	9, 369	0.8	.624	.018	1.3	.283	.030	1.5	.160	.036
Vowel * Context * Gender * Language	9, 369	2.0	.049	.048	0.9	.508	.021	0.7	.677	.016
Tests of Between-Subjects Effects										
Gender	1, 41	47.5	**<.001**	**.536**	62.2	**<.001**	**.603**	4,0	.053	.088
Language	1, 41	10.9	**.002**	**.211**	2.4	.132	.055	0.0	.944	.000
Gender * Language	1, 41	6.3	**.016**	**.132**	0.6	.458	.014	0.5	.477	.012

First formant (F1). The effect of Vowel is significant. Bonferroni post-hoc tests indicate that the four (half-)close vowels /i, ɪ, u, ʊ/ have the same F1; so do the semi-diphthongs /e, o/, and the pair /ɑ, ʌ/. The F1 of /ɛ/ and /æ/ differ significantly from one another as well as from any other vowel.

Context has no effect on F1, and does not interact with other factors, with the exception of a small third-order interaction between Vowel, Context and Gender. Female speakers have higher F1 values than males because they have smaller/shorter resonance cavities, so the Gender effect is predictably significant. Also, the Vowel-by-Gender interaction is significant. Smaller but still significant effects are found for the Language background of the speakers, and for the Language-by-Gender interaction. Finally, there is a small (but significant four-way interaction between Vowel, Context, Gender and Language background. Figure 6.6 plots the F1 center frequency for the 10 vowels that remain in the dataset, broken down by Gender (row panels) and by Language background of the EFL learner (column panels). The vowels are ordered from

left to right in descending order of F2 frequency as they were found in our American English control data (i.e., I used the same axis layout as in the next figure, which plots F2).

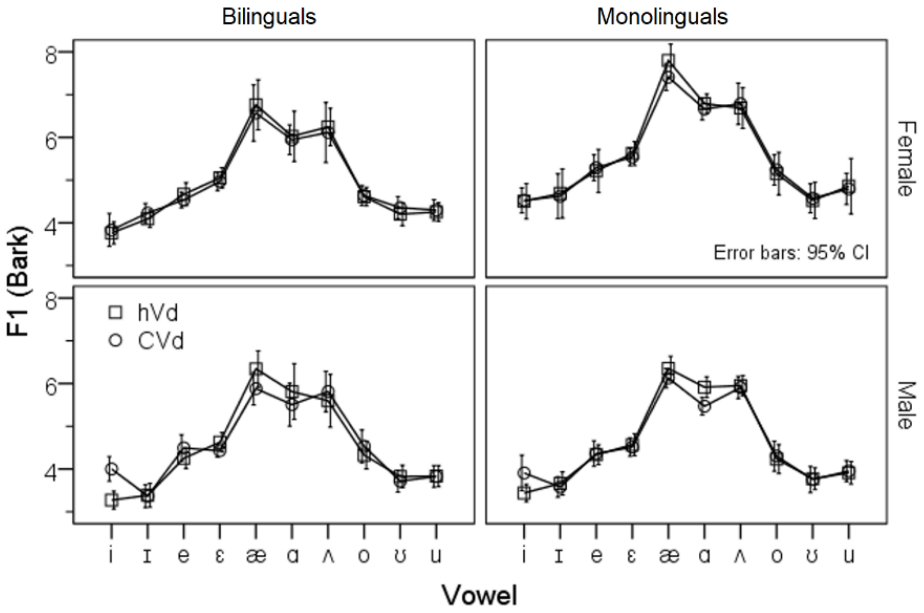


Figure 6.6. F1 center frequency (in Barks) of ten American English monophthongs produced by male (lower panels) and female (upper panels) monolingual Persian (right-hand panels) and early bilingual Azerbaijani/Persian (left-hand panels) EFL speakers. The vowels are ordered from left to right in descending magnitude of F2 (i.e., vowel backness). In each panel separate curves are plotted for /hVd/ and /CVd/ consonant frames. Error bars are the 95% confidence intervals of the means.

The effect of Context is very small visually, and can be seen only for the open vowels /æ, ɑ/, which explains the significant Vowel-by-Context interaction. Also, there is no visible difference in F1 between monolingual and bilingual male speakers; the female monolinguals, however, have systematically higher F1 values than their bilingual peers. This would explain the significant main effect of Gender as well as the interaction between Gender and Language.[23]

Second formant (F2). Figure 6.7 plots the F2 center frequency for the 10 vowels, broken down by Gender and by Language background. The vowels are ordered from left to right in descending magnitude of F2 as found in our American English control data.

[23] This is somewhat puzzling. No similar effect or interaction is seen for F2, so we can rule out the possibility that the high F1 values for monolingual women are caused by smaller resonance cavities (e.g., due to younger age). High F1 could then be a correlate of speaking softly (caused by shyness?): when we speak softly, we do not open our mouth as much as when we talk on a loud voice (F1 correlates with mouth openness).

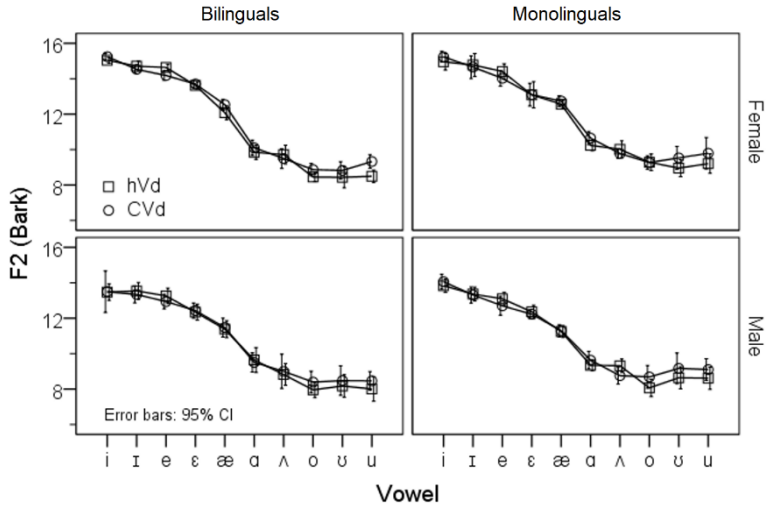


Figure 6.7. Center frequency of F2 (Barks) for 10 English monophthongs pronounced in /hVd/ words and in rhyming everyday keywords (/CVd/) by Iranian EFL learners, broken down by Gender and by Language background (monolingual Persian vs bilingual Azerbaijani/Persian). Error bars include the 95% confidence interval of the mean. Vowels are ordered from left to right in descending magnitude of F2 as found for American control data.

The effect of Vowel is very strong. The three (half-) close back vowels /u, ʊ, o/ do not differ from one another but all other pairwise vowel contrasts are significant. Gender is a predictably significant effect (see above). Also, the Vowel-by-Gender interaction reaches significance. Context exerts a small (but significant) effect on F2. Finally, the Vowel-by-Context interaction reaches significance. No other effects or interactions were found significant. Interactions between the main effects (even when significant) are hardly noticeable. There are only two effects that should be taken seriously: the effect of Vowel, and the effect of Gender. The vowel effect is what we are interested in. The effect of Gender will be neutralized in subsequent analyses through z-transformation within speakers (Lobanov normalization). The effect of Context (11.06 vs 10.96 Bark), even though significant, is very small indeed (± .05 Bark). The LDA in the next section will be run on the /hVd/ tokens only, since this is the only contexts for which we native control data (Wang & Van Heuven, 2006).

Duration. The effect of vowel was highly significant. Bonferroni post-hoc pairwise comparisons bear out that /ɪ/ is significantly shorter than any other vowel, while the semi-diphthongs /e, o/ are significantly longer than all other vowels. This latter finding suggests that duration is used by the EFL learners to differentiate lax /ɪ, ɛ, ʊ/ from their tense competitors /e, o/.

Vowel durations do not differ significantly between male and female speakers. There is no interaction between Gender and Vowel nor is there interaction between Gender and any other factor. Target vowels are pronounced longer in everyday keywords (132 ms) than in

/hVd/ items (126 ms). Moreover, the context interacts with the vowel type. Figure 6.8 shows the interaction.

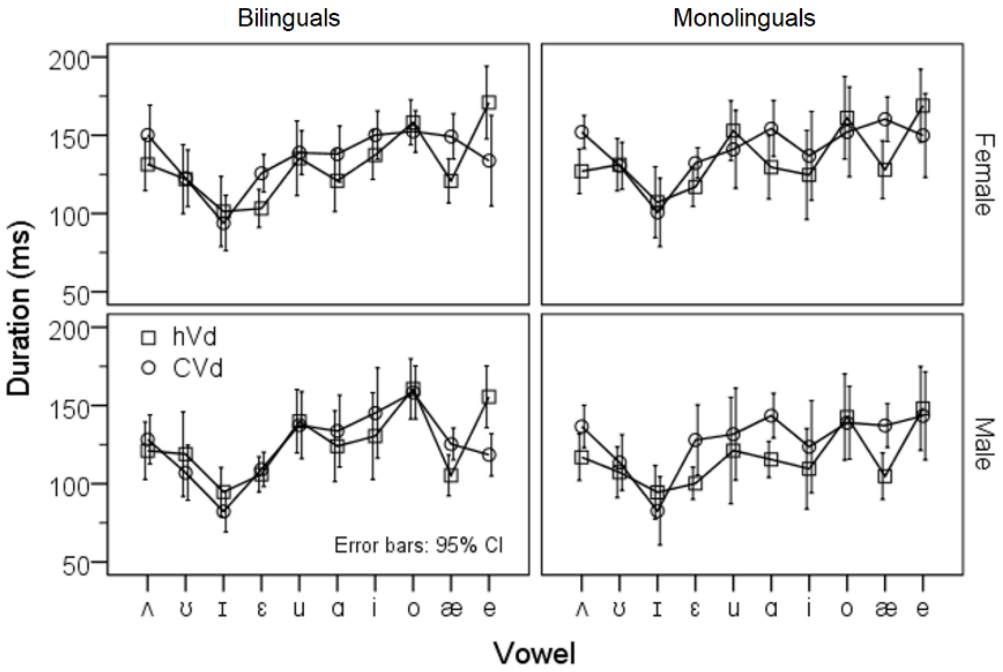


Figure 6.8. Vowel duration (ms) for ten American English monophthongs produced by four groups of Iranian learners of English as a foreign language. See Figure 6.7 for more information.

6.3.6. Multivariate analyses

In this section I will combine the acoustic vowel parameters (F1, F2, duration) in a multivariate analysis, in an attempt to automatically classify the ten American English vowel types as produced by the Iranian EFL learners. The ANOVA results above show that the three parameters we measured for our vowel tokens are very sensitive to differences between vowel types. The effect of Vowel in Table 6.2 is strongest for F2 ($p\eta^2$ = .956), second for F1 ($p\eta^2$ = .862) and third for Duration ($p\eta^2$ = .470). All other effects (and interactions) that can be seen in Table 6.2 are small (and often insignificant) in comparison and will not contribute significantly to automatic classification of the vowels. The one exception is the effect of Gender ($p\eta^2$ = .603 for F2, .536 for F1, insignificant for Duration) but this effect is not linguistically relevant, and will be factored out prior to the analysis through Lobanov normalization.

Two different classification algorithms were employed, i.e., Linear Discriminant Analysis (LDA, Klecka, 1981) and Multinomial Logistic Regression Analysis (MLRA, Hosmer & Lemeshow, 1989). This is the same choice that was made in Van Heuven et al. (2020). I once did the analyses separately for each of the four groups of speakers, and then – assuming that the groups do not differ from one another in any principled manner, I also ran the analyses for the four groups combined. To test the contribution of vowel duration

explicitly, all analyses were carried out once with only the spectral parameters, F1 and F2, and a second time with duration added. All analyses were done on the dataset with imputed values for missing cases (see above), after excluding the vowel /ɔ/ and subsequent within-speaker z-normalization of the acoustic parameters. Table 6.3 shows the percentage of correctly classified vowel tokens in each of the above conditions. The underlying confusion matrices can be found in Appendix 6.3 with hierarchical cluster trees to show the similarity structure of the ten vowels in the system (Appendix 6.4).

Table 6.3. Percent correct classification by Linear Discriminant Analysis and by Multinomial Logistic Regression Analysis of 10 American English vowels produced by four groups of Iranian learners of English as a foreign language, and for all groups combined. Percentages are listed for analyses with spectral parameters only (F1, F2) and with spectral parameters plus vowel duration. All predictors were z-normalized within speakers. Columns under Δ specify the difference due to addition of the duration parameter.

		LDA			MLRA			Δ MLRA – LDA	
		F1, F2	+ Dur	Δ	F1, F2	+ Dur	Δ	F1, F2	+ Dur
Female	Bilingual	68.1	72.3	+4.2	69.2	76.2	+7.0	+1.1	+3.9
	Monolingual	62.7	70.5	+7.8	66.4	70.9	+4.5	+3.7	+0.4
Male	Bilingual	59.1	67.3	+8.2	61.4	72.7	+11.3	+2.3	+5.4
	Monolingual	73.0	75.0	+2.0	74.5	79.0	+4.5	+6.0	+4.0
All		66.6	70.7	+4.1	67.6	72.4	+4.8	+1.0	+1.7

On the basis of the two spectral parameters, F1 and F2 formant center frequency, the AE vowels produced by our Iranian EFL learners can be automatically classified in between 59 and 73% correct by LDA. Given 10 vowels to choose from, this is approximately 7 times better than chance (= 10% correct). Automatic classification by MLRA is slightly better, with between 61 and 75% correct. Adding vowel duration as a third predictor improves the correct classification by 2 to 8 points for the LDA method, and by 5 to 11 points for the MLRA method. Note that these measures merely indicate how well the EFL learners separate their AE vowels. It does not mean that they separate them in the same way American native speakers do. Our hypothesis should be that native speakers keep the vowels more distinct, and may well have different centroids and boundaries between adjacent vowel categories.

6.3.7. Classifying non-native vowels by native models

Ideally, all the vowel tokens produced by the Iranian EFL learners should be offered for perceptual identification. However, since each listener would have to respond to 894 different stimuli (Table 6.1), even if each token would be presented only once, we decided not to involve human listeners at this stage of the project. Instead, I used the Linear Discriminant classifier (see above) as a substitute for a group of human native listeners, as done before by,

e.g., Strange et al. (2004), Wang & Van Heuven (2006), and many others. It has been found that an LDA classifier, when properly trained with vowel tokens produced by a representative group of native speakers, divides up the multidimensional vowel space approximately the same way native listeners of the language do. If we then use the classifier to identify new tokens, for instance tokens produced by non-native speakers, it will identify the new tokens as of they were produced by native speakers. This way, the LDA classifier is a substitute for the human native listener. In this section we will use the American native vowel tokens made available to us to train an LDA classifier, and then use the model to identify the vowel tokens produced by the monolingual and early bilingual Iranian EFL speakers. Given the earlier observations, we expect to find the same number and type of identification errors for both groups of EFL learners. All predictors were z-normalized within speakers. Leave-one-out cross-classification was applied when the native-speaker model was used to classify the native vowels.

Table 6.4. Percentage of correct vowel identification by Linear Discriminant Analysis with spectral parameters F1, F2 or with spectral parameters plus vowel duration. All predictors were z-normalized within speakers.

Training set	Test set	F1, F2	+ Dur	Δ
Native speakers	Native speakers	83.5	89.5	+6.0
Native speakers	All Iranian EFL learners	48.9	56.7	+7.8
Native speakers	Monolingual Persian EFL learners	48.8	56.9	+8.1
Native speakers	Bilingual Azerbaijani/Persian EFL learners	49.0	56.5	+7.5

As before, all predictors made a significant contribution to the model. With only the two spectral parameters the native speaker vowels were correctly classified at 84% correct, which better than 8 times chance-level performance (chance = 10% correct). Adding (normalized) vowel duration as the third predictor raises the percentage of correct classification to 90 (9 times better than chance). When the native classification model is then applied to the EFL tokens, the performance decreases dramatically, to 49% correct for spectral parameters only and 57% correct when duration is added as a predictor. As predicted, the results are virtually identical for the two groups of EFL learners: monolinguals and bilinguals differ by only tenths of a percentage point.

If we contrast the correct classification in Table 6.4 (classified by native speaker norms) with the classification by the LDA when trained by the EFL learner groups themselves in Table 6.3, it can be observed that the scores are considerably better in the latter situation: 67% correct for spectral parameters and 71% correct with duration added. This improvement relative to the results in Table 6.4 can be considered an instance of the Interlanguage Speech Intelligibility Benefit (ISIB). The claim is that listeners who share their native language with

the EFL speaker, intuitively know the peculiarities imparted by the shared native language to the foreign speech. Listeners who do not share the native language of the EFL speaker, cannot use these subtleties. This would also explain why listeners who share the native language with the EFL speaker tend to outperform native listeners of the target language (Bent & Bradlow, 2003; Wang & Van Heuven, 2015; Van Heuven, 2016).

A breakdown of the classification results by vowel type is given in Figure 6.9. The figure shows the percentages of correct vowel classification by the LDA when trained on American native vowel tokens, and tested on either the same group of native speakers (circles), when tested on the monolingual Persian EFL learners (triangles), and on the early bilingual Azerbaijani/Persian learners (squares). Separate curves are drawn for prediction based on spectral parameters only (open markers, dotted lines) and for prediction with vowel duration added (closed markers, solid lines.

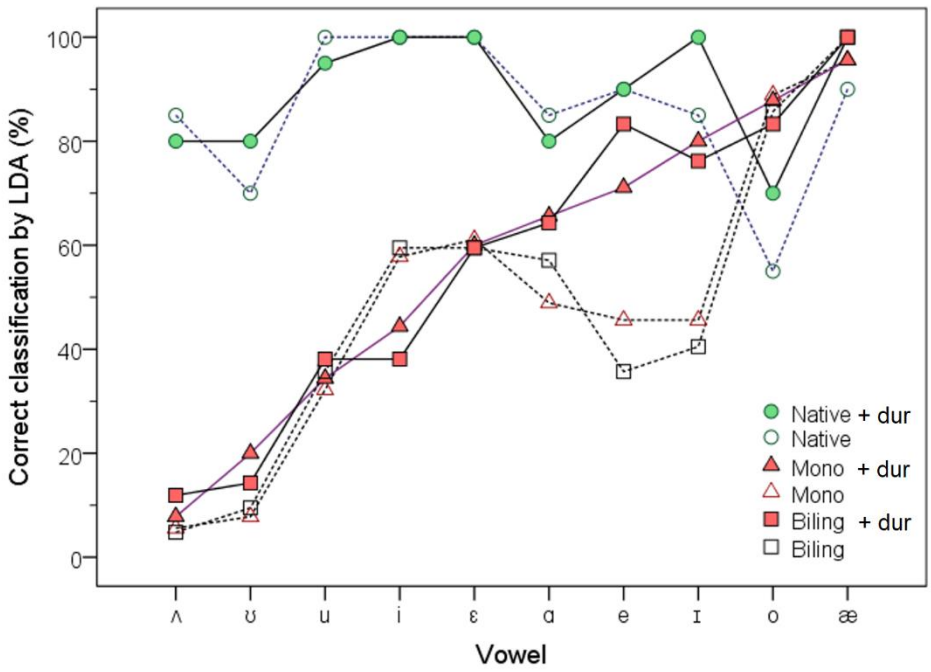


Figure 6.9. Correct classification (%) of ten American English vowels by Linear Discriminant Analysis trained on native vowel tokens and tested on the same tokens (20 speakers, circles), and on EFL tokens produced by monolingual Persian (21 speakers, triangles) and bilingual Azerbaijani/Persian (24 speakers, squares) learners. Classification is done with spectral parameters only (open symbols) or with vowel duration added (closed symbols). All predictors are z-normalized within speakers.

The first thing we notice is that, again, there is hardly any difference between the two groups of EFL learners. This confirms our earlier conclusion that the extra three central vowels of Azerbaijani offer no advantage to the bilingual EFL learners over the six-vowel inventory of Persian. Some EFL vowels are very poorly identified. The largest number of

mis-classifications (relative to the native classifications) are found for the vowel types /ʌ, ʊ, u, i, ɛ/. The vowel /æ/ is classified correctly as often as for native speakers. The vowels /e, ɪ/ are correctly classified only if duration is added as a predictor, which shows that this contrast critically depends on the temporal difference made by the EFL learners. A better view of the pronunciation problems of the EFL learners is afforded if we consider the confusion structure in the LDA classification.

The full confusion matrices of intended versus actually classified vowel are included in Appendix 6.3. On the basis of the matrices simplified confusion graphs can be constructed, which highlight the most frequently occurring mis-classifications (errors) in the pronunciation of the EFL speakers (which can also be compared with the confusion structure obtained for the native speakers). Figure 6.10 shows the confusion structure obtained for the classification of the American native-speaker classification based on all three acoustic predictors. The ten vowels are arranged as if in a traditional vowel diagram, with front vowels to the left, back vowel to the right, closed vowels at the top and open vowels at the bottom. The lax and centralized vowels are in the center of the diagram, and shaded for visual contrast.

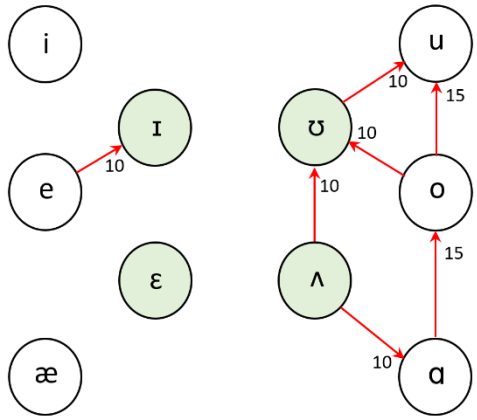


Figure 6.10. Vowel confusion structure for classification by LDA of ten American English monophthongs produced by and tested on 20 American native speakers. Predictors were F1, F2 and vowel duration. Confusions < 10% have been omitted. Lax/short vowels in shaded circles. Arrows point away from the intended vowel to the incorrectly identified vowel. The percentage of confusion is indicated at the arrow heads.

There is hardly any confusion among native-speaker vowels, and never more than in 15% of the identifications. There are no confusions between non-adjacent vowels, nor between front and back vowels. There is a tendency for back vowels to be more vulnerable to confusion than front vowels.

Figure 6.11 presents similar confusion graphs obtained for the monolingual Persian (panel A) and the early bilingual Azerbaijani/Persian (panel B) EFL learners.

