Early vs. late Serbian-English bilinguals’ responses
to two Australian English vowel contrasts

Lidija Krebs-Lazendic
(BA) Linguistics, Honours Class 1, University of Novi Sad, Serbia

A thesis submitted for the degree of Doctor of Philosophy
MARCS Auditory Laboratories
University of Western Sydney
August 2008
I hereby declare that this submission is my own work and to the best of my knowledge it contains no material previously published or written by another person, nor material which has been accepted for the award of any other degree or diploma at the University of Western Sydney, or at any other educational institution, except where due acknowledgment is made in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project’s design and conception is acknowledged.

___________________________________
Lidija Krebs Lazendic
ACKNOWLEDGEMENTS

It has been said that there are three types of people: those who watch things happen, those who wonder what happened (the vast majority of mankind find themselves in these two groups) and those who make things happen. During last couple of years I have been surrounded with people who make things happen. Without them, I would have never finished this thesis.

These people are:

My