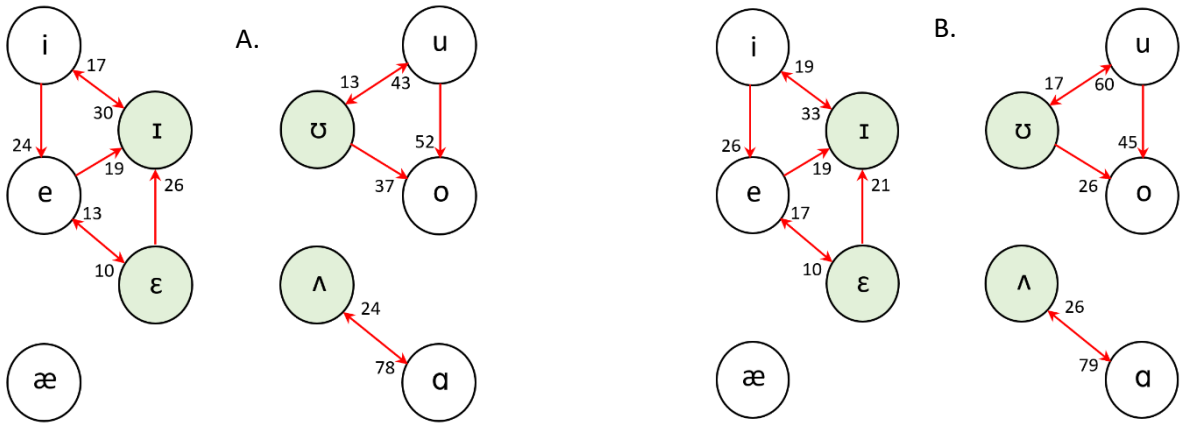


Figure 6.11. As for Figure 6.13 but for monolingual Persian (panel A, left) and for early bilingual Azerbaijani/Persian (panel B, right) learners of English as a foreign language.

The confusion structure is virtually the same for the two groups of Iranian EFL learners. It is also plain that there is substantial confusion in both groups. Yet the confusion is limited to specific clusters of adjacent vowels only. Moreover, there is never any confusion between front and back vowels. One cluster is formed by the close and mid front vowels /i, ɪ, e, ɛ/, where /ɛ/ would seem to be articulated somewhat too close while /i/ is articulated not close enough. Similarly, there is substantial confusion within the close and mid back vowels /u, ʊ, o/, where especially the vowel /u/ is a source of confusion. This vowel is very often classified as /o/, whereas short /ʊ/ is typically classified as long /u/. These observations provide numerical support for the informal observations that were made earlier on the basis of Figure 6.2. The last cluster is formed by the low back vowels /ʌ, ɑ/, a strongly asymmetrical confusion pair where the EFL learners produce /ɑ/ when /ʌ/ is intended. Their articulation of short /ʌ/ is insufficiently centralized (in fact, the /ʌ/ centroid is further back than the centroid of /ɑ/, see Figure 6.1).

6.4. Conclusions and discussion

In this chapter we examined the monophthongs of American English as produced by our two groups of Iranian adolescent learners of English as a foreign language, i.e., one group of monolingual Persian speakers and a second group of early Azerbaijani/Persian bilingual speakers. The articulation of the vowels was quantified in terms of acoustic correlates, whereby jaw aperture (the articulatory close-open dimension) was expressed as the first vowel formant (F1) and degree of backness (and lip rounding, which together with backness determines the length of the mouth cavity) was captured by the second vowel resonance, F2. Vowel duration was measured as a third acoustic property. The results obtained for the EFL

learners were compared with control data produced by American native speakers which were made available to us.

The results show, first of all that no differences were found in the realizations of the American English vowels between the monolingual and the bilingual EFL learners. The vowel systems of Persian and of Azerbaijani are very similar, with the exception that Azerbaijani has three close or mid-close central vowels where Persian has none. We entertained the hypothesis that the pronunciation of the American central vowel /ʌ/ might benefit from the presence of one or more central vowels in the Azerbaijani inventory, but our results do not support this hypothesis. Probably, the /ʌ/ vowel is too open to be a felt a reasonable target for one of the central vowels of Azerbaijani. This explanation clashes to some extent with the results found in Chapter 4. The /ʌ/ was assimilated to each of the three central vowels of Azerbaijani as much as it was assimilated to /o/. Nevertheless, the assimilation patterns were so weak that /ʌ/ remained Unclassified. The observed realization as /ɑ/ is not predicted by the assimilation patterns found in Chapter 4, which would predict a prevalent realization of /ʌ/ as /o/, which is not what was found. We conclude, therefore, that knowing a second vowel system at a native level provides no advantage to our bilingual EFL learners.

On the basis of the present results, I would predict that the intelligibility of my participants' English would be severely compromised by the flawed pronunciation of the vowels. Given the massive confusion structure, by which ten contrastive vowels are reduced to four clusters within which contrasts are weak or completely absent, words will be difficult to recognize, and the listener will have to depend much more on context (which is also poorly pronounced). This expectation is based on the pervasive confusion structure found by the automatic vowel classification through Linear Discriminant Analysis. It should be pointed out, in this matter, that the LDA may have overestimated the importance of vowel duration in determining the percentage over correct identifications. It is true that the Iranian EFL speakers generally differentiate between long and short vowels, but only in terms of duration ratios. In absolute terms all EFL vowels were so short that they would fall into the short-vowel category in English. By applying z-normalization within speakers, we abstracted away from absolute vowel durations, assuming that human listeners would do the same. This assumption, however, is speculative and requires additional research to find out whether human listeners (whether American native listeners, or non-native listeners) are quick to adjust duration boundaries once they are confronted with speakers who pronounce all vowels shorter than is normal for English).

Incorrect shortening of the AE vowels has also been found by Farran (2022) for Palestinian Arabic EFL learners. The probable cause for the incorrect vowel durations is a

failure on the part of the EFL learners to lengthen the vowel preceding coda /d/ (Van Heuven & Farran, 2022). In American English vowels should be at least 100 ms longer before coda /d/ than before coda /t/ (Peterson & Lehiste, 1960: table II). In the same materials as in my study, Farran's speakers produced vowel durations before coda /d/ that were equally long as before coda /t/. If I add the 100 ms appropriate before /d/ to the Iranian EFL vowel durations, most of the discrepancy between the Iranian EFL learners and the L1 AE speakers is eliminated.

When asked to rank-order the vowel pronunciation problems for Iranian EFL learners, I would suggest that the highest priority should be to teach the learners to produce a proper low central vowel /ʌ/, i.e., by making it clearly distinct from low back /ɑ/. The second-largest problem resides in the contrasts among the three (mid) high back vowels. Tense /u/ and its lax counterpart /ʊ/ are articulated at the same position in the vowel space by the Iranian EFL learners: the centroids are in the same position, and the spreading ellipses overlap completely for both groups of learners. To be true, the learners make a difference in duration but the difference turns out to be, counter to what we expected, smaller than in the native control data. In view of the asymmetric confusion found (Figure 6.11A-B), the learners should make an effort to make the lax /ʊ/ shorter rather than make tense /u/ longer. The learners should also be taught to front tense /u/, rather more the way the native control speakers do, so as to avoid confusion with /o/. Somewhat speculatively, it would also help if the learners were taught to diphthongize tense /o/ (and /e/ as well), the way native speakers do. In a future extension of my study, I will analyze the degree of diphthongization for the tense mid vowels, to see whether indeed the Iranian learners fail to diphthongize the tense mid vowels of American English. The third group of vowels that tend to be confused is the (mid) close front cluster. The confusions here are relatively mild, but it would certainly help if the learners were instructed to lower the articulation of mid-close /e/ and /ɛ/ so as to reduce confusion with the closed counterparts /i/ and /ɪ/.

A last point to be made here is that the confusion structure found on acoustic grounds in the present chapter strongly resembles the confusion structure that was revealed earlier in Chapter 5 on the basis of perceptual labeling of artificial vowels. In the next (conclusion) chapter I will compare the present production data with the earlier perceptual labeling data of Chapter 5 to see whether or not perception leads production in foreign language acquisition.

Chapter 7

Discussion & Conclusions

7.1. Introduction

In this final chapter I will summarize the experiments that were reported in the preceding chapters, and explain how each experiment was set up to provide missing pieces of a larger puzzle. Next, I will try to answer the twelve questions that were raised at the end of chapter 2. The questions will be repeated one by one in separate sections, and I will formulate the answer(s) based on the evidence that was obtained in the experimental chapters. When the most likely answer was predicted in Chapter 2, I will additionally conclude whether the hypothesis was confirmed or has to be rejected. In the last part of this chapter, I will consider the limitations of the present study, and formulate recommendations for future research and discuss possible implications of the present results for the teaching of English as a foreign language in Iran as a second or a third language.

7.2. Summary of experiments

The present dissertation aimed to find out how the native language or languages of monolingual Persian and bilingual Azerbaijani/Persian adolescents in secondary school influence the acquisition of English as a foreign language, i.e., a second language in the case of monolingual learners and a third language for the bilingual learners. The domain in which phenomena would be studied, was restricted to the acquisition of the phonology of English, specifically American English, which is the pronunciation norm used in the Iranian educational system. Within the realm of phonology, the experimental work concentrated on the acquisition of the vowel system, specifically the eleven monophthongs (also called pure vowels), thereby excluding phenomena related to diphthongs (vowel glides), consonants, syllable structure, and prosody (word and sentence stress, intonation).

In **Chapter 4**, we examined how the eleven American English (AE) monophthongs were assimilated into the six (in the case of Persian) or nine (in the case of Azerbaijani) vowels by 22 monolingual Persian EFL learners, and by 27 (early) bilingual Azerbaijani/Persian EFL learners. The bilingual learners performed the vowel matching task twice, once matching the AE vowels with the six Persian vowels, and a second time with the nine Azerbaijani vowels. In the assimilation task, the participant heard a token of an AE

vowel, and had to decide which of the (six or nine) vowels in their native language the foreign sound was most similar to, and rate the degree of similarity ('goodness') on a scale from 1 to 5 (= highly similar). The results of the perceptual assimilation task may serve to predict learning difficulties in the acquisition of nonnative sounds. Most difficult are nonnative sounds (here vowels) that match equally well with one vowel type in the learner's L1. The results obtained in Chapter 4 show that the monolinguals and the bilinguals assimilate the AE vowels into the six vowels of Persian in virtually the same way. When responding in Azerbaijani mode, however, the bilinguals selected their native vowel /y/ as the modal (i.e., most frequent) response for AE /u/ (showing that they perceived the fronting of AE /u/) and their own /œ/ for AE /o/ and /ʊ/, again indicating that the bilinguals were sensitive to the centralized nature of these AE vowels. The AE mid-low central vowel /ʌ/, however, could not be matched with any vowel in either Persian or Azerbaijani, giving it the status 'unclassified'.

Since even early bilinguals differ in the degree to which they command and use their two native languages, we aimed, in **Chapter 3**, to establish the relative dominance of Azerbaijani (their home language) over Persian (the national language of education and instruction). The participants in the assimilation experiment of Chapter 4 therefore filled in the Language Experience and Proficiency Questionnaire (LEAP-Q), in which they indicated, for each of their languages, since when they were exposed to them and how well they rated their performance in these languages when speaking, writing and reading. The results confirmed that all bilinguals were exposed to Azerbaijani before they were exposed to Persian. However, participants differed strongly in the degree of dominance of Azerbaijani over Persian, to the extent that some came close to being perfectly balanced bilinguals. In an excursion, we made an attempt to correlate several measures of relative language dominance with the consistency with which the bilinguals had performed the perceptual assimilation task in Chapter 4, assuming that stronger language dominance of Azerbaijani would be reflected by greater consistency in perceptual assimilation to Azerbaijani than to Persian.

In **Chapter 5** the same monolingual and bilingual EFL learners listened to 86 artificial vowel sounds in /mVf/ nonsense items differing in degree of jaw aperture (7 perceptually equal steps of one bark unit along the F1 resonance dimension) and constriction place (9 equal steps of 1 Bark along the F2 dimension). Twenty impossible combinations were excluded a priori, after which the remaining 43 vowels were generated in a short (200 ms) and a long (300 ms) version. Listeners had to decide for each synthetic vowel token, which one of the 11 AE monophthongs was most similar to it. This procedure allows us to map out the perceptual representation of the AE vowel system in the mind of the EFL learner. Specifically, the results

reveal which location in the vowel space the listener considers as the most typical realization of the vowel (the vowel centroid) and how this location is shifted under the influence of vowel duration. The results obtained in this chapter were compared with similar data collected from native listeners of American English (using the same materials and procedures), made available to us by colleagues from abroad. The results indicate that there are hardly any differences between the monolingual and bilingual EFL learners in the way they identify the artificial vowels as tokens of the eleven AE vowel types. However, for both groups of learners alike, their perceptual representation differs markedly from what is found for the native control listeners, the most conspicuous difference being the fact that duration is hardly used by the native listeners to differentiate between the tense and lax subsystems in the AE vowel inventory, while the learners ignore small differences in vowel quality and overestimate the importance of vowel duration.

In **Chapter 6**, the same participants produced the 11 monophthongs of American English (as well as a lot of other materials, which were not analyzed in the present dissertation) in everyday keywords and in /hVd/ words. The acoustic correlates of jaw aperture (also called vowel openness or vowel height) and constriction place (along the front-back dimension), i.e., F1 and F2 respectively, and vowel duration were measured and statistically analyzed. The results of the acoustic analysis of the EFL learners' AE vowel production (/hVd/ words only) could be compared in detail with similar data collected for native speaker of AE English, made available by the same source as in Chapter 5. The monolingual and bilingual Iranian EFL learners pronounce the AE vowels in the same manner. Both learner groups depart grossly from the L1 norms for American English vowels. In contradistinction to what was found in the perceptual representation in Chapter 5, all the EFL vowels have much shorter durations, while at the same time the EFL vowel space is much smaller than the native space. However, if we abstract from this temporal and spectral reduction – which may well be the same as what native speakers would do, if they were to pronounce the AE vowels as quickly as the EFL learners did, the organisation of the AE vowel system is highly isomorphic with the perceptual representation found in Chapter 5. Spectral differences between members of tense/lax vowel pairs are small (if they can be found at all), while relative vowel durations mirror those of the L1 speakers, and differentiate between spectrally close lax and tense vowels.

7.3. Answering research questions

Using the experimental results collected in Chapters 3 through 6, I will now try to answer the more specific research questions of this dissertation.

7.3.1. Perceptual assimilation of English vowels

Our first research question asked how (early) monolingual Persian and (early) bilingual Azerbaijani/Persian listeners categorize the pure vowels of American English as exemplars of the vowels of their native language(s). A specific hypothesis was formulated: The tense-lax counterparts of English vowels will be assimilated into the same native language categories.

The perceptual assimilation of the AE vowels into Persian and into Azerbaijani was reported in Chapter 4. Since there are 11 AE vowels and fewer vowels in Persian (6) and Azerbaijani (9), no one-to-one assimilation can be expected. The overall results show that only AE /æ/ was uniquely assimilated to its counterpart /æ/, both in Persian and in Azerbaijani. AE tense /u/ was uniquely assimilated into Persian /u/ but not into Azerbaijani /u/. All other peripheral AE vowels were generally matched onto two competing vowels in the participants' L1s, so that poor perceptual discrimination is predicted for these diffuse assimilations. Depending on differences in goodness of fit between the AE vowels that map onto the same vowel category in the participants' L1, severe and lasting (Same Category scenario) or shorter-term (Category Goodness scenario) learning problems are predicted. Table 7.1 summarizes the results.

Table 7.1. Summary of perceptual assimilation of AE vowels to Persian and Azerbaijani, by monolingual and bilingual EFL learners. Legend for AE vowels: bold+underlined = Good token of L1 category, underlined only = Fair token, normal print = Poor token, parenthesized = Unclassified but with clear modal response. CG = Category Goodness scenario, TC = Two Categories scenario, SC = Same Category scenario, UC = Uncategorized Categorized scenario.

Monolinguals			Bilinguals					
AE	Persian	Scenario	AE	Persian	Scenario	AE	Azerbaijani	Scenario
i ɪ	i	CG	**i** ɪ	i	CG	i ɪ	i	CG
e ɛ	e	CG	e ɛ	e	CG	e ɛ	e	SC
æ	æ	TC	æ	æ	TC	æ	æ	TC
(ʌ)	-	UC	(ʌ)	-	UC	(ʌ)	-	UC
ɑ ɔ	ɑ	SC	**ɑ** ɔ	ɑ	CG	**ɑ ɔ**	ɑ	SC
o ʊ	o	SC	o ʊ	o	SC	o (ʊ)	œ	CG
u	u	TC	**u**	u	TC	u	y	TC

Assimilation into Persian is almost identical for monolinguals and bilinguals for all seven vowel pairs in the table. One exception is observed: /ɑ, ɔ/ both assimilate to Persian /ɑ/ but in a CG scenario for the bilinguals against an SC scenario for the monolinguals. Although a difference in learning problem is predicted on account of this, the practical consequences will be negligible, since these vowels have merged into one category for most native speakers of American English (the low back vowel merger). AE /u/ is uniquely assimilated into Persian /u/

by the monolinguals, so that any AE vowel pair with /u/ will be in a Two Category scenario (see Chapter 4) for which no discrimination problems are predicted (i.e., no learning problem).

The hypothesis that the tense and lax counterparts of spectrally close vowel pairs assimilate into the same L1 category, is confirmed for the AE pairs /i, ɪ/, /e, ɛ/ and /o, ʊ/. Because AE /æ/ and /u/ were bi-uniquely assimilated to a single vowel category in the L1, two other tense-lax vowel pairs, i.e., /æ, ɛ/ and /u, ʊ/, qualify as TC contrasts – even though these pairs have often proved problematic for EFL learners with other L1 backgrounds than Persian and Azerbaijani (see, e.g., Wang & Van Heuven, 2006 for Dutch and Mandarin learners of English).

7.3.2. Difference in perceptual assimilation between monolingual and bilingual learners

Our second research question was closely related to the first. It asks if there is a difference the perceptual assimilation found between the monolingual and bilingual listeners, and if so, whether the difference between the groups can be explained by a difference between the vowel systems of Azerbaijani and Persian. We have seen, in the preceding section, that there is hardly any difference in the way the monolinguals and the bilinguals assimilate the AE vowels into Persian. Apparently, the bilinguals have a clearly defined mental image of the vowels of Persian, which they can keep separate from the competing vowels in Azerbaijani, their dominant language. It would indicate that the Persian vowels are available as substitutes for the vowels of English. The results are different, however, when we compare the assimilation of AE vowels to Persian and to Azerbaijani, which comparison can only be made for the bilinguals. Here we may observe a remarkable difference between the two response modes. The difference concerns the assimilation of the AE mid-high back vowels /o, ʊ/ and /u/, which are assimilated to /o/ and u/, respectively in Persian, but to /œ/ and /y/ in Azerbaijani. Although Azerbaijani has rounded back vowels /o/ and /u/, the bilinguals prefer the central counterparts /œ/ and /y/ as the L1 vowel category that fits the AE vowels best. This behaviour indicates that the AE vowels are centralized. Fronting of AE /u/ has been amply documented in the literature ever since the classical study by Peterson and Barney (1952). The lax vowel /ʊ/ is centralized – as is characteristic of lax vowels, while AE /o/ is phonetically a semi-diphthong, which has a position close to /ʊ/ as its starting point (Ladefoged & Disner, 2012: 43-45). It seems obvious that the structure of the vowel system of Azerbaijani, with central alternatives for the back vowels prompts the bilinguals to avail themselves of these options, when the nonnative vowels are perceived as closer to the centralized alternatives. This shows, again, that the bilinguals are able to keep their vowel systems, of Azerbaijani and of Persian, separate, when asked to perform a perceptual assimilation task. The hypothesis formulated in Chapter 2, that the

central(ized) AE vowels /ʌ, ʊ/ may be categorized separately by the bilinguals when instructed to assimilate the AE vowels into the vowels of Azerbaijani, turns out to be incomplete, and is not fully confirmed by our results. Not only is AE /ʊ/ categorized as a central vowel but also the tense vowels /o/ and /u/. The AE mid-low central vowel /ʌ/ is perceived as too remote from any vowel in either Persian or Azerbaijani to be categorized at all.

7.3.3. Relationship between language dominance and perceptual assimilation

We asked what differences there are in relative language dominance between Azerbaijani and Persian in the early bilinguals, and whether these differences in any way reflected by their performance in the perceptual assimilation task. The results of the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al. 2007) revealed that all early bilinguals in our sample of participants acquired Azerbaijani in the parental home before they were exposed, from age four onwards, to Persian at school. The bilinguals also indicated that they were (somewhat) more proficient in primary language use (speaking, listening) in Azerbaijani than in Persian. However, the difference in strength of the two languages differed across individuals, so that they could be ordered along a scale of relative language dominance of Azerbaijani over Persian. We operationalized the quality of the participant's performance on the perceptual assimilation task as the consistency with which they assigned AE vowels to the six vowels of Persian or (in a second run) to the nine vowels of Azerbaijani. Each stimulus vowel was presented on two occasions, so that we could establish whether or not the participant assigned the same stimulus to the same response category on both occasions. The consistency index could then be defined as the proportion of stimulus repetitions assigned to the same category twice out of all stimulus pairs. We then tested following hypothesis: the more dominant Azerbaijani over Persian, the larger the difference in consistency with which the early bilinguals perform the perceptual assimilation task in Azerbaijani mode relative to Persian. The results showed that the bilinguals were less consistent, and took more time, in their task performance for Azerbaijani than for Persian. Choosing from nine response alternatives is more difficult than a choice from six. This, however, does not preclude the possibility to establish the difference in task consistency. The difference in task consistency was then correlated with many potentially useful indicators of relative language dominance of Azerbaijani over Persian (as established from the LEAP-Q answers). In the final analysis one compound indicator turned out to be a significant indicator, predicting 35% of the variance in the difference in task consistency of the perceptual assimilation. This optimal prediction was afforded by a weighted combination of three LEAP-Q difference (Δ Azerbaijani over Persian)

measures, i.e., self-perceived accentedness, accentedness perceived by others, and intensity of exposure to the languages.

Although this result confirms the hypothesis, we are reluctant to advance consistency in perceptual assimilation as a strong and useful correlate of (relative) language dominance in early bilinguals. It seems, in retrospect, that a consistency in perceptual assimilation of foreign sounds to one's native language is not the most suitable way to assess the quality of one's implicit knowledge of the native sound system. More direct tests would be better alternatives, such as perceptual identification of systematically varied synthesized stimuli (as in Van Zanten & Van Heuven, 1984; Van Heuven & Van Houten, 1989; Van Heuven, 2017).

7.3.4. Perceptual representation of AE vowels by monolingual and bilingual EFL learners

In § 7.3.1 we established how monolingual Persian and early bilingual Azerbaijani/ Persian adolescents assimilate the AE vowels into their native language(s). The perceptual assimilation task presupposes no prior exposure to the language the test vowels are taken from. The assimilation patterns found merely suggest how well pairs of sounds in the unknown language will be discriminated by a learner of the unknown language. In Chapter 5, however, we assume that the Iranian EFL learners have developed some awareness of the phonetics of the AE vowel system, after some six years of English lessons. Consequently, our next research question was: "How do English learners with Azerbaijani and/or Persian as their native language(s) perceive the vowels of American English?", or more concretely: "What does the perceptual representation of the AE vowels look like in the mental conception of the Iranian EFL learners; what perceptual targets have they developed, in terms of vowel quality and vowel duration?" We wished to test two hypotheses, i.e., the learners have set up category prototypes (preferred, ideal targets) with different specifications than those entertained by American native listeners, and which lie closer to one another in a way that can be related to the learners' native language(s). Moreover, not only will the distance between adjacent vowels be different (and too small), there will also be more uncertainty about where the spectral and temporal boundaries are between adjacent vowels. Crucially, the non-native listeners will show a rather poorly defined perceptual categories for the AE vowels in which the contrast between adjacent vowel categories is primarily based on a difference in duration.

The perceptual representation of the AE vowels by the two groups of EFL learners was shown in Figures 5.3-4. These figures map out the AE vowel space as it is mentally conceived by the EFL learners. The positions of the phonetic symbols in the F1-by-F2 plots represent the

vowel centroids, i.e., the ideal, prototypical vowel qualities entertained by the learners for each of the 11 AE monophthongs. It can be seen in these figures that the vowels /i, ɪ/ and /u, ʊ/ are virtually in the same spot in the vowel space, indicating that the learners believe (or are not aware) that the vowel colour ('quality') of AE /i/ should differ from /ɪ/, and that /u/ should have a different quality than /ʊ/. If the learners differentiate between the members of the pairs, it must be cued by a difference in vowel duration. The remaining eight vowels are spaced along the outer edge of the vowel space. The AE central vowel /ʌ/ has its target (centroid) along the back edge of the vowel space, in between the back vowels /ɑ, ɔ/, which are also close together.

It is also seen in Figures 5.3-4 that the AE vowels are poorly defined. Each vowel centroid has a large ellipse around it, such that the ellipses belonging to spectrally adjacent vowels overlap to a large extent. The greater the overlap between ellipses, the shallower the perceptual boundaries separating the vowels. Complete overlap of two ellipses means that the listener makes no difference between the vowels, so that the perceptual representations of the vowel colour for the pair is identical.

The two groups of EFL learners select approximately the same vowel durations as appropriate or typical of each of the 11 AE monophthongs. The vowel durations selected by the two learner groups are strongly and significantly correlated (r = .901). The lax vowels of AE English are assigned relatively short durations, whereas longer durations are selected for most of the tense vowels. However, the selected mean duration for tense /e/ and /æ/ were roughly as short as for the lax vowels. The duration of the AE high vowels /i, u/ were the longest of all, whereas their lax counterparts /ɪ, ʊ/ were given the shortest durations. This implies that the tense-lax contrast in the pairs /i, ɪ/ and /u, ʊ/ is largely carried by the difference in duration, and not by a difference in vowel quality.

We conclude, then, that the perceptual vowel representations entertained by the monolingual and bilingual groups are virtually the same, and show properties that are unlikely to be observed in native speakers of American English. The spectral distance between the tense and lax counterparts of the high vowels is very small, and in the case of the front pair /i, ɪ/ close to zero.

7.3.5. Difference in perceptual representation of AE vowels between L1 and L2 listeners

A crucial question raised in this dissertation is how the perception of the vowels of English, summarized in § 7.3.4, differs from the way native speakers of American English perceive their vowels. Native AE listeners will have more sharply defined perceptual

representations of the vowels, with contrasts between adjacent vowels in the vowel space primarily based on spectral differences rather than on duration.

Figure 7.1 shows a summary of the figures in Chapter 5, but averaged over long and short vowel stimuli, thereby focussing on the spectral properties only. The figure shows the centroids only, in separate panels for the two groups of learners and for the American control listeners. The seven long ('tense') vowels in each panel are joined by the complex hull, which sketches the size of the vowel system. The shaded inner quadrilateral joins the four short ('lax') vowels.

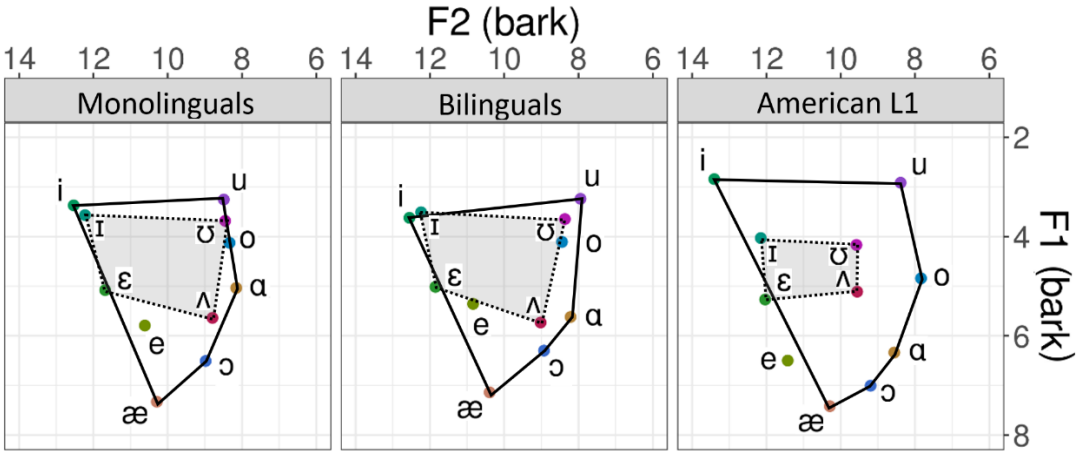


Figure 7.1. Perceptual representation of the vowel quality (location in F1-by-F2 plot in Barks) of the 11 American English monophthongs entertained by three groups of listeners.

Figure 7.1 clearly shows that the members of the high (closed) long-short ('tense-lax') vowel pairs are not (/i-ɪ/) or only marginally (/u-ʊ/) different from one another, with their centroids roughly halfway between the close versus half-close counterparts found for the native AE listeners. The figure also shows that the short AE vowels /ʊ, ʌ/ have central target positions in the vowel space, while the targets of all short (lax) vowels lie close to one another, showing that these vowels form a close-knit subsystem in the mid-central area of the vowel space. The monolingual learners locate the targets for /ʊ, ʌ/ on the back edge of the vowel space, close to the imaginary line that joins /u-ɑ-ɔ/. The bilinguals tend to have the targets for /ʊ, ʌ/ slightly more centralized – as might be predicted from the results of the perceptual assimilation study, which indicated that the bilinguals seem aware of the more centralized position of AE /ʊ, ʌ/ but the effect is small, not convincing and a trend at best. Moreover, the target assumed for tense /u/ is more centralized for the monolingual Persian EFL learners than for the bilinguals, even though Persian has no central high vowel.

The degree of similarity/discrepancy between the vowel systems in Figure 7.1. can be quantified by computing squared Euclidean distances (ED^2, see § 7.3.12 for explanation and

details) between the corresponding vowels in two configurations, after z-normalizing the vowel systems. This shows that the configurations of the two learner groups are highly similar, with a mean ED^2 = .030 Z between the two learner systems, against .179 Z (monolinguals) and .222 Z (bilinguals) between the learner systems and the native speaker configuration. The learner groups, then, differ some 7 times more from the American native speakers than they differ from each other. The effect is significant by a Repeated Measures ANOVA with speaker group as a within-items factor, $F(2, 18) = 11.2$, $p = .004$, $p\eta^2 = .553$. Both learner groups differ from the native group, but do not differ from each other (Bonferroni post-hoc test with $\alpha = .05$).

7.3.6. Native-language interference in perceptual representation

Our next question asked to what extent the differences found in the preceding Section 7.3.5 can be explained by properties of the native language, or native languages, of the nonnative listeners. Our hypothesis was that the lax AE vowels /ʊ/ and /ʌ/ would be better distinguished from their tense counterparts in the perceptual identification by the early bilinguals than by the monolinguals due to the existence of central vowels in Azerbaijani.

The lack of spectral contrast between the tense and lax members was predicted from the absence of such contrasts in the learners' native language. Neither Persian nor Azerbaijani have a tense/lax distinction in their vowel systems. In contradistinction to the lack of spectral contrast, the members of the tense-lax AE vowel pairs are clearly differentiated in the perceptual representation by their duration. The two vowel pairs with the greatest difference in vowel durations selected for tense vs lax members were precisely the pairs /i, ɪ/ and /u, ʊ/.

Inspection of Figure 7.1, however, does not suggest that the extra central vowels in the Azerbaijani inventory have engendered a clearly better, more authentic (i.e., more native-like) conception of the vowel contrasts in American English. This impression is formally supported by the statistics presented in the preceding section, which bear out that there are no significant differences between the vowel pairs in the perceptual representation of the monolinguals and bilinguals, while both perceptual representations differ significantly from that of the American native listeners. This falsifies the hypothesis that the presence of central vowel would be helpful to set up vowel targets for central or centralized AE vowels such as /y/, /ʊ/ and /ʌ/. We must conclude, accordingly, that the early bilinguals relied exclusively on their Persian native vowels as a source of transfer, or, possibly, only used their native Azerbaijani vowel targets in so far as these also occur on Persian. This is to some extent surprizing, since we also found indications in the perceptual assimilation results, that the bilinguals seemed aware that tense AE /u/ was more like Azerbaijani /y/ than /u/.

7.3.7. Acoustic realization of AE vowels by monolingual and bilingual EFL learners

The questions discussed in the preceding sections, and the hypotheses formulated, can be repeated for the production of the AE vowels by the two groups of EFL learners, in comparison with native speakers of American English. These questions and hypotheses were asked and tested in Chapter 6. Figure 7.2 summarizes the findings, averaged over male and female speakers. It shows the positions of the centroids of the American English vowels in the vowel quality space, as in Figure 7.1 above, for monolingual, bilingual and American native speakers of English. In the figure the location of the tense AE vowel /ɔ/ is omitted, since only 10 tokens (out of 90 attempts) were produced (see Table 6.1).

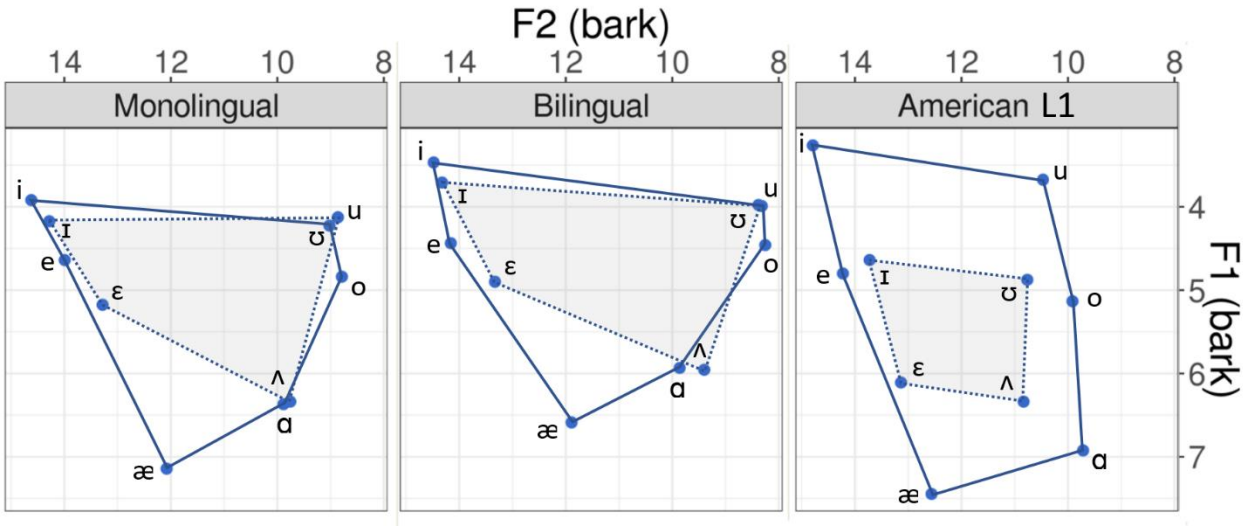


Figure 7.2. Location of centroids (F1 and F2 center frequencies, in Barks) of 10 AE vowels produced by monolingual Persian, bilingual Azerbaijani-Persian and American L1 speakers of English.

The members of the tense-lax pairs /i, ɪ/ and of /u, ʊ/ are produced at virtually the same place in the vowel space, by monolingual and bilingual EFL learners alike. Also, /e/ and /o/ are produced as almost close vowels, and approximate the locations of /i/ and /u/, respectively. The lax mid-vowel /ɛ/ is rather close to /e/ in the learners' speech, so that it will be spectrally distinct from /æ/ but runs the risk of serious confusion with the high(-mid) front vowels. The vowels /ʌ, ɑ/ are produced at the same location by the learners, along the back edge of the vowel space, even though only /ɑ/ should be pronounced as a back vowel.

Tense and lax vowels are produced as long and short, respectively, by the EFL learners (Figure 6.7) but tense /æ/ is too short, while lax /ʌ/ is too long. In fact, /ʌ/ is equally long as tense /ɑ/, so that /ʌ/ and /ɑ/ is fully merged in the EFL vowel production by both learner groups.

The relative configuration of the AE vowel system in the speech of the two learner groups is virtually the same. In absolute values, however, the AE vowels produced by the bilinguals tend to be lower, so that the entire configuration seems somewhat shifted to the right-

hand top corner of the diagram. As explained earlier (Chapter 6, § 6.3.2), this is probably caused by different physiological characteristics of the two learner groups. The younger the speaker, the smaller the cavities in the vocal tract, and the higher the resonances in these cavities (higher formant values). Also, men (and boys after puberty) have larger cavities then women, generating lower resonances. The monolinguals and the American native speakers were equally divided between genders but the girls outnumbered boys by 16 to 11 in the bilingual group. This may well account for the shift of the vowel space towards the right-hand top corner of the diagram. Note, incidentally, that no such difference is seen in the mapping of the perceptual representation of the AE vowels for the same two groups of speakers in Figure 7.1.

A formal quantification of the similarity/discrepancy between the vowel systems in Figure 7.2. can be made by computing squared Euclidean distances (ED^2, see § 7.3.12 for explanation and details) between the corresponding vowels in two configurations, after z-normalizing the vowel systems. Such a quantification shows that the configurations of the two learner groups are highly similar, with a mean ED^2 = .014 Z between the two learner systems, against .191 Z (monolinguals) and .203 Z (bilinguals) between the learner systems and the native speaker configuration. The learner groups differ some 14 times more from the American native speakers than they differ from each other. The effect is significant by a Repeated Measures ANOVA with speaker group as a within-items factor, $F(2, 18) = 13.0$, $p = .001$, $p\eta^2 = .591$. Both learner groups differ significantly from the native group, but do not differ from each other (Bonferroni post-hoc test with $\alpha = .05$).

7.3.8. Difference in acoustic realization of AE vowels by L1 and L2 speakers

Let us now compare the production of the AE vowels by the two learner groups (considered as one group, since differences between monolinguals and bilinguals were negligible) and the American L1 control speakers, whose vowel production is summarized in the rightmost panel of Figure 7.2. The L1 vowels appear to be organised in a system with four degrees of height, which are roughly equidistant along the (bark-transformed) F1 dimension, with high vowels /i, u/, high-mid /e, ɪ, ʊ, o/, low-mid /ɛ, ʌ/ and low /æ, ɑ/. The four lax vowels form a separate subsystem in the realization by the American native speakers, where /ɪ/ and /ʊ/ are not paired with high /i/ and /u/, as in the nonnative realizations, but with mid /e/ and /o/, respectively. The L1 low-mid vowel /ɛ/ is halfway between /e/ and /æ/, thereby optimally avoiding confusion with high-mid /e, ɪ/ and low /æ/, while nonnative /ɛ/ is articulated rather close to /e/ which may well cause perceptual confusion.

L1 /ʌ/ is pronounced as a central vowel and is quite distinct from its nearest neigbor /ɑ/, in height (F1) as well as in backness (F2). This differs markedly from the nonnative articulation of /ʌ/, which is identical (both in quality and in duration) to /ɑ/. Finally, the high(-mid) vowels (/u, ʊ, o/) are clearly centralized by the native speakers, with F2 values higher than for /ɑ/, while these are articulated as back vowels by the learners, with F2 values lower than for /ɑ/.

These observations explain why the overall shapes ('convex hulls') of the vowel spaces differ between the learners and the native speakers. The learners reduce the vowel space along the height (F1) dimension because they realize the tense and lax members of the high vowel pairs the same, right in the middle of the native high /i, u/ and high-mid /ɪ, ʊ/. In the F2 dimension (constriction place), the opposite is seen: here the native speakers centralize the high(-mid) back vowels, so that the vowel space becomes slenderer, while the learners' vowel space shows a broader extension along the F2 dimension.

The tense and lax members of the spectrally close vowel pairs are very well separated in terms of duration in the articulation of the American native speakers (Figure 6.8). Tense and lax vowels are produced as long and short, respectively, by the EFL learners as well (Figure 6.7), but tense /æ/ is too short, while lax /ʌ/ is too long.

7.3.9. Native language interference in the AE vowel production by EFL learners

To what extent can the deviations from the American L1 norms found in the preceding section (§ 7.3.8) be explained as interference, i.e., negative transfer, from the EFL learners' native language(s)? We advanced the hypothesis here that the absence of a tense-lax vowel contrast in both Persian and Azerbaijani, as well as the absence of a length contrast, would be reflected as insufficient contrast between the high(-mid) front vowels /i, ɪ/ for both learner groups, and between the high(-mid) back vowels /u, ʊ/ for Persian EFL learners – but not necessarily for the bilinguals on account of the fact that the latter group may rely on their central vowels to differentiate between back rounded /u/ and centralized /ʊ/. We also predicted that low-mid central /ʌ/ would be merged with low back /ɑ/ by monolingual Persian EFL learners but necessarily by the bilinguals, since the latter have a notion of central vowels in Azerbaijani. However, if a distinction could be found between the members of lax and tense vowel pairs, the contrast would be signalled primarily by a difference in duration rather than by a difference in vowel quality (i.e., formant structure), because differences in duration are less vulnerable to perceptual desensitization after L1 acquisition than are vowel quality differences (Bohn, 1995).

The results found in Chapter 6, and summarized in § 7.3.8, confirm most of these hypotheses. The members of the tense-lax vowel pairs are articulated in almost exactly the

same location in the vowel space, and are only differentiated by duration. This is found for monolinguals and bilinguals alike, so that the second part of the hypothesis, that the bilinguals may have recourse to their central vowels, is falsified.

7.3.10. Predicting incorrect perceptual representation from perceptual assimilation

In the preceding sections, we have summarized the evidence that shows that both groups of Iranian EFL learners, monolinguals and bilinguals alike, entertain an (almost identical) incorrect perceptual representation of the vowels of American English. An important question, raised in Chapter 2, asks how well the problems in perceptual identification of the AE vowels by the two groups of learners can be predicted from the results of the perceptual assimilation task in Chapter 4

. The hypothesis to be tested is that AE vowel pairs that are implicated in the Same Category (SC) scenario, Category Goodness (CG) scenario or Uncategorized (UC, NC, UN, NN), scenarios will be vulnerable to perceptual confusion, while vowel pairs in a Two Category (TC) scenario will be properly distinguished in the learner's perceptual representation. The Perceptual Assimilation Model is not explicit about the relative order of difficulty predicted by the scenarios in between SC and TC, so we will assume that these scenarios yield intermediate discrimination problems for the L2 learners. In Chapter 5 we established the degree of perceptual confusion for all AE 55 vowel pairs, In Table 5.3 for the monolingual Persian participants and in Table 5.4 for the bilingual Azerbaijani/Persian EFL learners. The confusion structure was summarized in Figure 5.10B for the monolinguals and in Figure 5.11B for the bilinguals. In these figures I omitted confusions that occurred in less than 10% of the responses, so that I consider these as not confused in the learners' perceptual representation of the AE vowels. Somewhat arbitrarily, when confusions were found in at least 20% of the responses, I will consider them serious. Any confusions between two vowels that are found in 10% to 20% of the responses I will consider intermediate. When confusions are symmetrically confused, the highest of the two confusion percentages counts. Table 7.2 summarizes the results that are needed to test the hypothesis. In the rows, it lists the three possible predictions based on the PAM results (SC = serious problem, TC = no problem, other = intermediate problem), while rows list the results found in the confusion graphs. In panel A this is done for the monolinguals, in panel B for the bilinguals using predictions based on assimilation to Persian, in panel C on assimilation to Azerbaijani. The table lists all vowel pairs with confusions ≥ 10 explicitly. Minor confusions (< 10% of the responses) are counted but not listed explicitly. Correct predictions by PAM are found along the main diagonal, in

green cells. If the predictions made by PAM are correct well above chance, as would be shown by a significant degree of association (φ) we may consider the hypothesis confirmed.

Table 7.2. Crosstabulation of discrimination difficulty predicted by PAM (in rows) against confusion in perceptual identification (%, in columns) of 11 American English monophthongs by monolingual Persian and by early bilingual Azerbaijani/Persian EFL learners. Frequent confusions (≥ 10%) are spelled out, non-confused vowel pairs (< 10%) are merely counted.

PAM-predicted difficulty	Monolinguals				Bilinguals							
	Persian				Persian				Azerbaijani			
	≥ 20	≥ 10	< 10	Σ	≥ 20	≥ 10	< 10	Σ	≥ 20	≥ 10	< 10	Σ
Serious (SC)	ɒ-ʊ 2 ɑ-ɔ	e-ɪ 1	0	3	i-ɪ 3 u-ʊ ɑ-ɔ	e-ɛ 1	0	4	ɑ-ɔ 1	e-ɛ 1	0	2
Intermediate (other scenarios)	ʌ-æ 2 ʌ-ɑ	ɪ-ɛ 3 ʌ-e ʌ-ɔ	7	12	ʌ-ɑ 2 ʌ-ɔ	ʌ-ɒ 2 ʌ-æ	6	10	i-ɪ 5 ɪ-u ɪ-ʊ u-ʊ ʌ-ɑ ʌ-ɔ	ɪ-u 7 ɪ-ɛ u-ɒ e-ɔ ɒ-ʊ ʌ-ɒ ʌ-æ	28	40
None (TC)	i-ɪ 4 u-ʊ æ-ɛ æ-ɔ	i-ɛ 3 u-ɒ e-ɔ	33	40	ɪ-ʊ 2 æ-ɔ	i-ɛ 6 ɪ-u ɪ-ɛ u-ɒ e-ɔ ɒ-ʊ	33	41	æ-ɔ 1	i-ɛ 1	11	13
Total (Σ)	8	7	40	55	7	9	39	55	7	9	39	55

The association between the PAM predictions and the three categories of confusability is moderate but significant for the monolinguals (φ = .477, $\chi^2(4) = 12.5$, p = .007, one-tailed) and for the bilinguals when they assimilate the AE vowels to Persian (φ = .579, $\chi^2(4) = 18.5$, p < .001, one-tailed). When the bilinguals assimilate the AE vowels to the nine vowels of Azerbaijani, however, the association is no longer significant, (φ = .336, $\chi^2(4) = 6.2$, p = .092, one-tailed), i.e., a trend at best.

In all, the predictions of perceptual confusion of the AE vowels by derived from the perceptual assimilation of the nonnative vowels to the vowels in the learner's native language(s) are not impressively accurate. Yet, the accuracy is better than can be expected from mere chance decisions, as long as the assimilation was done to the vowels of Persian. The accuracy is at chance level only when the bilinguals had to assimilate the AE vowels to the inventory of their first native language, Azerbaijani. Here the large number of unclassified AE vowels placed 40 (out of 55) AE vowel pairs in the intermediate-difficulty category, which prediction failed in 28 of the 40 pairs. For more discussion see § 7.3.11.

7.3.11. Predicting incorrect AE vowel articulation from perceptual assimilation

We may now ask the same question, and advance the same hypotheses as in the previous section for the power of perceptual assimilation of nonnative vowels to the Iranian participants' native vowel categories as predictor of difficulty of creating sufficient contrasts among the nonnative target vowel categories in speech production. I will answer the question in the same way as in the preceding section by cross-tabulating the PAM predictions against the confusion structure obtained for the vowel production results by the monolingual and bilingual EFL learners, as reported in Figure 6.14A-B. Table 7.3 summarizes the results.

Table 7.3. Crosstabulation of discrimination difficulty predicted by PAM (scenarios in rows) against confusions (%, in columns) in automatic classification by LDA of 11 American English monophthongs produced by monolingual Persian and by early bilingual Azerbaijani/Persian EFL learners. See Table 7.2 for more information.

PAM-predicted difficulty	Monolinguals				Bilinguals assimilating to							
	Persian				Persian				Azerbaijani			
	≥ 20	≥ 10	< 10	Σ	≥ 20	≥ 10	< 10	Σ	≥ 20	≥ 10	< 10	Σ
Serious (SC)	ʊ-o 1	e-ɪ 1	1	3	i-ɪ 1	u-ʊ 2 e-ɛ	1	4	0	e-ɛ 1	1	2
Intermediate (other scenarios)	ɪ-ɛ 2 ʌ-ɑ	e-ɛ 1	9	12	ʌ-ɑ 1	0	9	10	i-ɪ 5 u-o ɪ-ɛ ʊ-o ʌ-ɑ	e-ɪ 2 u-ʊ	33	40
None (TC)	i-ɪ 4 i-e u-ʊ u-o	0	36	40	u-o 4 ʊ-o i-e ɪ-ɛ	e-ɪ 1	36	41	i-e 1	0	12	13
Total (Σ)	7	2	46	55	6	3	46	55	6	3	46	55

The association between the PAM predictions and the three categories of confusability is moderate but significant for the monolinguals (φ = .468, $\chi^2(4) = 12.1$, p = .009, one-tailed) and for the bilinguals when they assimilate the AE vowels to Persian (φ = .578, $\chi^2(4) = 18.3$, p < .001, one-tailed). When the bilinguals assimilate the AE vowels to the nine vowels of Azerbaijani, the association is weaker but still significant, (φ = .401, $\chi^2(4) = 8.8$, p = .033, one-tailed). The significant association is mainly due to the success of PAM in predicting which pairs will not be implicated in confusions. This is the great majority of the pairs, most of which involve vowels that are not adjacent to one another in the vowel space, and therefore will not easily be confused. PAM predictions of vowel pairs with strong or moderate confusion are much less successful.

The conclusion follows that PAM predicts the lack of contrastiveness between members of AE vowel pairs in the speech production of the Iranian EFL leaners better than

what can be expected on the basis of chance alone but the prediction is far from perfect. The reason why the PAM hypothesis is not convincingly confirmed may be that the Iranian EFL speakers have learned some of the contrastive properties of the AE vowels so that some errors that might be found for absolute beginners no longer occur and/or that new errors have come up in the student's interlanguage.

7.3.12. Correspondence between perceptual representation and production of AE vowels

The results of Chapter 5 show how our two groups of EFL learners conceive of the American English vowel system, in terms of the location of the vowels in the vowel quality space and their duration. In Chapter 6, I measured the location of the AE vowels in the vowel quality space (in terms if formants F1 and F2) and their duration in the speech production of the same individuals. We now come to the final question that was raised in Chapter 2: To what extent is the organisation of the AE vowels in the perceptual representation established in Chapter 5 reflected in the structure of the vowel system as measured in Chapter 6. In other words: To what extent are perception and production of the AE vowels by the EFL learners correlated? Our hypothesis is that: (non)native deviations in the perceptual representation will correlate with (non)nativeness in the production, assuming that incorrect or deviant perceptual targets are the underlying cause of the (same) deviant articulations in the learner's speech production.

To answer this question, and test the hypothesis, we first of all disregard the vowel /ɔ/ because it has too many missing values (see § 7.3.7), and because the native speakers do not differentiate between this vowel and its nearest neighbour /ɑ/ (the low back vowel merger). In the vowel systems we measured for the perceptual representations, we take the average of the locations of /ɔ/ and /ɑ/. Visual inspection of the topography of the vowel systems in the perceptual representations (Figure 7.1) shows that a point midway the centroid of /ɔ/ and /ɑ/ yields a better match with the /ɑ/ in the production data (Figure 7.2) than either /ɔ/ or /ɑ/ alone. Figure 7.3 shows the location of the remaining 10 vowels, for each of the three speaker groups in separate panels, in the perceptual representation (red squares) and in the speech production (green circles). The tense vowels in each set are connected by a convex hull, as an indication of the global size and shape of the vowel system.

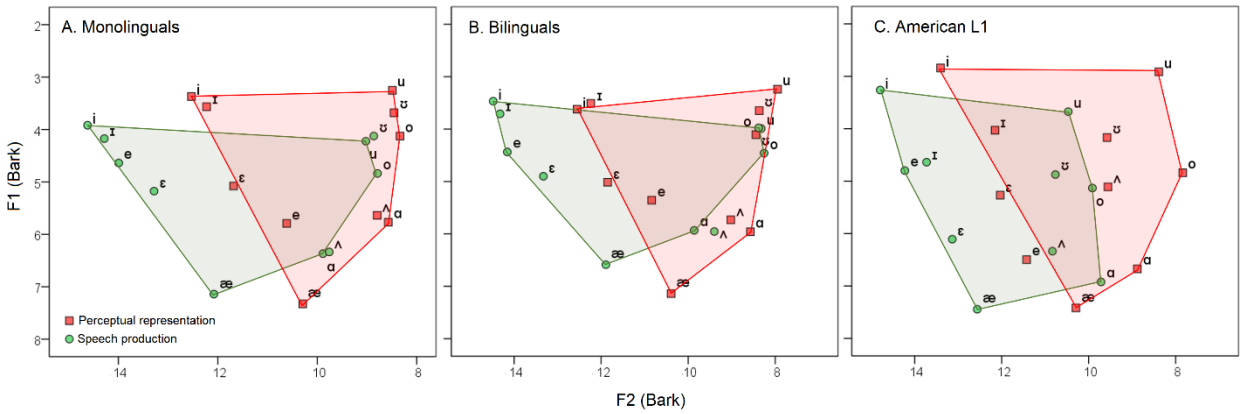


Figure 7.3. Vowel quality (F1 by F2 in Barks) of ten monophthongs of English (excluding /ɔ/) in the perceptual representation and in speech production by 22 Persian monolinguals, 27 Azerbaijani/Persian bilinguals and 20 American native speakers. The tense vowels are joined by a convex hull.

Visual comparison reveals that the production systems (green polygons) are reduced along the vertical dimension relative to the perceptual representations (red polygons), and are shifted to the left by roughly 2 Bark units. These global differences have no linguistic significance and can best be abstracted away from by normalizing the configurations by applying a z-transformation to each set of 10 datapoints along both axes. What then remains are the optimal configurations for evaluating the degree of fit between perceptual representation and production data.[24]

The vowels in the tense-lax pair /i, ɪ/ are very close to one another in both the perceptual representation and in the production data for both learner groups, while they are distant from each other in the native speaker and listener data. The same observation can be made for the members of the /u, ʊ/ pair, and for the /ɑ, ʌ/ pair. The members of these pairs are close together in the vowel space, i.e., insufficiently distinct in the perceptual representation and in the production data, for both learner groups, while they are distant from each other in the native speaker/listener systems.

I will now quantify the overall discrepancy between the vowel systems in the perceptual representation and production data for each of the three groups of participants, and then compare the magnitudes of the overall discrepancies. This is done by computing the squared Euclidean Distance (ED^2) between the corresponding vowel in two vowel charts under comparison, which is defined as:

$$ED^2_{perc,\ prod} = (zF1_{perc} - zF1_{prod})^2 + (zF2_{perc} - zF2_{prod})^2,$$

[24] I assume that a similar normalization is performed by the human listener. The rather simple z-normalization is appropriate here since the topographies have the same orientation in the vowel space and require no rotation.

where the subscripts *perc* and *prod* refer to the same vowel in the two different topographies. We compute ED^2 for each of the ten comparable vowel pairs, and then sum the ten squared distances into $\sum ED^2$, which is the overall measure of discrepancy between the two systems under comparison. We may test the significance of the difference between any two topographies are by a oneway Repeated Measures Analysis of Variance on the ten squared distances per system (with Bonferroni correction for multiple comparisons in the post-hoc analysis of pairwise contrasts).

Table 7.4 tabulates the mean ED^2 values, averaged over nine vowels (not only excluding /ɔ/ but also /e/, which was wrongly used by the listeners in the perception experiment, and was an undue source of error) between the perceptual representation and the speech production data of the three groups of participants (all nine combinations of perceptual representation and speech production.

Table 7.4. Discrepancy (in mean squared Euclidean distance in z-transformed F1-by-F2 (Barks) plane) between perceptual representation and production data of nine American English vowels (excluding /ɔ/ and /e/, see text) measured for three groups of speakers (all nine combinations). The rightmost three columns specify the *F* ratio, *p* value and effect size obtained by a one-way RM-ANOVA on the three conditions listed in the row.

Perceptual representation	Speech production by			RM-ANOVA		
	Monolinguals	Bilinguals	American L1	$F(2, 16)$	p	$p\eta^2$
Monolinguals	.078	.112	.191	2.7	.136	.253
Bilinguals	.074	.096	.223	5.9	.036	.426
American L1	.284	.300	.148	2.5	.148	.240

The overall tendency observed in Table 7.4 is that the perceptual representations and the production locations of the nine test vowels are rather similar, and do not differ much between and across the two EFL learner groups, with mean z-normalized distances of roughly .1 z per vowel. The discrepancy between either of the nonnative representations and the American native production data is roughly twice as large, even though the difference is significant only in the comparison of the perceptual representation acquired by the bilinguals, which does not differ from the produced vowel systems of the monolingual EFL learners (mean ED^2 = .074 z) nor from their own produced system (mean ED^2 = .096 z) while the perceptual representation of both learner groups deviates significantly ($p < .05$ by Bonferroni post-hoc test) from the production of the native vowels.

These findings are support the hypothesis that the perceptual representation of the vowels matches the configuration of the same vowels as produced by the same group of individuals. This finding suggests that the deviations from the native norm in the production of the American English vowels by our learner groups are caused by similar incorrect

perceptual targets. It should be pointed out, however, that a causal relationship cannot be shown by the present correlational data.

7.4. Insights gained from present research

In this dissertation I examined the perceptual representation and the production of the monophthongal vowels of American English for a group of monolingual Persian learners of English as a foreign language, and for a comparable group of early bilingual Azerbaijani/Persian EFL learners, and compared these with parallel data obtained for native speakers and listeners of the target language. I also collected data on the perceptual assimilation of the 11 AE monophthongs to the native vowel systems of the two learner groups, i.e., to the 6 vowels of Persian (for both learner groups) and also to the 9 vowels of Azerbaijani (for the bilinguals), as a way to check whether assimilation patterns might predict or explain problems in the perception or production of the AE vowels by the EFL learners, even after some 6 years of English lessons at school. The early bilinguals considered themselves highly fluent in both Azerbaijani (home language) and Persian (official language at school), as established by the Language Experience and Proficiency Questionnaire (Marian et al., 2007). Differences in language dominance of Azerbaijani over Persian correlated only weakly with the consistency with which the bilinguals performed the perceptual assimilation task in their two languages. The insight gleaned from this is that, counter to our expectation, a perceptual assimilation task is not well suited as a sensitive measure of familiarity with a phonological system.

The approach taken in Chapter 5 to establish the perceptual representation of a vowel system has not been used widely in the literature. In fact, the use of synthesized vowels, varying systematically in phonetic quality in perceptually equal steps, as well as in duration is an innovation in itself. The technique allows us to reconstruct the spectral targets that the listener has set up for the various vowel categories in the target language (the prototypes) as well as the approximate location of the boundaries that separate the categories. Because not only the dimensions of jaw aperture (mouth opening, F1 frequency) and constriction place (backness and lip rounding, F2) were systematically varied but also the duration of the vowel, the ensemble of vowel types also allows the researcher to establish the trading relationships between vowel quality and length. The application of this novel technique revealed that American native listeners hardly use the duration of their vowels as a distinguishing feature, and almost exclusively rely on differences in vowel quality to distinguish between the monophthongs of English. The Iranian learners of English, both monolingual Persian speakers and early bilingual Azerbaijani/Persian speakers turned out to rely on vowel quality only to

make gross distinctions among the English vowels in terms of six spectral categories in a 3 (high, mid, low) by 2 (front, back) arrangement. These are equivalent to the six basic vowel categories of Persian, as well as the six peripheral vowel types of Azerbaijani. Although the additional central vowels in Azerbaijani were shown to be accessible as response categories in the perceptual assimilation task, they did not significantly affect the perceptual representation of the English vowels in a way that differed between the monolinguals and the bilinguals. Any distinctions between tense vs lax vowels within the six basic vowel quality categories were predominantly made on the basis of vowel duration rather than by relying on the smaller shifts in vowel quality that were relied upon by the American native listeners. This result confirms the validity of Bohn's (1995) desensitization hypothesis that after perceptual sensitivity to vowel quality is lost sooner after L1 acquisition than to vowel duration.

In Chapter 6 I studied the acoustic properties (F1, F2 and duration) of the vowels of American English as produced by the same individuals as the EFL learners in Chapter 5. Although the vowel systems produced by the learner groups were smaller in size in absolute frequency measurements, a high degree of congruence was found after normalization between the perceptual representations and the locations of the vowel centroids in the speech production for each of the three groups of speakers (monolinguals, bilinguals, American native speakers). This makes it very likely that the (correct or incorrect) perceptual representation of the AE vowels guides the speaker towards the (correct or incorrect) production of the vowel, rather than the other way around.

The spectral reduction, especially in the vowel-height dimension (jaw aperture), which was observed in the learners' production of the AE vowels was paralleled by a similar reduction in vowel duration. For reasons that are not clear at this moment, the EFL learners produced much smaller durations for the AE vowels than the native speakers did, even though the learners made approximately the same relative differences in vowel duration (e.g., shorter durations for lax vowels than for their tense counterparts) as were found for the native speakers. In the perceptual representation the same mean vowel duration was observed for nonnatives and natives alike; there was no overall shorter vowel duration on the part of the nonnatives. However, the nonnatives exaggerated the vowel durations: the preferred durations were shorter for short/lax vowels and longer for tense vowels than was found with the native speakers. This, again, illustrates that the perceptual representation that we established in Chapter 5 provides a better and more direct window on the vowel system of (native and nonnative) speakers than measuring acoustic properties in their speech production.

Predictions from perceptual assimilation behavior about problematic contrasts among English vowels were only partly successful. In hindsight, one may ask whether there is any point in trying to predict learning problems in foreign language acquisition at all. Since the various models advanced in the literature tend to fail as often as they make successful predictions, we might as well dispense with these models, and establish problem areas by trial and error. Here, measuring the perceptual representation (in terms of perceptual prototypes and boundaries in an ensemble of synthesized exemplars, would be an efficient and effective way of identifying problem areas in the acquisition of the pronunciation of a nonnative language.

7.5. Limitations and recommendations for future research

In the present dissertation, only the perceptual representation and production of the monophthongs of American English were studied. Although I argued that correct, authentic pronunciation of the vowels is the most important source of unintelligibility of nonnative speakers whose native languages has a small vowel inventory, there is no doubt that other types of pronunciation errors will also seriously affect a nonnative's intelligibility in English. In the (near) future we should therefore also study other areas of difference in the phonologies of Azerbaijani, Persian and English, and their effects on EFL speakers' intelligibility and their understanding of English. These areas would include the correct recognition and production of diphthongs, various types of consonants, syllable structures, word stress patterns, sentence stress and melody.

At this stage we do not know how the intelligibility of Persian-accented (or Azerbaijani-accented) English compares to that of other types of nonnative Englishes. The quality of the Iranian learners of English was evaluated without engaging human native listeners, whether American native listeners or international listeners who use English as a lingua franca. The intelligibility of our EFL learners was inferred by comparing the acoustic properties of the speech production and the perceptual representation with those of native speaker/listeners. However, native listeners were never asked to judge the quality of the nonnative speech in opinion testing, nor were they functionally tested on their understanding of the nonnative speech. Future studies should make an effort to gauge the results of the automatic classification of the English vowels (as a shortcut to human listening) against the actual (opinion) scores of human listeners, as was done, for instance, by Wang (2007).

The data of the present studies (Chapters 3 to 6) were collected per speaker on the same day, and for the groups of monolingual and bilingual EFL learners within a very limited timeframe of one month. At the moment the data were collected, the EFL learners had been

exposed to English in a school situation during six years. This organization of the data collection precludes any study of the dynamics of EFL acquisition. Alternatives would have been to enroll multiple groups of EFL learners, at different stages of L2 acquisition, so that we could compare the performance of beginners in their first year, with more experienced learners in later years (i.e., a study in apparent time). Methodologically, it would be better still to follow cohorts of learners, on an individual basis, from day one until the final school exam, so that any developments in the individual students' interlanguage could be studied in detail. This would be a large-scale undertaking which can probably be organized only through cooperation by a team of teachers and researchers in multiple locations in the country. The data collected should be made available to all interested parties, and may contribute to our understanding of the L2 (and L3) acquisition process not only of the phonological aspect of English as a foreign language, but also of other aspects of EFL acquisition, such as listening comprehension, and even reading and writing.

7.6. Some pedagogical implications

One of the most prominent findings in the present dissertation is that the early bilingual Azerbaijani/Persian learners of English as a foreign language appear to have no advantage of early monolingual Persian EFL learners, at least not in the area of pronunciation of the vowels of English, even though the bilinguals have access to three more (central) vowels in their native language than the Persians. Now that we have established that the L3 learners behave largely like their L2-learning peers, there is no reason to advocate separate teaching materials and exercises for monolingual and bilingual learners – at least not in the context of the Iranian educational system.

Before such a "one-size-fits-all" policy can be adopted, however, we must make sure that the absence of any measurable advantage of having access to two different native vowel systems could be due to an artefact. In the classroom the English teacher is the primary model for the pronunciation of the foreign language. Students will not be motivated to pronounce the foreign language more authentically than their teacher does – assuming that the learners are able to discern that the teacher deviates from native speakers of the target language which they may here in recorded materials and exercises. The teachers use Persian as the language of instruction in the classroom. If the teacher is a monolingual Persian, the pronunciation model of English will be Persian-accented English, and even if the teacher is a bilingual Azerbaijani/Persian, they will have been trained by monolingual Persian speakers of English.

The perception and production of AE vowels of the nonnative speakers is poor, even

after six years of English lessons. The survey published by the Educational Language Testing consortium (2017) lists Iran at place 79 in a world-wide ranking of 169 countries. The position is based on the TOEFL test, which measures speaking, listening, writing and reading skills in English as a foreign language. I suggest more practice in discrimination drills and computer-assisted supervised pronunciation teaching. The AE vowels produced by our Iranian EFL learners were correctly classified by the Linear Discriminant Analysis at 57% (Table 6.4), which is our best estimate of what human native listeners would achieve. The correct classification of the same vowels produced by the American native speakers was 90%. Van Heuven and Gooskens (2017: 137) report results for six more groups of English L2 speakers, whose AE vowel tokens were automatically classified using the same native-speaker training set as in the present dissertation. The nine groups in total can be rank ordered as in Table 7.5.

Table 7.5. Correctly identified vowel tokens (%) by Linear Discriminant Analysis trained on American L1 vowel tokens for nine speaker groups.

L1 of speaker	Score	L1 of speaker	Score
1. American English	89.5	6. Mandarin	60.5
2. Danish	81.5	7. Azerbaijani/Persian	56.9
3. Norwegian	78.5	8. Persian	56.5
4. Swedish	78.0	9. Hungarian	51.5
5. Dutch	74.5		

The table suggests that the authenticity of the American English vowels by Iranian monolinguals and bilinguals can be improved by some 20 percentage points.

The results of the present dissertation suggest (but do not prove) that the most viable way to improve the quality of the AE vowel production of the Iranian EFL learners is to start by shaping the perceptual representation of the AE vowel system in terms of the location of the prototypes in the vowel quality space, and the boundaries separating the categories. This can probably be done most effectively by a perceptual training program. L2 learners should first be resensitized to discriminate between small (and normally insignificant) differences in vowel quality. This can be achieved by asking students to imitate, using their own vocal organs, arbitrary vowel qualities produced by a speech synthesizer, while receiving visual and numerical feedback, from a computer system, on the success of their imitations. Once the (adolescent or adult) student has learned to detect and successfully imitate small (subphonemic) differences in vowel quality, the next phase would be to learn to discriminate between tokens of AE vowels in adjacent positions in the AE vowel space. The participants need to be trained to perceive vowel quality differences between tokens that are members of the same vowel category in the L1. They have to learn to perceive differences between allophones of the same

phoneme. This would imply discrimination drills with computer-generated feedback. The student hears a large number of exemplars of two adjacent vowel types, differing in quality, duration and diphthongization, spoken by different native speakers of (American) English, and has to decide for every single token whether it is a token of category A or B, with immediate feedback. In the third and final stage of the training process the learner would hear an exemplar of a target vowel (in context, which they have to imitate. The computer then provides a numerical evaluation of the success of the imitation, provides visual feedback by showing the location of the model vowel and of the learner's imitation in the vowel space on screen, and gives specific instructions to the learner what they should do to approximate the target more closely. Programs that perform these functions have been developed and are available at low cost or no cost at all (see, e.g., Afshar, 2021; Povel & Wansink, 1986; Smakman, 2015, 2020), but are not widely used in secondary education. Learning how to use the software would be an important step for teachers of English as a foreign language towards improving the pronunciation standards in Iranian secondary school teaching of English as a foreign language.

Before embarking on any large-scale implementation of computer-assisted learning environments, however, it would be expedient to carry out strategic research on which aspects of Persian- (or Azerbaijani-) accented English are most detrimental to the learners' intelligibility in English. This type of study would involve recordings of representative spoken utterances (or larger texts) by Iranian speakers of English, with a strong Persian accent. The accented speech can then be selectively improved by replacing all the vowels by corrected exemplars (by electronically exchanging accented and native realizations of the same vowel (produced by a perfect Persian/English bilingual, or by speech resynthesis), and then establish the intelligibility of the utterances (see above). Similar corrections of accented sounds can be applied to only the consonants, or to selected consonant-vowel sequences (i.e., syllables). Also, incorrect word and sentence stresses can be replaced by corrected patterns, which an also be done with other aspects of prosody (sentence melody). This type of research has been done as part of attempts to learn how defective speech produced by deaf persons should be improved (e.g., Maassen & Povel, 1986). More recently, the impact of artificial correction of segments and prosody on intelligibility and perceptual evaluation of nonnative speech has been researched by, e.g., Rognoni (2014) and Capliez (2016a, b). The degree to which intelligibility is improved by each of these artificial corrections tells policy makers and curriculum developers, which aspects of pronunciation teaching would require most attention. The curriculum should then be adjusted to reflect the communicative priorities of vowels, consonants and prosodic structure.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Edinburg University Press.

Afshar, N. (2021). Book review of Carley Paul and Inger M. Mees: American English Phonetics and Pronunciation Practice. *Hungarian Journal of Applied Linguistics*, 21(2), 1–7. http://alkalmazottnyelvtudomany.hu/wordpress/wp-content/uploads/Afshar_rec.docx.pdf

Afshar, N. & Van Heuven, V. J. (2022). Perceptual assimilation of English vowels by monolingual and bilingual learners in Iran. *Argumentum*, 18, 172–191. DOI:10.34103/ARGUMENTUM/2022/9

Aghazada, M., Goncharova, A. & Chernyavskiy, S. I. (2021). Azerbaijani Turks in Iran: from the History to the Modernity. *Voprosi Istorii = Historical Journal*, 5(1), 145–156. DOI: 10.31166/VoprosyIstorii202105Statyi15 УДК 93/94

Ansarin, A. A. (2004). An acoustic analysis of Modern Persian vowels. Proceedings of the 9th Conference on Speech and Computer (SPECOM), September 20-22, 2004, St. Petersburg. https://www.isca-speech.org/archive_open/specom_04/spc4_315.pdf

Aronov, R., McHugh, B. D. & Molnar, T. (2017). A pilot acoustic study of Modern Persian vowels in colloquial speech. *Proceedings of the Linguistic Society of America*, 2, 1–7. DOI: 10.3765/plsa.v2i0.4059

Baese-Berk, M. M. & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, 89, 23–36. DOI: 10.1016/j.jml.2015.10.008

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, 81, 981–1005. DOI: 10.3758/s13414-019-01725-4

Baker, W. & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL – International Review of Applied Linguistics in Language Teaching*, 44, 231–250. DOI: 10.1515/iral.2006.010

Best, C. T., McRoberts, G. W. & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 45–60. DOI: 10.1037//0096-1523.14.3.345

Best, C. T. & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20, 305–330.

Best, C. T. (1994a). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds), *The development of speech perception: The transition from speech sounds to spoken words*. Cambridge, MA: MIT Press, 167–224.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*. Timonium, MD: York Press, 167–200.

Best, C. T., Faber, A. & Levitt, A. G. (1996). Perceptual assimilation of non-native vowel contrasts to the American English vowel system. *Journal of the Acoustical Society of America*, 99(4), 2602. DOI: 10.1121/1.415316

Best, C. T., McRoberts, G. W. & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, 109, 775–794. DOI: 10.1121/1.1332378

Best, C. T.; Tyler, M. D (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Munro, M. J. & Bohn, O.-S. (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 13–34. https://www.academia.edu/24865551/

Bigdeli, N., & Sadeghi, V. (2020). Perceptual evidence for the phonological adaptation of English vowels in Persian sound system. *Scientific Journal of Language Research*, 12, 274–295. DOI: 110.22051/jlr.2019.23487.1629

Birdsong, D. (2014). Dominance and age in bilingualism. *Applied Linguistics*, 35(4), 374–392. DOI: 10.1093/applin/amu031

Boersma, P. & Van Heuven, V. J. (2001). Speak and unSpeak with Praat. *Glot International*, *5*, 341–347. https://www.fon.hum.uva.nl/paul/papers/speakUnspeakPraat_glot2001.pdf

Boersma, P. & Weenink, D. (2019). Praat, a system for doing phonetics by computer. www.praat.org

Bohn, O. S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore, MD: York Press, 279–304.

Bohn, O.-S. & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11, 303–328.

Brown, A. (1991). Functional load and the teaching of pronunciation. I: Brown, A. (Ed.), *Teaching English pronunciation: A book of readings*. London: Routledge, 221–224.

Capliez, M. (2016a). Acquisition and learning of English phonology by French speakers: on the roles of segments and suprasegments. Ph. D. dissertation, Université de Lille III – Charles de Gaulle. https://www.researchgate.net/publication/311431450

Capliez, M. (2016b). Prosody- vs. segment-based teaching: Impact on the perceptual skills of French learners of English. *Language, Interaction and Acquisition*, 7(2), 212–237. DOI: 10.1075/lia.7.2.03cap

Celce-Murcia, M., Brinton, D. M. & Goodwin, J. M. (2010). *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* Cambridge: Cambridge University Press.

Clements, G. N. & Sezer, E. (1982). Vowel and consonant disharmony in Turkish. In H. van der Hulst & N. Smith (eds) *The structure of phonological representations,* Part II, 213–255. Dordrecht: Foris.

Collins, B. S. & Mees, I. M. (1984). *The sounds of English and Dutch*. Leiden: Leiden University Press.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

Dauer, R. (1983). Stress-timing and syllable-timing re-analysed. *Journal of Phonetics*, 11(1), 51–62. DOI: 10.1016/S0095-4470(19)30776-4

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C. & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19(5), 381–385. DOI: 10.1016/j.cub.2009.01.017

DeCasper, A. J. & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208, 1174–1176. DOI: 10.1126/science.7375928

DeCasper, A. J. & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, 9, 133–150. DOI: 10.1016/0163-6383(86)90025-1

Delattre, P. (1965). *Comparing the phonetic features of English, French, German and Spanish: an interim report.* Heidelberg: J. Gross Verlag.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Berlin/New York: Mouton de Gruyter.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, 27, 315–330. DOI: 10.1111/j.1467-1770.1977.tb00124.x

Eckman, F. R. (1985). Some theoretical and pedagogical implications of the markedness differential hypothesis. *Studies in Second Language Acquisition*, 13, 23–41. DOI: 10.1017/S0272263100009700

Educational Testing Service (2017). *Test and score data summary for TOEFL iBT® Tests January 2017 – December 2017 Test Data*. http://www.ets.org.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. LOT Dissertation Series nr. 113. Utrecht: LOT. https://www.lotpublications.nl/Documents/113_fulltext.pdf

Everitt, B. (1993). *Cluster analysis*. London: Arnold.

Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *The European Journal of Neuroscience*, 15(2), 399–402. DOI: 10.1046/j.0953-816x.2001.01874.x

Farran, B. M. M. (2022). The perception and production of American English sounds by Palestinian Arabic adolescents. Doctoral dissertation, University of Pannonia, Veszprém.

Flege, J. E. (1987). The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47–65. DOI: 10.1016/S0095-4470(19)30537-6

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange, (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. Timonium, MD: York Press, 233–277.

Flege, J. E. & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress*. Cambridge University Press, 3–83. DOI: 10.1017/9781108886901.002

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39(3), 399–423. DOI: 10.2307/3588487

Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica*, 38(1–3), 35–40. DOI: 10.1159/000260013

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14(1), 3–28. DOI: 10.1016/S0095-4470(19)30607-2

Frieda, E. M., Walley, A. C., Flege, J. E. & Sloane, M. E. (1999). Adults' perception of native and nonnative vowels: implications for the perceptual magnet effect. *Perception and Psychophysics*, 61(3), 561–577. DOI: 10.3758/bf03211973

Gafos, A. & Dye, A. (2011). Vowel harmony: Opaque and transparent vowels. In M. van Oostendorp, C. J. Ewen, E. Hume & K. Rice (Eds.), *The Blackwell Companion to Phonology*, 2164–2189. Blackwell Publishing.

Ghaffarvand Mokari, P., Gholi Famian, A. & Ghafoori, N. (2013). An acoustic study of production and perception of English vowels by Azeri English learners. *Journal of Basic and Applied Scientific Research*, 3(7), 84–91. https://www.researchgate.net/publication/263963222

Ghaffarvand Mokari, P. & Werner, S. (2016). An acoustic description of spectral and temporal characteristics of Azerbaijani vowels. *Poznan Studies in Contemporary Linguistics*, 52(3), 503–518. DOI: 10.1515/psicl-2016-0019

Ghaffarvand Mokari, P. & Werner, S. (2017). Perceptual assimilation predicts acquisition of foreign language sounds: The case of Azerbaijani learners' production and perception of Standard Southern British English vowels. *Lingua*, 185, 181–195. DOI: 10.1016/j.lingua.2016.07.008

Ghaffarvand Mokari, P., Werner, S. & Talebi, A. (2017). An acoustic description of Farsi vowels produced by native speakers of Tehrani dialect. *The Phonetician*, 114, 6–23. https://www.researchgate.net/profile/Payam-Ghaffarvand-Mokari/publication/319261459

Guion, S. G., Flege, J. E., Akahane-Yamada, R. & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, 107(5), 2711–2724. DOI: 10.1121/1.428657

Heeringa, W. & Van de Velde, H. (2018). Visible Vowels: A Tool for the visualization of vowel variation. Proceedings CLARIN Annual Conference 2018, 8-10 October, Pisa, CLARIN ERIC, 120–123. https://office.clarin.eu/v/CE-2018-1292-CLARIN2018_ConferenceProceedings.pdf

Hepper, P.G., Scott, D. & Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, 11(3), 147–153. DOI: 10.1080/02646839308403210

Hillenbrand, J., Getty, L. A., Clark, M. J. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97(5), 3099–3111. DOI: 10.1121/1.411872

Hillenbrand, J. M., Clark, M. J. & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America*, 108(6), 3013-3022. DOI: 10.1121/1.1323463

Hockett, C. F. (1955). *A manual of phonology*. Baltimore, MD: Waverly Press.

Hosmer, D. W. & Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.

House, A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, 33(9), 1174–1178. DOI: 10.1121/1.1908941

Howlader, M. R. (2010). Teaching English pronunciation in countries where English is a second language: Bangladesh perspective. *ASA University Review*, 4, 233–244. www.asaub.edu.bd/data/asaubreview/v4n2sl20.pdf

Iverson, P. & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, 99(2), 1130–1140. DOI: 10.1121/1.415234

Jenkins, J. (1998). Which pronunciation norms and models for English as an international language? *ELT Journal*, 52, 119–126.

Jenkins, J. (2000). *The phonology of English as an international language: New models, new norms, new goals.* Oxford: Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirical researched pronunciation syllabus for English as an international language. *Applied Linguistics* 23(1), 83–103.

Jenkins, J. (2002). *World Englishes*. London: Routledge.

Jeong, H. & Thorén, B. (2018). Intelligibility of the alveolar [s] replacing the initial interdental /θ/ in English words. *Proceedings of Fonetik 2018: The XXXth Swedish Phonetics Conference*, 39–42.

Jones, T. (2019). A corpus phonetic study of contemporary Persian vowels in casual speech. *University of Pennsylvania Working Papers in Linguistics*, 25, article 15. https://repository.upenn.edu/pwpl/vol25/iss1/15

Karimzad, F., Shosted, R. K. & Peymani, P. (2015). The sound system of Tabrizi Azeri. Unpublished Manuscript. University of Illinois at Urbana-Champaign, & University of Tabriz

Kaushanskaya, M. (2020). The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition*, 23(5), 945–950. DOI: 10.1017/S1366728919000038

Klecka, W. R. (1980). *Discriminant analysis*. Beverly Hills & London: Sage.

Kuhl, P. K. (1988). Auditory perception and the evolution of speech. *Human Evolution*, 3(1–2), 19–43. DOI: 10.1007/BF02436589

Kuhl, P. K. (1991). Human adults and human infants show a 'perceptual magnetic effect' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, 93–107. DOI: 10.3758/BF03212211

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences USA*, 97, 11850–11857. DOI: 10.1073/pnas.97.22.1185

Kuhl, P., Conboy, B., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and Native Language Magnet Theory Expanded (NLM-E). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 979–1000. DOI: 10.1098/rstb.2007.2154

Kuhl, P. K. & Iverson, P. (1995). Linguistic experience and the 'perceptual magnet effect'. In W. Strange & J. J. Jenkins (Eds.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, 121–154.

Labov, W. Ash, S. & Boberg, Ch. (2006). *The Atlas of North American English*. Berlin: Mouton-de Gruyter, 187–208.

Ladefoged, P. & Disner, S. F. (2012). *Vowels and consonants*. Chichester: Wiley.

Ladefoged, P. & Johnson, K. (2011). *A course in Phonetics*. Boston: Wadsworth.

Lado, R. (1957). *Languages across cultures. Applied Linguistics for language teachers*. Ann Arbor, MI: University of Michigan Press.

Lazerte, S. (2021). How a soccer team is amplifying the voices of Azerbaijani-Turks in Iran. *The Caspian Post*, 6 August 2021). https://caspianpost.com/en/post/perspectives/how-a-soccer-team-is-amplifying-the-voices-of-azerbaijani-turks-in-iran

Leather, J., & James, A. (1991). The Acquisition of Second Language Speech. *Studies in Second Language Acquisition*, 13(3), 305–341.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5(3), 253–264.

Lehman, W. P. & Heffner, R.-M. S. (1940). Notes on the length of vowels (III). *American Speech*, 15(4), 377–380. https://www.jstor.org/stable/487070

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. DOI: 10.1037/h0020279

Lisker, L. & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384–422.

Lisker, L. & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In B. Hála, M. Romportl & P. Janota (eds.), *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague: Academia, 563–567.

Luk, G. & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25(5), 605–621. DOI: 10.1080/20445911.2013.795574

Maassen, B. & Povel, D.-J. (1985). The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech. *Journal of the Acoustical Society of America*, 78(3), 877–886. DOI: 10.1121/1.392918

Majidi, M.-R., & Ternes, E. (1999). Persian (Farsi). *Handbook of the International Phonetic Association*. Cambridge University Press, 124–125. https://archive.org/details/rosettaproject_pes_phon-2

Mannell, R., Cox, F. & Harrington, J. (2009) *An introduction to Phonetics and Phonology*, Macquarie University. http://clas.mq.edu.au/speech/phonetics/

Marian, V., Blumenfeld, H. K. & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, 50(4), 940–967. DOI: 10.1044/1092-4388(2007/067)

Melnik-Leroy, G.A., Turnbull, R. & Peperkamp, S. (2021). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/–/y/ contrast. *Second Language Research*. DOI: 10.1177/0267658320988061 (publication ahead of print)

Mirahadi, S. S., Kamran, F., Mansuri, B., Tohidast, S. A., Rashtbari, K., Panahgholi, E. & Tagipur, A. (2018). Investigation of the formant structure of Persian vowels in Persian-Azari bilingual adults. *Archives of Rehabilitation*, 19(2), 142–148. DOI: 10.32598/rj.19.2.142

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25(3), 481–521. DOI: 10.2307/3586981

Munro, M. J. & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34, 520–531. DOI: 10.1016/j.system.2006.09.004

Nagle, C. L. & Baese-Berk, M. M. (2021). State of the scholarship. Advancing the state of the art in L2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*, 1–26. DOI: 10.1017/S0272263121000371 (publication ahead of print)

Peivasti, S. M. (2012). An acoustic analysis of Azerbaijani vowels in Tabrizi dialect. *Journal of Basic and Applied Scientific Research*, 2, 7181–7184. https://www.textroad.com/pdf/JBASR/J.%20Basic.%20Appl.%20Sci.%20Res.,%202(7)7181-7184,%202012.pdf

Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24(2), 175–184. DOI: 10.1121/1.1906875

Peterson, G. E. & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32(6), 693–703. DOI: 10.1121/1.1908183

Pillai, S. & Delavari, H. (2012). The production of English monophthong vowels by Iranian EFL learners. *Poznań Studies in Contemporary Linguistics*, 48, 473–493. DOI: 10.1515/psicl-2012-0022

Piske, T., MacKay, I. & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics,* 29, 191–215. DOI: 10.1006/jpho.2001.0134

Povel, D.-J. & Wansink, J. (1986), A computer-controlled vowel corrector for the hearing impaired. *Journal of Speech and Hearing Research,* 29(1), 99–105. DOI: 10.1044/jshr.2901.99

Querleu, D., Lefebvre, C., Titran, M., Renard, X., Morillion, M. & Crepin, G. (1984). Réactivité du nouveau né de moins de deux heures de vie à la voix maternelle [Reaction of the newborn infant less than 2 hours after birth to the maternal voice]. *Journal of Gynecology & Obstetrics and Biolology of Reproduction*, 13, 125–134.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L. & Crépin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 29, 191–212. DOI: 10.1016/0028-2243(88)90030-5

Raphael, L. & Bell-Berti, F. (1975). Tongue musculature and the feature of tension in English vowels. *Phonetica*, 32(1), 61–73. DOI: 10.1159/000259686

Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192. DOI: 10.1146/annurev.neuro.27.070203.144230

Rognoni, L. (2014). The phonetic realization of narrow focus in English L1 and L2. Data from production and perception. Ph. D. Dissertation, Università degli Studi di Padova. https://www.academia.edu/6761475/

Sadeghi, V. & Bigdeli, N. (2018). انطباق واجی واکه های انگلیسی با فارسی در چارچوب انگارۀ شنیداری [Phonological adaptation of English vowel with Persian vowel based on perceptual assimilation]. *Journal of Research in Linguistics*, 10, 43–60. https://www.sid.ir/FileServer/JF/35413971803.pdf

Sağın-Şimşek, Ç. & König, W. (2012). Receptive multilingualism and language understanding: intelligibility of Azerbaijani to Turkish speakers. *International Journal of Bilingualism*, 16(3), 315–331. DOI: 10.1177/1367006911426449

Sakai, M. (2016). (Dis)connecting perception and production: Training adult native speakers of Spanish on the English /i/-/ɪ/ distinction. Ph. D. dissertation, Georgetown University. https://repository.library.georgetown.edu/bitstream/handle/10822/1042879/Sakai_georgetown_0076D_13518.pdf

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1), 187–224. DOI: 10.1017/S0142716417000418

Salehi, M. & Neysani, A. (2017). Receptive intelligibility of Turkish to Iranian-Azerbaijani speakers. *Cogent Education*, 4(1), 1–15. DOI: 10.1080/2331186X.2017.1326653

Sebastián-Gallés, N. & Baus, C. (2005). On the relationship between perception and production in L2 categories. In: A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones*. New York: Erlbaum, 279–292.

Sebastián-Gallés, N. & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, 72(2), 111–123. DOI: 10.1016/s0010-0277(99)00024-4

Shaffer, B. (2021). *Iran is more than Persia. Ethnic politics in the Islamic Republic*. Washington, DC: PDD Press. https://www.fdd.org/wp-content/uploads/2021/04/fdd-monograph-iran- is-more-than-persia.pdf

Sisinni, B. & Grimaldi, M. (2009). Second language discrimination vowel contrasts by adults speakers with a five vowel system. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association*, Brighton, UK, 1679–1682.

Smakman, D. (2015). *Accent Building. A British English pronunciation course for speakers of Dutch*. Leiden: Leiden University Press.

Smakman, D. (2020). *Clear English pronunciation*. Abingdon, New York: Routledge.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research, Methods, Instruments, & Computers*, 31(1), 137–149. DOI: 10.3758/BF03207704

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K. & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, 26(4), 311–344. DOI: 10.1006/jpho.1998.0078

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A. & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America*, 109(4), 1691–1704. DOI: 10.1121/1.1353594

Strange, W., Bohn, O.-S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, 115(4), 1791–1807. DOI: 10.1121/1.1687832

Strange, W., Bohn, O.-S., Nishi, K. & Trent, S. A. (2004). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, 118(3), 1751–1762. DOI: 10.1121/1.1992688

Thorin, J., Sadakata, M., Desain, P. & McQueen, J. M. (2018). Perception and production in interaction during non-native speech category learning. *Journal of the Acoustical Society of America*, 144(1), 92–103. DOI: 10.1121/1.5044415

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, 88(1), 97–100. DOI: 10.1121/1.399849

Tremblay, M. C., & Sabourin, L. (2012). Comparing the behavioral discrimination abilities of monolinguals, bilinguals and multilinguals. *Journal of the Acoustical Society of America*, 132(5), 3465–3474. DOI: 10.1121/1.4756955

Trofimovich, P. & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28(1), 1–30. DOI: 10.1017/S0272263106060013

Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen & M. H. Sørensen, (Eds.), *A sound approach to language matters – In honor of Ocke-Schwen Bohn*. Dept. of English, School of Communication & Culture, Aarhus University, 607–630. https://www.researchgate.net/publication/333124848

Gerland, P., Andreev, K., Bassarsky, L. Bay, G., Cruz Castanheira, H., Gaigbe-Togbe, V., Gu, D., Hertog, S., Li, N., Ribeiro, I., Spoorenberg, T., Ueffing, P., Wheldon, M. & Zeifman, L. (2019). *World Population Prospects 2019. Volume I: Comprehensive Tables*. New York: United Nations, Department of Economic and Social Affairs, Population Division. https://population.un.org/wpp/Publications/Files/WPP2019_Volume-I_Comprehensive-Tables.pdf

Van Heuven, V. J. (1986). Some acoustic characteristics and perceptual consequences of foreign accent in Dutch spoken by Turkish immigrant workers. In J. van Oosten & J. F. Snapper (Eds.), *Dutch Linguistics at Berkeley, papers presented at the Dutch Linguistics Colloquium held at the University of California, Berkeley on November 9th, 1985*, Berkeley: The Dutch Studies Program, U.C. Berkeley, 67–84. https://openaccess.leidenuniv.nl/handle/1887/2831

Van Heuven, V. J. (2016). An acoustic characterisation of English vowels produced by American, Dutch, Chinese and Hungarian speakers. *Hungarian Journal of Applied Linguistics*, 16(2), 1-20. DOI: 10.18460/ANY.2016.2.004

Van Heuven, V. J. (2017). Perception of English and Dutch checked vowels by early and late bilinguals. Towards a new measure of language dominance. In: S. E. Pfenninger & J. Navracsics (Eds.) *Future research directions for Applied Linguistics* (Second Language Acquisition 109). Bristol, Buffalo, Toronto: Multilingual Matters. 73–98. DOI: 10.21832/9781783097135-006

Van Heuven, V. J. (2022). Stress deaf and color blind. Native language background and perceptual categories. In: J. M. van de Weijer (Ed.), *Representing phonological detail*. Volume 2. Berlin, New York: Mouton de Gruyter. (in press). https://www.researchgate.net/publication/362223574

Van Heuven, V. J., Afshar, N. & Disner, S. F. (2020). Mapping perceptual vowel spaces in native and foreign language: Persian learners of English compared with American native speakers. In: Sz. Bátyi & Zs. Lengyel (Eds.), *Kétnyelvüség: Magyar és nem Magyar*

kontextus. Tanulmányok Navracsics Judit köszöntésére/Bilingualism: Hungarian and non-Hungarian context. Studies in honor of Judit Navracsics. Veszprém: Pannon Egyetem, 113–130. https://www.researchgate.net/publication/347252430 https://scholarlypublications.universiteitleiden.nl/access/item%3A2728595/view

Van Heuven, V. J. & Farran, B. M. M. (2022). The role of duration in the articulation and mental representation of American English vowels by Palestinian Arabic learners of English. Paper presented at the 23rd Summer School of Psycholinguistics, 23-25 May 2022, University of Pannonia, Veszprém, Hungary. https://www.researchgate.net/publication/361570474

Van Heuven, V. J. & Gooskens, C. (2017). An acoustic analysis of English vowels produced by speakers of seven different native-language backgrounds. In: M. Wieling, M. Kroon, G. van Noort & G. Bouma (Eds.), *From semantics to dialectometry*. *Festschrift in honour of John Nerbonne*. London: College Publications, 137–147. http://hdl.handle.net/1887/57185

Van Heuven, V. J. & Kirsner, R. S. (2004). Phonetic or phonological contrasts in Dutch boundary tones? In L. Cornips & J. Doetjes, (Eds.), *Linguistics in the Netherlands 2004*. Amsterdam/Philadelphia: John Benjamins, 102–113. DOI: 10.1075/avt.21.13heu

Van Heuven, V. J. & Van Houten, J. E. (1989). Vowel labelling consistency as a measure of familiarity with the phonetic code of a language or dialect. In: M. E. H. Schouten & P. Th. van Reenen (Eds.) *New methods in dialectology*. Dordrecht: Foris, 121–130.

Van Zanten E. & Van Heuven V. J. (1984). The Indonesian vowels as pronounced and perceived by Toba Batak, Sundanese and Javanese speakers. *Contributions of the Royal Institute of Anthropology [Bijdragen tot de Taal-, Land- en Volkenkunde]*, 140(4), 497–521. DOI: 10.1163/22134379-90003411

Walker, R. (2001). Pronunciation for international intelligibility. *English Teaching Professional*, 21, 1–7. https://englishglobalcom.files.wordpress.com/2013/12/pronunciation-for-international-intelligibility2.pdf

Walker, R. 2012. Vowel harmony in Optimality Theory. *Language and Linguistics Compass* 6(9), 575–592. DOI: 10.1002/lnc3.340

Wang, H. & Van Heuven, V. J. (2006). Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. In J. M. van de Weijer & B. Los (eds.) *Linguistics in the Netherlands 2006*. Amsterdam: John Benjamins, 237–248. DOI: 10.1075/avt.23.23wan

Wang, H., & Van Heuven, V. J. (2014). Is a shared interlanguage beneficial? Mutual intelligibility of American, Dutch and Mandarin speakers of English. In R. van den Doel & L. Rupp, (Eds.), *Pronunciation Matters. Accents of English in The Netherlands and elsewhere*. VU University Press, 175–194. http://hdl.handle.net/1887/3146383

Wang, H. & Van Heuven, V. J. (2015). The Interlanguage Speech Intelligibility Benefit as bias toward native-language phonology. *i-Perception* 6(6): 1–13. DOI: 10.1177/2041669515613661

Wang, H. & Van Heuven, V. J. (2018). Relative contribution of vowel quality and duration to native language identification in foreign-accented English. *Proceedings of the 2nd International Conference on Cryptography, Security and Privacy, 13–16 March 2018, Guiyang, China*. New York: Association for Computing Machinery, 16–20. DOI: 10.1145/3199478.3199507.

Wang, X., & Chen, J. (2019). English speakers' perception of Mandarin consonants: The effect of phonetic distances and L2 experience. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences,*

Melbourne, Australia. Australasian Speech Science and Technology Association, 250–254. https://assta.org/proceedings/ICPhS2019/papers/ICPhS_299.pdf

Wang, X., & Chen, J. (2020). The Acquisition of Mandarin consonants by English learners: The relationship between perception and production. *Languages*, *5*(2), 1–15. DOI: 10.3390/languages5020020

Watkins, K. E., Strafella, A. P. & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8), 989–994. DOI: 10.1016/s0028-3932(02)00316-0

Wilson, S. M., Saygin, A. E. P., Sereno, M. I. & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701–702. DOI: 10.1038/nn1263.

Appendices

Appendix A4.1. Acoustic details of American English vowel tokens used in the PAM test.

The vowel tokens were segmented from the first clear glottal pulse following the [h]-noise until the near-silence of the [d] following the vowel. Mean formant frequencies F1 and F2 were measured (using the Burg algorithm in Praat with three formants in a 0 to 3 KHz frequency band), from the vowel onset until either F1 or F2 showed the beginning of a transition to the following [d]. For the semi-diphthongs /e/ and /o/, mean F1 and F2 were computed for the first 50% of the vowel duration only. The results are shown in Table A4.1.

Table A4.1. Stimulus analysis of 22 vowel tokens used in PAM test. F1, F2 (Hz) and duration (ms) of eleven vowel tokens produced by two male American native speakers in /h..d/ context.

Vowel	Speaker 1			Speaker 2		
	F1	F2	Dur	F1	F2	Dur
i	300	2378	220	296	2163	184
ɪ	399	1989	113	403	1778	136
e	382	2157	293	436	1890	240
ɛ	506	1887	184	524	1688	128
æ	649	1730	273	667	1637	216
u	352	1184	245	353	1212	205
ʊ	433	1339	209	465	1201	160
o	442	1047	289	424	1213	205
ɔ	550	967	278	589	873	268
ɑ	713	1150	249	620	992	258
ʌ	595	1154	172	523	1186	150

The formants were psychophysically scaled to Bark units so that equal distances in the F1-F2-plane correspond to equal auditory distances in vowel quality using the formula in Traunmüller (1990). The resulting vowel plot is shown in Figure A1. Note the similarity between the acoustic vowel chart here and the articulatory IPA diagram in Figure 1C.

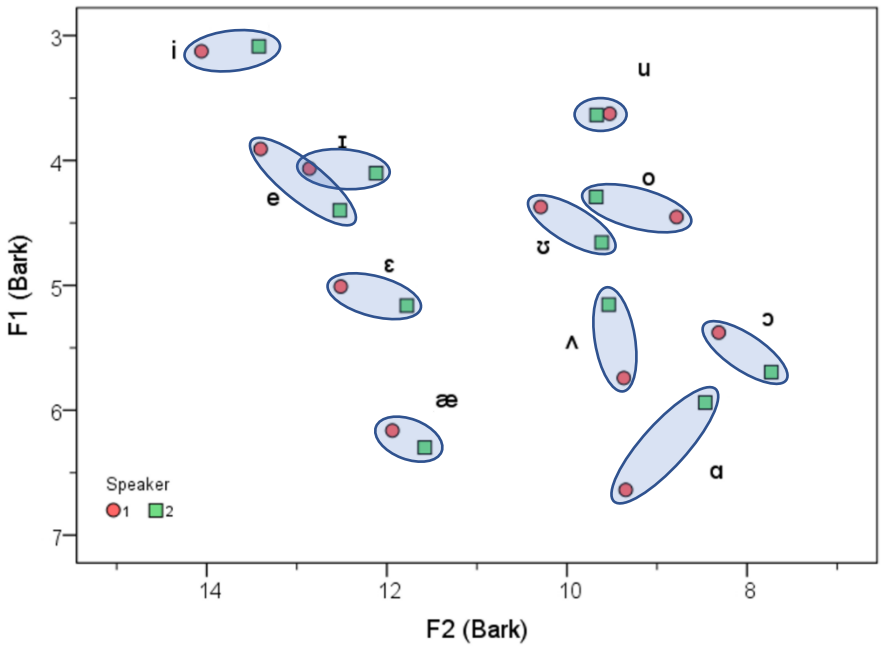


Figure A4.1. Vowel tokens of Table 1 plotted in the acoustic vowel space defined by F1 (top to bottom, Barks) and F2 (right to left, Barks). Ellipses were drawn by hand and have no theoretical status.

A plot of the vowel durations is shown in Figure A4.2.

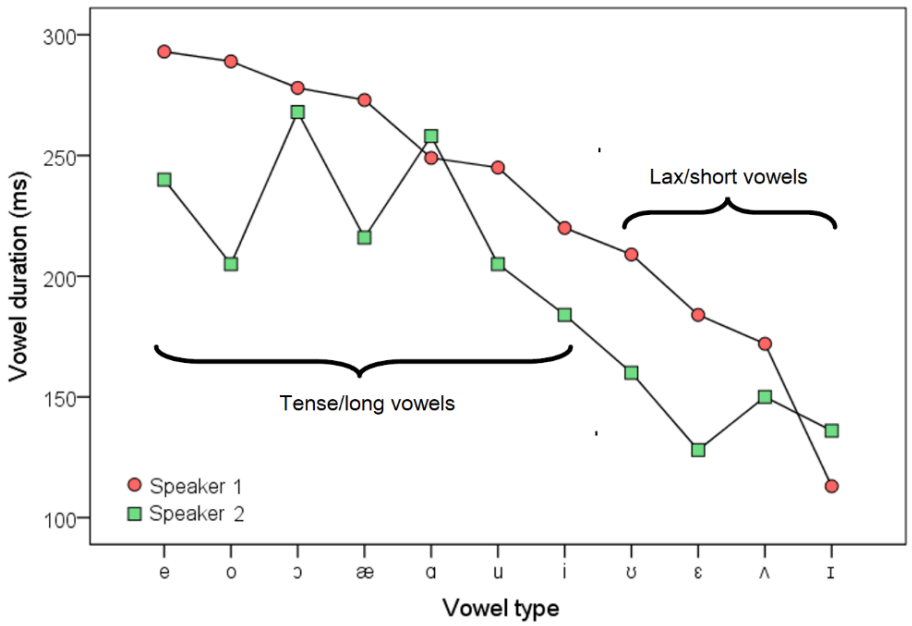


Figure A4.2. Duration (ms) of 11 American English monophthongs produced by two male native speakers. Vowel types are plotted from left to right in descending order of the duration realized by speaker 1.

There is a split in duration between the seven phonetically tense and long vowels and the four lax and short vowels. There are two vowel pairs in Figure A1 the members of which are spectrally close to one another. These are the pairs /ɪ, e/ and /ʊ, o/. These members will nevertheless be distinct by the difference in duration, and by the slight change in quality in the time course of the semi-diphthongs /e/ and /o/.

Appendix A4.2. Praat MFC script for PAM test.

"ooTextFile"
"ExperimentMFC 7"
blankWhilePlaying? <no>
stimuliAreSounds? <yes>
stimulusFileNameHead = "Sounds/"
stimulusFileNameTail = ".wav"
stimulusCarrierBefore = ""
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 0.5 seconds
stimulusMedialSilenceDuration = 0
stimulusFinalSilenceDuration = 1.5 seconds
numberOfDifferentStimuli = 22
"had_01" ""
"hawed_01" ""
"hayed_01" ""
"head_01" ""
"heed_01" ""
"hid_01" ""
"hod_01" ""
"hoed_01" ""
"hood_01" ""
"hud_01" ""
"whod_01" ""
"had_11" ""
"hawed_11" ""
"hayed_11" ""
"head_11" ""
"heed_11" ""
"hid_11" ""
"hod_11" ""
"hoed_11" ""
"hood_11" ""
"hud_11" ""
"whod_11" ""

numberOfReplicationsPerStimulus = 2
breakAfterEvery = 0
randomize = <PermuteBalancedNoDoublets>
startText = "This is a listening experiment.

After hearing a sound, choose the vowel that is most similar to what you heard.

Click to start."

runText = "Choose the vowel that you heard."
pauseText = "You can have a short break if you like. Click to proceed."
endText = "The experiment has finished."
maximumNumberOfReplays = 0
replayButton = 0 0 0 0 "" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""

responsesAreSounds? <no> "" "" "" "" 0 0 0

Left column for Azerbaijani

numberOfDifferentResponses = 12
0.15 0.25 0.8 0.9 "il" 40 "" "i"
0.35 0.45 0.8 0.9 "ül" 40 "" "ü"
0.55 0.65 0.8 0.9 "ıl" 40 "" "ı"
0.35 0.45 0.4 0.5 "" 40 "" ""
0.55 0.65 0.4 0.5 "" 40 "" ""
0.75 0.85 0.4 0.5 "al" 40 "" "a"
0.75 0.85 0.8 0.9 "ul" 40 "" "u"
0.15 0.25 0.6 0.7 "el" 40 "" "e"
0.35 0.45 0.6 0.7 "öl" 40 "" "ö"
0.55 0.65 0.6 0.7 "" 40 "" ""
0.75 0.85 0.6 0.7 "ol" 40 "" "0"
0.15 0.25 0.4 0.5 "əl" 40 "" "æ"

Right column for Persian

numberOfDifferentResponses = 6
0.2 0.4 0.8 0.9 "\FIpictures/sir.png" 40 "" "i"
0.2 0.4 0.6 0.7 "\FIpictures/ser.png" 40 "" "e"
0.2 0.4 0.4 0.5 "\FIpictures/sær.png" 40 "" "æ"
0.6 0.8 0.8 0.9 "\FIpictures/sur.png" 40 "" "u"
0.6 0.8 0.6 0.7 "\FIpictures/sor.png" 40 "" "o"
0.6 0.8 0.4 0.5 "\FIpictures/sar.png" 40 "" "a"

numberOfGoodnessCategories = 5
0.25 0.35 0.10 0.20 "1 (poor)" 24 ""
0.35 0.45 0.10 0.20 "2" 24 ""
0.45 0.55 0.10 0.20 "3" 24 ""
0.55 0.65 0.10 0.20 "4" 24 ""
0.65 0.75 0.10 0.20 "5 (good)" 24 ""

Appendix A5.

Table A5.1. Biographic data on 20 native listeners of American English listeners who participated in the control experiment described in Van Heuven et al. (2020) and in Chapter 5.

#	Code#	Name	Gender	Age	L1	Place of birth	Where raised
1.	101	Bob1220	M	78	English	Chicago IL	Chicago IL
2.	102	Cathy1220	F	68	English	Culver City	Culver City CA
3.	103	Elaine1220	F	76	English	Cleveland	Cleveland, OH
4.	105	Phil1220	M	60	English	San Francisco	San Francisco CA
5.	110	Anjalee122	F	19	English	Palo Alto CA	San Jose CA
6.	111	Carson122	M	21	English	Connecticut	New Jersey
7.	113	Mac121	M	21	English	Wyoming	Wyoming
8.	121	Gabriel123	M	20	English	Florida	various US states
9.	122	Jaron123	M	19	English	California	California
10.	123	Jordan123	F	21	English	New Jersey	New Mexico
11.	126	Aamani128	F	19	English	California	California
12.	127	Aamuro129	M	18	English	Colorado	Colorado
13.	128	Andrew128	M	21	English	Washington	Washington DC
14.	129	Charli128	M	21	English	Washington	Washington DC
15.	130	Daniel128	F	21	English	Virginia	Zurich
16.	131	Jenna129	M	22	English	USA	USA
17.	132	Sawyer129	M	20	English	North CA	North CA
18.	133	Selden129	M	21	English	New York	New York
19.	134	Sophia129	F	19	English	South CA	South CA
20.	135	Tim129	M	22	English	South CA	Calgary, Canada

Figure A5.2. Oscillograms (amplitude against time), spectrograms and formant tracks (frequency against of time, gray shades represent intensity) of selected synthesized /mVf/ stimuli for Chapter 5. Graphs in left column have vowel duration of 200 ms, in right column 300 ms. Stimulus numbers refer to Figure 5.1.

Upper row:	Oscillogram of stimulus 1.1.
Second row from top:	Spectrogram with formant tracks overlaid (F1..F5) for stimulus 1.1 ([i]-like)
Second row from bottom:	Spectrogram with formant tracks overlaid (F1..F5) for stimulus 1.9 ([u]-like)
Bottom row:	Spectrogram with formant tracks overlaid (F1..F5) for stimulus 7.5 ([a]-like)

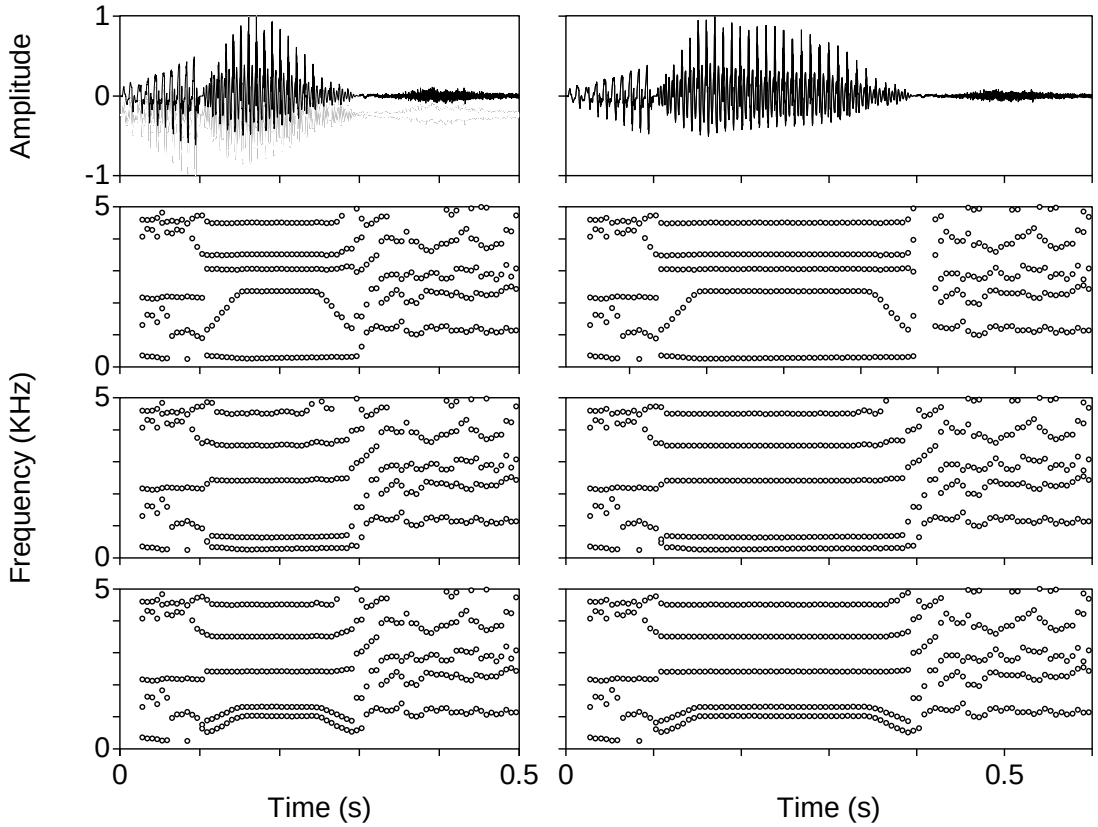


Table A5.3. Number of responses given to 86 synthesized vowel stimuli by 20 native listeners of American English broken down by vowel duration (short = 200 ms, long = 300 ms) and by formant frequencies (F1 and F2 in Hz). The eleven response categories are listed in the second row. The modal response per stimulus vowel is in the green-shaded cell, and bolded if chosen in > 50%. Tied modes are indicated in yellow-shaded cells. The casting vote then goes to the response category that maximizes contiguous areas in the vowel space.

Spectrum		Vowel duration = 200 ms											Vowel duration = 300 ms										
F1	F2	i	ɪ	e	ɛ	æ	ɑ	ʌ	ɔ	o	ʊ	u	i	ɪ	e	ɛ	æ	ɑ	ʌ	ɔ	o	ʊ	u
237	2357	**19**			1								**19**			1							
237	2031	**19**	1										**19**									1	
237	1746	2	6		1		1	2				8	4	3		2			1	1		2	7
237	1497	1	1								6	**12**							3		1	4	**12**
237	1278	1			1			2		1	1	**14**	1						1			3	**15**
237	1086	1									1	**18**							1			2	**17**
237	915										2	**18**											**20**
237	764									2	1	**17**									1	1	**18**
237	628										2	**18**										1	**19**
339	2357	**16**	3	1									**19**		1								
339	2031	**10**	9									1	**12**	5	1	2							
339	1746	3	7			1					4	5	4	7		1			1		1	3	3
339	1497		9								5	6	1	6			2		4			2	5
339	1278	2						1			6	**11**										7	**13**
339	1086										6	**14**							1			1	**18**
339	915							1			4	**15**									1	3	**16**
339	764											**20**										1	**19**
339	628										2	**18**						1			1	2	**16**
447	2031		**10**		10								1	6	2	**11**							
447	1746		**12**		6	1						1	1	6	1	**10**				1		1	
447	1497		5	1	2			4			5	3		4	2	3			4			4	3
447	1278		2		3		1	3	1		8	2		2		1			5			**11**	1
447	1086							4			**11**	5			1		1		6		1	**11**	
447	915						1	3		9	6	1						1	5		9	4	1
447	764							1	3	**11**	5							1	2	3	**13**	1	
565	2031		3	2	**15**									1	3	**15**						1	
565	1746		1	1	**18**										4	**15**						1	
565	1497				**15**		1	4						1	3	9	2			1		4	
565	1278				3		1	7	1	2	5	1		1		1	2	2	9			5	
565	1086				1		3	7		2	7						1	2	7	2	1	7	
565	915						4	6	3	4	3							8	1	3	8		
565	764						7	2	3	6	1	1					1	8	2	5	4		
694	1746			4	**12**	3			1						2	7	9			2			
694	1497			1	**12**	3		1	3				1		3	3	**11**		1	1			
694	1278				2	2	1	9		1	5			1	2	1	6	2	3	5			
694	1086						5	**10**	1	1	3						1	6	1	**11**	1		
694	915					3	6	1	9		1						3	**10**		5	2		
838	1497			4	3	**13**									2		**15**			3			
838	1278					**9**	3	1	6	1					1		**10**	7		2			
838	1086					7	3		**9**	1							**10**	3		7			
998	1497			1		**16**			3						2		**12**			6			
998	1278		1	2		**10**	3		4						2		9	4		5			
998	1086			2		5	4		8	1					1		9	4		6			
All		74	70	19	105	73	44	69	55	42	100	209	82	43	33	82	104	59	58	69	44	83	203

Table A5.4. Number of responses given to 86 synthesized vowel stimuli by 21 monolingual Persian learners of American English broken down by vowel duration (short = 200 ms, long = 300 ms) and by formant frequencies (F1 and F2 in Hz). In grey-shades cells the (multi-)modal response is less than 25%. For more information see Table A5.2

Spectrum		Vowel duration = 200 ms											Vowel duration = 300 ms										
F1	F2	i	ɪ	e	ɛ	æ	ɑ	ʌ	ɔ	o	ʊ	u	i	ɪ	e	ɛ	æ	ɑ	ʌ	ɔ	o	ʊ	u
237	2357	7	**11**	1							2		**12**	7					1				1
237	2031	6	**14**								1		**13**	7	1								
237	1746	6	10	1						2		2	6	9		3							3
237	1497		10	1	1			2		1	3	3	5	3	2	1					1	3	6
237	1278	2	1	1	2		1	1		2	6	5	2	3			1	1		1	5	4	4
237	1086		1					1		4	7	8	1	1						1	5	5	8
237	915			1			2	1		5	3	9	1					3	1	1	2	3	10
237	764						4		1		9	7							2	1	2	2	**14**
237	628		1	1						2	**13**	4						1	1		3	4	**12**
339	2357	7	**12**		1						1		**13**	7		1							
339	2031	6	**12**		1					1		1	9	10								1	1
339	1746	4	**15**								1	1	**12**	7		1							1
339	1497	3	8	1	2					1	4	2	3	5		2		1	1			4	5
339	1278	3	2		2					1	**11**	2	4	2				2			3	3	7
339	1086		1	1			1			6	10	2			1	1		2			2	6	9
339	915						3	2			10	6		1	1			1	1		2	8	7
339	764				1		1	1		2	**13**	3									9	6	6
339	628						1			6	**12**	2						1			6	7	7
447	2031		**13**		8								4	5	2	7			1	1		1	
447	1746	3	6	3	6			1			1	1	5	3	3	8					1	1	
447	1497	2	7		7			1		1		3	5	3	1	8					1	3	
447	1278		4	1	6			4		2	3	1	1	2	3	4			1		3	3	4
447	1086		2	2			1	3	1	4	4	4		1		1		1			**12**	2	4
447	915		1				1	2	1	8	4	4						2	2	1	6	5	5
447	764	1	1	1			4	1		10	3					1		1	1	2	8	6	2
565	2031	1	6	3	10	1								2	5	**12**						2	
565	1746	1	3	4	**11**						1	1	4	1	3	10		1			1		1
565	1497			4	9	4	1			1	1	1	1	4	2	10	3		1				
565	1278	1	2	2	4	6	1			2	2	1	1	2	1	3	4		2	1	2	3	2
565	1086		1	1	1	2	3	5	2	2	3	1	1		2		6	1	4	1	3	3	
565	915		1	2			4	4	2	5	3				2			2	3	7	6		1
565	764					1	2	6	6	1	5				1			5	5	6	2	1	1
694	1746	1	2	2	8	8							3	2	5	2	9						
694	1497		1	5	4	9				1		1	1		5	2	**11**		1	1			
694	1278		1	4	3	8		1	2	2					1	1	**12**			4	1	2	
694	1086			3		5	2	4	5		1	1			4	1	1	3	6	4	1	1	
694	915			2		3	1	6	4	2	3				2		3	4	7	5			
838	1497			4	1	**14**		1	1						3		**13**			4		1	
838	1278		1	3		10	1	1	3	2					3		**13**		1	3	1		
838	1086			5		8	2	1	3	1	1				2		7	2	2	4		4	
998	1497			2		**16**	1		2						4	1	**12**		1	2		1	
998	1278			5		**11**	1	2	1		1				2		**15**	2	2				
998	1086			3		**11**	1	1	3	1	1			1	1		**12**	2	1	2	1		1
All		54	150	69	88	117	39	52	37	78	143	76	107	88	62	80	122	38	48	52	89	95	122

Table A5.5. Number of responses given to 86 synthesized vowel stimuli by 27 early bilingual Azerbaijani/Persian learners of American English broken down by vowel duration (short = 200 ms, long = 300 ms) and by formant frequencies (F1 and F2 in Hz). In grey-shades cells the (multi-)modal response is less than 25%. For more information see Table A5.2

Spectrum		Vowel duration = 200 ms											Vowel duration = 300 ms										
F1	F2	i	ɪ	e	ɛ	æ	ɑ	ʌ	ɔ	o	ʊ	u	i	ɪ	e	ɛ	æ	ɑ	ʌ	ɔ	o	ʊ	u
237	2357	7	**20**										13	11						1	2		
237	2031	7	**19**								1		13	10	2	2							
237	1746	2	**19**	1	1				1		2	1	8	11		4						1	3
237	1497	1	10	1	4	1		2		2	4	2	3	6	1	1			3		6	2	5
237	1278		2	2	3			4		4	8	4		3	2	1	1		2	1	6	7	4
237	1086				1	1		4		1	10	10				3			5		7	7	5
237	915		1	1	1			3		5	9	7	1					2	2		4	7	11
237	764						2	4			10	11							1		6	5	**15**
237	628		1				1			5	**14**	6							1		5	6	**15**
339	2357	5	**20**	1	1								**14**	11		2							
339	2031	5	**20**		1						1		11	12	1	2					1		
339	1746	3	**14**	2	3		1	2			1	1	12	9	1	3			2				
339	1497	3	8	2	2			2		1	6	3	4	7	1	1			3	1	3	6	1
339	1278	1	6	1	1			7		2	8	1	2	3		1			4	1	6	4	6
339	1086		2	1	3			3	1	2	13	2		1	1				3	1	4	7	10
339	915						1	1		3	**16**	6		1			1		2		4	6	13
339	764							2		2	**14**	9				1		1	1		6	7	11
339	628		1				2	1		6	11	6						1	1	1	6	6	12
447	2031	3	13		10							1	11	7	3	5			1				
447	1746	2	6	3	12					1	1	2	9	2		13					1	1	1
447	1497	1	11	3	3	1		2		1	3	2	8	3	1	7		1	1		2	4	
447	1278		10		5		1	2		2	4	3	1	6	3	3			2	2	3	3	4
447	1086		3		2		2	2	1	5	9	3		2		1	1	2	3		9	3	6
447	915		1	1	1		3	3		7	7	4							3	1	12	4	7
447	764		1	1			5	2	2	9	3	4				1		3	3	3	11	3	3
565	2031	1	6	1	**17**				1		1		1	1	3	**20**					1	1	
565	1746	2	4	2	**17**	1						1	5		3	**15**	1				2		1
565	1497	1	4	2	11	5	2				1	1		4	4	13	5				1		
565	1278	1	6	2	3	3		5	1	3	1	2	1	3	2	3	5		4	1	5	1	2
565	1086		2	1	1	2	2	9	3	4	2	1			1		2	5	3	7	6	2	1
565	915		1	1		2	3	8	6	3	3		1		1		2	5	9	5	2	1	1
565	764					2	3	7	7	2	6							8	2	10	4	2	1
694	1746		1	2	10	12		1	1					1	6	6	13		1				
694	1497	1		5	4	**14**			2		1				2	5	**17**		1	2			
694	1278		2	2	5	12		4	2					1	2	1	**15**			7	1		
694	1086					7	3	8	6		2	1				1	2	5	8	6	3	2	
694	915			2		4	3	10	6		2						5	2	8	10		2	
838	1497	2		1		**22**			1	1						1	**19**		1	4	1	1	
838	1278		2			**16**	1	4	2	2				1	2		**15**		2	5	1		1
838	1086					9	4	8	4	1	1						9	6	5	5		2	
998	1497					**20**	1	2	4								**21**		2	4			
998	1278		1	3		**15**	2	3	2		1				1		**15**	3	2	3	1	2	
998	1086					11	3	5	6	1	1				2		9	7	4	3	2		
All		48	217	44	122	160	45	120	59	75	177	94	118	116	45	116	158	51	95	84	134	105	139

Appendix A6

Table A6.1. List of stimulus vowels in common key words (A) and in /hV(r)d/ carrier (B).

	A	B
1.	Now say **need** again.	Now say **heed** again.
2.	Now say **kid** again.	Now say **hid** again.
3.	Now say **played** again.	Now say **hayed** again.
4.	Now say **bed** again.	Now say **head** again.
5.	Now say **bad** again.	Now say **had** again.
6.	Now say **rude** again.	Now say **who'd** again.
7.	Now say **good** again.	Now say **hood** again.
8.	Now say **road** again.	Now say **hoed** again.
9.	Now say **sawed** again.	Now say **hawed** again.
10.	Now say **god** again.	Now say **hod** again.
11.	Now say **card** again.	Now say **hard** again.
12.	Now say **mud** again.	Now say **hud** again.
13.	Now say **word** again.	Now say **heard** again.
14.	Now say **slide** again.	Now say **hide** again.
15.	Now say **employed** again.	Now say **hoyed** again.
16.	Now say **loud** again.	Now say **how'd** again.
17.	Now say **beard** again.	Now say **here'd** again.
18.	Now say **toured** again.	Now say **hoored** again.
19.	Now say **shared** again.	Now say **haired** again.
20.	Now say **bed** again.	Now say **head** again.

Appendix A6.2. Descriptive statistics of vowel production data

Table A.6.2A. Mean, standard deviation, range, minimum and maximum for F1, F2 (Hz) and duration (ms) of AE vowels produced by monolingual Persian EFL learners, aggregated and broken down by gender of speaker. Targets were everyday /CVd/ keywords. Maximum *N* = 21.

Monolingual Persians			American English vowel in /CVd/ context										
Gender	Statistic	Parm	i	ɪ	e	ɛ	æ	ɑ	ɔ	o	ʊ	u	ʌ
Male	Mean	F1	383	347	433	450	645	562	603	437	366	382	617
		F2	2386	2135	1951	1855	1570	1210	1309	1035	1132	1128	1046
		Dur	124	83	143	108	137	144	155	136	114	122	137
	SD	F1	63	35	47	47	42	35	45	46	46	35	46
		F2	206	197	214	157	121	139	96	175	233	159	113
		Dur	41	31	39	23	20	20	12	33	25	30	19
	Min	F1	289	273	369	388	581	517	544	388	298	321	537
		F2	1994	1803	1557	1691	1457	1008	1168	876	869	901	872
		Dur	70	54	95	69	116	106	139	93	91	95	115
	Max	F1	513	389	506	508	703	625	652	538	476	435	688
		F2	2729	2409	2338	2228	1872	1508	1380	1374	1579	1396	1189
		Dur	221	136	228	150	172	174	165	186	160	181	179
	Range	F1	224	116	137	120	122	108	108	150	178	114	151
		F2	735	606	781	537	415	500	212	498	710	495	317
		Dur	151	82	133	81	56	68	26	93	69	86	64
	N		10	10	10	9	10	10	4	9	10	9	10
Female	Mean	F1	451	464	540	572	825	718	646	542	457	482	737
		F2	2833	2617	2375	2249	1957	1420	1280	1142	1195	1257	1235
		Dur	137	101	150	140	160	154	161	154	131	141	152
	SD	F1	50	92	55	48	70	53		64	57	64	97
		F2	183	336	235	130	134	122		106	187	269	81
		Dur	42	32	40	19	21	27		45	22	37	16
	Min	F1	385	359	447	471	741	642	646	492	332	369	503
		F2	2433	1872	2061	2042	1772	1233	1280	886	918	862	1059
		Dur	91	66	76	101	133	106	161	107	90	82	126
	Max	F1	552	637	661	668	929	821	646	687	560	619	894
		F2	3092	2984	2795	2501	2275	1587	1280	1283	1524	1742	1361
		Dur	223	180	231	164	204	190	161	247	175	198	176
	Range	F1	167	278	214	197	188	179		195	228	250	391
		F2	659	1112	734	459	503	354		397	606	880	302
		Dur	132	114	155	63	71	84		140	85	116	50
	N		11	11	11	11	11	11	1	10	11	11	11
All	Mean	F1	419	409	489	517	739	643	612	492	414	437	680
		F2	2620	2388	2173	2071	1772	1320	1303	1091	1165	1199	1145
		Dur	131	92	147	126	149	149	156	146	122	133	145
	SD	F1	65	91	74	78	108	91	43	77	69	73	97
		F2	297	367	309	244	234	167	84	149	207	230	135
		Dur	41	32	39	26	23	24	11	40	25	35	19
	Range	F1	263	364	292	280	348	304	108	299	262	298	391
		F2	1098	1181	1238	810	818	579	212	498	710	880	489
		Dur	153	126	155	95	88	84	26	154	85	116	64
	N		21	21	21	20	21	21	5	19	21	20	21

Table A.6.2B. Mean, standard deviation, range, minimum and maximum for F1, F2 (Hz) and duration (ms) of AE vowels produced by monolingual/Persian EFL learners, aggregated and broken down by gender of speaker. Targets were /hVd/ words. Maximum *N* = 21.

Monolingual Persians			American English vowel in /hVd/ context										
Gender	Statistic	Parm	i	ɪ	e	ɛ	æ	ɑ	ɔ	o	ʊ	u	ʌ
Male	Mean	F1	333	356	429	466	675	618	614	438	368	379	622
		F2	2307	2151	2065	1900	1559	1151	1192	923	1029	1041	1147
		Dur	110	95	148	104	105	116	158	147	107	110	117
	SD	F1	29	40	36	61	54	43	56	34	38	40	42
		F2	182	178	160	199	115	87	110	130	150	158	99
		Dur	36	24	38	17	21	16	5	42	23	33	21
	Min	F1	280	306	383	377	601	542	550	382	314	311	556
		F2	2035	1898	1800	1705	1380	1047	1102	763	801	862	978
		Dur	71	62	93	80	71	99	153	99	73	55	82
	Max	F1	370	420	501	559	774	673	656	508	459	432	673
		F2	2599	2455	2306	2364	1799	1372	1314	1202	1291	1283	1306
		Dur	195	130	207	138	134	150	162	217	141	161	150
	Range	F1	90	114	118	182	173	131	106	126	145	121	117
		F2	564	557	506	659	419	325	212	439	490	421	328
		Dur	124	68	114	58	63	51	9	118	68	106	68
	N		10	10	10	10	10	10	3	8	10	9	10
Female	Mean	F1	447	472	507	588	886	735	771	533	452	462	723
		F2	2786	2665	2568	2249	1905	1330	1469	1141	1083	1124	1287
		Dur	125	107	173	119	128	130	124	156	131	148	127
	SD	F1	74	101	43	81	89	48		94	74	63	93
		F2	172	343	142	205	102	96		137	131	158	145
		Dur	45	34	34	34	28	30		36	25	23	21
	Min	F1	351	384	434	479	742	667	771	427	339	353	491
		F2	2422	1852	2340	1972	1800	1227	1469	953	899	922	1003
		Dur	63	56	135	65	97	79	124	121	94	116	104
	Max	F1	576	663	559	757	1019	831	771	751	644	601	818
		F2	3017	3033	2820	2574	2137	1516	1469	1361	1350	1351	1561
		Dur	204	186	227	185	175	165	124	234	171	194	173
	Range	F1	225	279	125	278	277	164		324	305	248	327
		F2	595	1181	480	602	337	289		408	451	429	558
		Dur	141	130	92	120	78	86		113	77	78	69
	N		10	11	10	11	11	11	1	10	11	10	11
All	Mean	F1	390	417	468	530	786	679	653	491	412	423	675
		F2	2547	2420	2317	2083	1740	1245	1261	1044	1057	1085	1220
		Dur	117	101	160	112	117	123	150	152	120	130	122
	SD	F1	80	97	56	94	130	75	91	87	72	67	88
		F2	300	377	297	266	206	128	165	171	140	160	142
		Dur	40	29	37	28	27	25	18	38	26	34	21
	Range	F1	296	357	176	380	418	289	221	369	330	290	327
		F2	982	1181	1020	869	757	469	367	598	549	489	583
		Dur	141	130	134	120	104	86	38	135	98	139	91
	N		20	21	20	21	21	21	4	18	21	19	21

Table A.6.2C. Mean, standard deviation, range, minimum and maximum for F1, F2 (Hz) and duration (ms) of AE vowels produced by monolingual Persian EFL learners, aggregated and broken down by gender of speaker. Targets were everyday /CVd/ keywords. Maximum *N* = 24.

Azeri/Persian bilinguals			American English vowel in /CVd/ context										
Gender	Statistic	Parm	i	ɪ	e	ɛ	æ	ɑ	ɔ	o	ʊ	u	ʌ
Male	Mean	F1	393	325	447	436	614	568		452	362	374	607
		F2	2351	2148	2019	1949	1619	1186		988	1010	997	1108
		Dur	145	82	118	109	125	134		158	107	137	128
	SD	F1	46	42	52	38	73	90		65	40	40	89
		F2	206	221	183	182	199	155		162	238	132	288
		Dur	43	20	20	23	15	34		25	26	32	23
	Min	F1	325	271	363	389	504	383		338	297	286	445
		F2	1951	1777	1695	1666	1359	954		805	786	789	758
		Dur	84	53	93	71	107	94		115	64	95	90
	Max	F1	485	387	523	497	730	692		567	427	435	783
		F2	2619	2428	2319	2220	1907	1434		1380	1615	1243	1823
		Dur	245	117	148	148	153	205		203	160	190	169
	Range	F1	160	116	160	108	226	309		229	130	149	338
		F2	668	651	624	554	548	480		575	829	454	1065
		Dur	161	64	55	77	46	111		88	96	95	79
	N		11	11	11	11	11	11		11	11	11	11
Female	Mean	F1	376	418	452	493	711	623		465	432	424	651
		F2	2834	2552	2426	2308	1898	1305		1054	1061	1157	1187
		Dur	150	94	134	126	149	138		154	122	137	150
	SD	F1	68	41	35	53	146	73		46	48	47	151
		F2	140	114	163	225	144	148		99	145	123	186
		Dur	26	29	48	35	24	30		22	30	23	32
	Min	F1	290	361	402	418	486	502		401	323	374	432
		F2	2661	2332	2119	2008	1637	1092		920	865	929	946
		Dur	118	61	72	65	115	99		121	76	107	105
	Max	F1	498	499	524	589	942	721		531	501	515	945
		F2	3215	2721	2661	2902	2136	1576		1247	1290	1315	1559
		Dur	190	155	237	195	186	184		205	178	173	204
	Range	F1	208	138	122	171	456	219		130	178	141	513
		F2	554	389	542	894	499	484		327	425	386	613
		Dur	72	94	165	130	71	85		84	102	66	99
	N		13	13	13	13	13	13		12	13	12	13
All	Mean	F1	383	375	450	467	667	598		459	400	400	631
		F2	2613	2367	2240	2144	1771	1251		1022	1038	1080	1151
		Dur	148	89	127	119	138	136		156	115	137	140
	SD	F1	59	62	42	54	126	84		55	56	50	126
		F2	298	265	267	273	219	160		134	190	149	236
		Dur	34	25	38	31	23	31		23	29	27	30
	Range	F1	208	228	161	200	456	338		229	204	229	513
		F2	1264	944	966	1236	777	622		575	829	526	1065
		Dur	161	102	165	130	79	111		90	114	95	114
	N		24	24	24	24	24	24		23	24	23	24

Table A.6.2D. Mean, standard deviation, range, minimum and maximum for F1, F2 (Hz) and duration (ms) of AE vowels produced by bilingual Azerbaijani/Persian EFL learners, aggregated and broken down by gender of speaker. Targets were /hVd/ words. Maximum *N* = 24.

Azeri/Persian bilinguals			American English vowel in /hVd/ context										
Gender	Statistic	Parm	i	ɪ	e	ɛ	æ	ɑ	ɔ	o	ʊ	u	ʌ
Male	Mean	F1	315	327	420	463	675	605		429	374	374	581
		F2	2189	2213	2119	1937	1595	1224		912	952	926	1062
		Dur	130	95	156	107	105	120		161	119	140	121
	SD	F1	32	43	39	65	83	125		54	43	39	110
		F2	220	230	199	212	175	236		107	162	169	159
		Dur	41	23	29	19	19	33		29	40	30	27
	Min	F1	275	270	346	355	539	404		324	322	310	371
		F2	1830	1871	1748	1508	1349	1008		790	758	719	805
		Dur	87	59	105	70	83	95		131	58	102	81
	Max	F1	372	392	483	602	807	757		502	462	447	708
		F2	2456	2517	2328	2255	1829	1790		1094	1270	1218	1245
		Dur	214	132	210	132	139	186		226	178	179	174
	Range	F1	97	122	137	247	268	353		178	140	137	337
		F2	626	646	580	747	480	782		304	512	499	440
		Dur	127	73	105	62	56	91		95	120	77	93
	N		11	11	11	11	11	10		11	11	11	11
Female	Mean	F1	367	403	467	527	736	636	650	466	415	419	663
		F2	2751	2623	2592	2233	1774	1253	1150	995	996	1003	1230
		Dur	137	101	171	112	121	121	195	159	122	129	131
	SD	F1	45	37	52	72	133	122		37	48	41	92
		F2	110	167	116	169	170	142		76	177	101	174
		Dur	26	37	38	30	23	33		24	36	34	28
	Min	F1	287	363	403	427	533	439		416	295	359	444
		F2	2579	2143	2412	1967	1519	1109		882	811	852	920
		Dur	102	50	131	66	93	76		107	65	82	91
	Max	F1	428	506	574	661	911	784		520	471	486	759
		F2	2934	2780	2770	2516	1970	1633		1142	1425	1188	1584
		Dur	190	177	266	164	167	164		195	192	180	181
	Range	F1	141	143	171	234	378	345		104	176	127	315
		F2	355	637	358	549	451	524		260	614	336	664
		Dur	88	127	135	98	74	88		88	127	98	90
	N		13	13	13	13	13	13	1	12	13	12	13
All	Mean	F1	343	368	446	498	708	623	650	448	396	397	625
		F2	2494	2435	2375	2098	1692	1240	1150	955	976	966	1153
		Dur	134	98	164	110	114	120	195	160	121	134	127
	SD	F1	47	55	52	75	115	121		48	49	45	107
		F2	330	285	287	239	192	184		99	168	140	185
		Dur	33	31	35	25	23	32		26	37	32	27
	Range	F1	153	236	228	306	378	380		196	176	176	388
		F2	1104	909	1022	1008	621	782		352	667	499	779
		Dur	127	127	161	98	84	110		119	134	98	100
	N		24	24	24	24	24	23	1	23	24	23	24

Table A.6.2E. Mean, standard deviation, range, minimum and maximum for F1, F2 (Hz) and duration (ms) of AE vowels produced by American native speakers, aggregated and broken down by gender of speaker. Targets were /hVd/ words. Maximum $N = 20$. [25]

American L1 speakers			American English vowel in /hVd/ context										
Gender	Statistic	Parm	i	ɪ	e	ɛ	æ	ɑ	ɔ	o	ʊ	u	ʌ
Male	Mean	F1	287	425	464	570	712	644	722	489	448	325	613
		F2	2400	2068	2141	1882	1745	1120	1089	1124	1310	1213	1308
		Dur	235	190	268	186	260	232	268	262	177	215	176
	SD	F1	45	39	44	44	106	114	56	104	51	62	51
		F2	191	222	262	218	151	148	164	148	175	253	104
		Dur	46	35	49	35	40	48	30	52	45	38	36
	Min	F1	240	381	401	519	577	452	625	307	367	179	523
		F2	2139	1763	1901	1621	1554	887	876	911	1099	899	1157
		Dur	182	124	193	100	205	117	220	163	113	171	114
	Max	F1	383	515	527	663	939	774	795	618	519	416	669
		F2	2735	2483	2775	2417	2043	1368	1461	1418	1551	1752	1445
		Dur	318	255	347	217	328	273	317	334	266	285	243
	Range	F1	143	134	126	144	362	322	170	311	152	237	146
		F2	596	720	874	796	489	481	584	507	452	853	287
		Dur	136	131	154	117	123	156	97	171	153	114	129
	N		10	10	10	10	10	10	10	10	10	10	10
Female	Mean	F1	359	511	508	720	948	865	814	561	542	404	737
		F2	2828	2383	2659	2193	1996	1305	1203	1378	1544	1514	1576
		Dur	228	178	264	185	256	222	264	226	172	236	163
	SD	F1	23	60	63	45	105	81	91	80	66	39	95
		F2	206	107	191	119	181	90	100	238	116	324	143
		Dur	38	25	40	31	43	53	31	42	36	25	32
	Min	F1	311	449	420	657	775	729	602	401	404	338	606
		F2	2496	2167	2330	2046	1740	1157	1033	1120	1417	1172	1375
		Dur	176	147	195	145	191	158	215	163	125	201	101
	Max	F1	390	602	627	785	1127	960	902	664	622	453	939
		F2	3141	2570	2874	2392	2304	1446	1322	1725	1775	2024	1858
		Dur	291	226	321	253	347	343	322	316	255	286	198
	Range	F1	79	154	206	128	352	231	300	263	218	115	332
		F2	646	403	545	346	563	290	289	605	358	852	483
		Dur	115	79	126	108	156	185	107	153	130	85	97
	N		10	10	10	10	10	10	10	10	10	10	10
All	Mean	F1	323	468	486	645	830	755	768	525	495	365	675
		F2	2614	2225	2400	2038	1871	1212	1146	1251	1427	1364	1442
		Dur	232	184	266	185	258	227	266	244	175	225	169
	SD	F1	51	66	57	88	159	149	87	97	75	64	98
		F2	292	234	347	234	207	153	144	233	188	322	184
		Dur	41	31	43	32	40	49	30	50	40	33	34
	Range	F1	150	221	226	266	550	508	300	357	255	273	416
		F2	1002	807	973	796	750	559	584	814	676	1125	701
		Dur	142	131	154	153	156	226	107	171	153	115	142
	N		20	20	20	20	20	20	20	20	20	20	20

[25] The data in this appendix were made available by Prof. dr. Hongyan Wang of Shenzhen University, P. R. China.

Table A.6.3 Confusion matrices for intended vowel by predicted vowel. Automatic classification by LDA with leave-one-out cross-validation. The left part of the table uses two predictors (F1, F2), the right part adds vowel duration as a third predictor. Green cells on main diagonal contain correct predictions (%), red cells contain the most frequent errors (only when correct prediction ≤ 75%).

		Predicted vowel type, from F1 & F2										Predicted vowel type, from F1, F2 & duration									
		i	ɪ	e	ɛ	æ	ɑ	ʌ	o	ʊ	u	i	ɪ	e	ɛ	æ	ɑ	ʌ	o	ʊ	u
Female bilingual	i	**84.6**	11.5	3.8								**88.5**	7.7	3.8							
	ɪ	26.9	50.0	19.2	3.8							3.8	**88.5**	3.8	3.8						
	e		30.8	53.8	15.4							3.8	15.4	69.2	11.5						
	ɛ			11.5	**88.5**									7.7	**92.3**						
	æ					**100.0**										**100.0**					
	ɑ						69.2	30.8									69.2	30.8			
	ʌ						34.6	57.7	3.8	3.8							23.1	69.2	3.8	3.8	
	o								**84.6**	7.7	7.7								**84.6**	15.4	
	ʊ								23.1	42.3	34.6						3.8		11.5	46.2	38.5
	u								7.7	30.8	61.5								11.5	34.6	53.8
Female monolingual	i	72.7	13.6	13.6								68.2	18.2	13.6							
	ɪ	40.9	36.4	13.6	9.1							31.8	59.1		9.1						
	e	9.1		68.2	22.7							9.1		59.1	27.3		4.5				
	ɛ			22.7	**77.3**									13.6	**86.4**						
	æ					**86.4**	13.6									**86.4**	13.6				
	ɑ				4.5	9.1	40.9	36.4	4.5		4.5		4.5		4.5	54.5	31.8		4.5		
	ʌ						18.2	72.7			9.1					22.7	68.2		9.1		
	o						4.5	4.5	**77.3**	4.5	9.1					4.5	4.5	**81.8**	4.5	4.5	
	ʊ								4.5	54.5	40.9							4.5	59.1	36.4	
	u								18.2	27.3	54.5							18.2	27.3	54.5	
Male bilingual	i	54.5	40.9	4.5								**77.3**	18.2	4.5							
	ɪ	27.3	68.2	4.5								13.6	**86.4**								
	e	9.1		68.2	22.7							9.1		59.1	27.3		4.5				
	ɛ			22.7	**77.3**									13.6	**86.4**						
	æ					**86.4**	13.6									**86.4**	13.6				
	ɑ				4.5	9.1	40.9	36.4	4.5		4.5		4.5		4.5	54.5	31.8		4.5		
	ʌ					4.5	13.6	68.2	9.1		4.5					4.5	9.1	72.7	4.5	4.5	4.5
	o						4.5		72.7	4.5	18.2							4.5	72.7		22.7
	ʊ								9.1	45.5	45.5								4.5	**77.3**	18.2
	u								18.2	50.0	31.8								18.2	27.3	54.5
Male monolingual	i	**80.0**	15.0	5.0								**85.0**	10.0	5.0							
	ɪ	15.0	**80.0**	5.0								10.0	**85.0**		5.0						
	e	5.0		70.0	25.0							5.0		75.0	20.0						
	ɛ	5.0	5.0	20.0	70.0							5.0		5.0	**90.0**						
	æ					**100.0**										**100.0**					
	ɑ						**80.0**	20.0									**80.0**	20.0			
	ʌ						25.0	75.0									25.0	75.0			
	o								**80.0**	10.0	10.0								**90.0**	5.0	5.0
	ʊ								5.0	75.0	20.0								5.0	65.0	30.0
	u								15.0	50.0	35.0								10.0	45.0	45.0
All groups combined	i	68.9	22.2	8.9								74.4	16.7	7.8	1.1						
	ɪ	23.3	63.3	8.9	4.4							12.2	**80.0**	2.2	5.6						
	e	7.8	12.2	58.9	21.1							8.9	6.7	63.3	21.1						
	ɛ	1.1	2.2	14.4	**81.1**	1.1						1.1	2.2	13.3	**83.3**						
	æ				2.2	**94.4**	3.3								1.1	**95.6**	3.3				
	ɑ				1.1	2.2	63.3	28.9	3.3		1.1			1.1		1.1	66.7	28.9	1.1	1.1	
	ʌ					2.2	17.8	68.9	6.7	1.1	3.3					2.2	17.8	68.9	5.6	4.4	1.1
	o						2.2	1.1	**80.0**	8.9	7.8						3.3	2.2	**80.0**	5.6	8.9
	ʊ								16.7	53.3	30.0						1.1		11.1	61.1	26.7
	u								16.7	40.0	43.3						1.1		12.2	35.6	51.1

Table A6.4. Confusion matrices for intended vowel by predicted vowel. Automatic classification by Multinomial Logistic Regression Analysis. For more information, see caption of Table A6.3.

		Predicted vowel type, from F1 & F2										Predicted vowel type, from F1, F2 & duration									
		i	ɪ	e	ɛ	æ	ɑ	ʌ	o	ʊ	u	i	ɪ	e	ɛ	æ	ɑ	ʌ	o	ʊ	u
Female bilingual	i	**84.6**	11.5	3.8								**88.5**	7.7	3.8							
	ɪ	26.9	50.0	19.2	3.8							3.8	**88.5**	3.8	3.8						
	e		30.8	53.8	15.4							3.8	15.4	69.2	11.5						
	ɛ			11.5	**88.5**									7.7	**92.3**						
	æ					**100.0**										**100.0**					
	ɑ						69.2	30.8									69.2	30.8			
	ʌ						34.6	57.7	3.8	3.8							23.1	69.2	3.8	3.8	
	o								**84.6**	7.7	7.7								**84.6**	15.4	
	ʊ								23.1	42.3	34.6						3.8		11.5	46.2	38.5
	u								7.7	30.8	61.5								11.5	34.6	53.8
Female monolingual	i	72.7	13.6	13.6								68.2	18.2	13.6							
	ɪ	40.9	36.4	13.6	9.1							31.8	59.1		9.1						
	e	9.1	22.7	45.5	22.7							4.5	4.5	**77.3**	13.6						
	ɛ			13.6	**86.4**								9.1	13.6	**77.3**						
	æ					**100.0**										**100.0**					
	ɑ						63.6	27.3	9.1								63.6	31.8	4.5		
	ʌ						18.2	72.7			9.1						22.7	68.2		9.1	
	o						4.5	4.5	**77.3**	4.5	9.1						4.5	4.5	**81.8**	4.5	4.5
	ʊ								4.5	54.5	40.9								4.5	59.1	36.4
	u								18.2	27.3	54.5								18.2	27.3	54.5
Male bilingual	i	54.5	40.9	4.5								**77.3**	18.2	4.5							
	ɪ	27.3	68.2	4.5								13.6	**86.4**								
	e	9.1		68.2	22.7							9.1		59.1	27.3		4.5				
	ɛ			22.7	**77.3**									13.6	**86.4**						
	æ					**86.4**	13.6									**86.4**	13.6				
	ɑ				4.5	9.1	40.9	36.4	4.5		4.5			4.5		4.5	54.5	31.8		4.5	
	ʌ					4.5	13.6	68.2	9.1		4.5					4.5	9.1	72.7	4.5	4.5	4.5
	o						4.5		72.7	4.5	18.2							4.5	72.7		22.7
	ʊ								9.1	45.5	45.5								4.5	**77.3**	18.2
	u								18.2	50.0	31.8								18.2	27.3	54.5
Male monolingual	i	**80.0**	15.0	5.0								**85.0**	10.0	5.0							
	ɪ	15.0	**80.0**	5.0								10.0	**85.0**		5.0						
	e	5.0		70.0	25.0							5.0		75.0	20.0						
	ɛ	5.0	5.0	20.0	70.0							5.0		5.0	**90.0**						
	æ					**100.0**										**100.0**					
	ɑ						**80.0**	20.0									**80.0**	20.0			
	ʌ						25.0	75.0									25.0	75.0			
	o								**80.0**	10.0	10.0								**90.0**	5.0	5.0
	ʊ								5.0	75.0	20.0								5.0	65.0	30.0
	u								15.0	50.0	35.0								10.0	45.0	45.0
All groups combined	i	68.9	22.2	8.9								74.4	16.7	7.8	1.1						
	ɪ	23.3	63.3	8.9	4.4							12.2	**80.0**	2.2	5.6						
	e	7.8	12.2	58.9	21.1							8.9	6.7	63.3	21.1						
	ɛ	1.1	2.2	14.4	**81.1**	1.1						1.1	2.2	13.3	**83.3**						
	æ				2.2	**94.4**	3.3								1.1	**95.6**	3.3				
	ɑ				1.1	2.2	63.3	28.9	3.3		1.1			1.1		1.1	66.7	28.9	1.1	1.1	
	ʌ					2.2	17.8	68.9	6.7	1.1	3.3					2.2	17.8	68.9	5.6	4.4	1.1
	o						2.2	1.1	**80.0**	8.9	7.8						3.3	2.2	**80.0**	5.6	8.9
	ʊ								16.7	53.3	30.0						1.1		11.1	61.1	26.7
	u								16.7	40.0	43.3						1.1		12.2	35.6	51.1

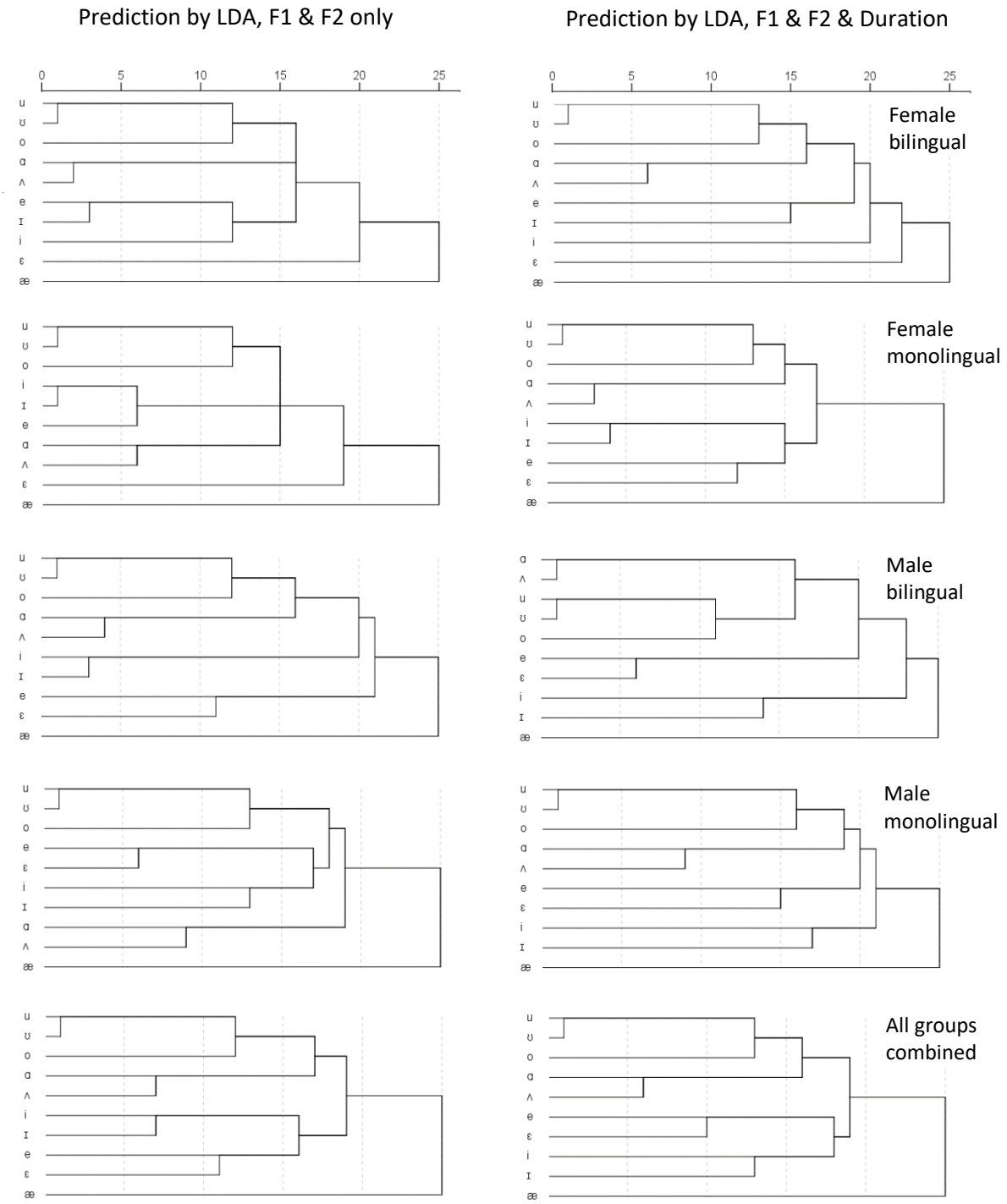


Figure A6.5. Hierarchical tree structures for vowel confusion determined by Linear Discriminant Analysis. Average linking was used with Euclidean distance. Similarity rescaled between 0 and 25. Left panels are based on predictions from spectral parameters F1 and F2 only, right panels on spectral parameters plus vowel duration. From top to bottom the panels are for the female bilinguals, female monolinguals, male bilinguals, male monolinguals, and (bottom row) for all groups combined.

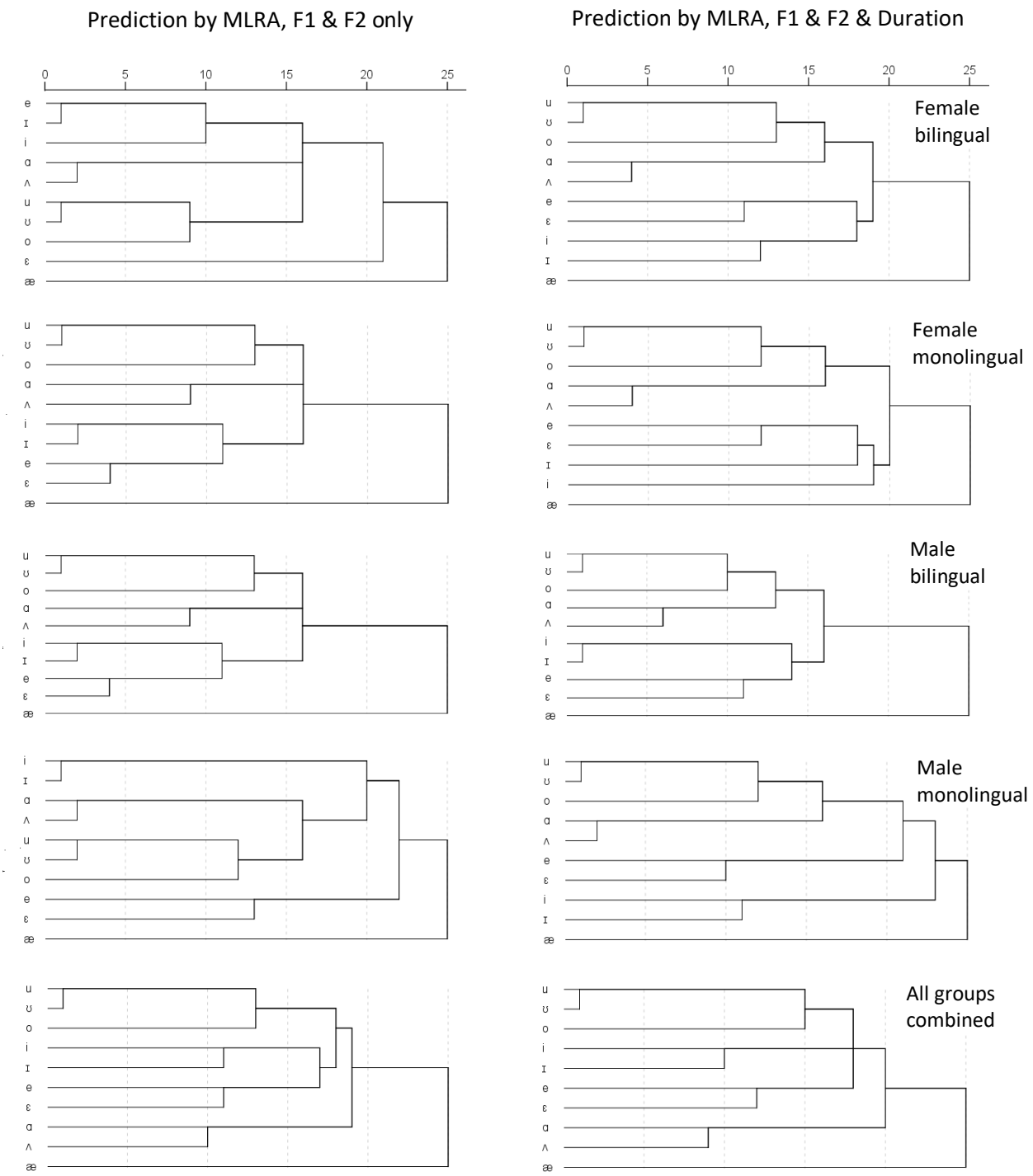


Figure A6.6. Hierarchical tree structures for vowel confusion determined by Multinomial Logistic Regression Analysis. For more information see previous caption.

Table A6.7 Confusion matrices for intended vowel by vowel predicted by model trained on native American vowel tokens. Automatic classification by LDA with leave-one-out cross-validation when model is trained and tested on American native speaker data (top panel). In the second, third and bottom panel, the native model is tested on the non-native tokens produced by Persian monolinguals, Azerbaijani/Persian bilinguals, and by both groups combined. The left part of the table uses two predictors (F1, F2), the right part adds vowel duration as a third predictor. Green cells on main diagonal contain correct predictions (%), red cells contain the most frequent errors (only when correct prediction ≤ 75%).

		Predicted vowel type, from F1 & F2										Predicted vowel type, from F1, F2 & duration									
		i	ɪ	e	ɛ	æ	ɑ	ʌ	o	ʊ	u	i	ɪ	e	ɛ	æ	ɑ	ʌ	o	ʊ	u
American English	i	**100**										**100**									
	ɪ		**85.0**	15.0									**100**								
	e		5.0	**90.0**	5.0								10.0	**90.0**							
	ɛ				**100**										**100**						
	æ				10.0	**90.0**										**100**					
	ɑ						**85.0**		15.0								**80.0**	5.0	15.0		
	ʌ						10.0	**85.0**		5.0							10.0	**80.0**		10.0	
	o						5.0	5.0	55.0	20.0	15.0						5.0		70.0	10.0	15.0
	ʊ								15.0	70.0	15.0				5.0				5.0	**80.0**	10.0
	u										**100**									5.0	**95.0**
Monolingual	i	57.8	5.6	35.6	1.1							44.4	30.0	24.4	1.1						
	ɪ	37.8	45.6	14.4	2.2							16.7	**80.0**	1.1	2.2						
	e	1.1	35.6	45.6	17.8								18.9	71.1	10.0						
	ɛ		28.9	7.8	61.1	2.2						1.1	25.6	13.3	60.0						
	æ				2.2	**95.6**		2.2							2.2	**95.6**		2.2			
	ɑ				1.1	3.3	48.9	42.2	4.4					1.1		4.4	65.6	24.4	2.2	2.2	
	ʌ					2.2	82.2	5.6	10.0						1.1	1.1	77.8	7.8	7.8	4.4	
	o						1.1	3.3	**88.9**	2.2	4.4						2.2	2.2	**87.8**	4.4	3.3
	ʊ								48.9	7.8	43.3								36.7	20.0	43.3
	u								53.3	14.4	32.2								52.2	13.3	34.4
Bilingual	i	59.5	2.4	35.7	2.4							38.1	33.3	26.2	2.4						
	ɪ	40.5	40.5	14.3	4.8							19.0	**76.2**		4.8						
	e		40.5	35.7	23.8								7.1	**83.3**	9.5						
	ɛ		31.0	4.8	59.5	4.8						2.4	21.4	16.7	59.5						
	æ					**100**										**100**					
	ɑ						57.1	38.1	4.8							4.8	64.3	26.2	4.8		
	ʌ						88.1	4.8	7.1								78.6	11.9	4.8	4.8	
	o						2.4	4.8	**85.7**	2.4	4.8						2.4	4.8	**83.3**	4.8	4.8
	ʊ								33.3	9.5	57.1								26.2	14.3	59.5
	u								47.6	16.7	35.7								45.2	16.7	38.1
All EFL speakers	i	56.3	8.3	35.4								50.0	27.1	22.9							
	ɪ	35.4	50.0	14.6								14.6	**83.3**	2.1							
	e	2.1	31.3	54.2	12.5								29.2	60.4	10.4						
	ɛ		27.1	10.4	62.5								29.2	10.4	60.4						
	æ				4.2	**91.7**		4.2							4.2	**91.7**		4.2			
	ɑ				2.1	6.3	41.7	45.8	4.2					2.1		4.2	66.7	22.9		4.2	
	ʌ					4.2	77.1	6.3	12.5						2.1	2.1	77.1	4.2	10.4	4.2	
	o							2.1	**91.7**	2.1	4.2						2.1		**91.7**	4.2	2.1
	ʊ								62.5	6.3	31.3								45.8	25.0	29.2
	u								58.3	12.5	29.2								58.3	10.4	31.3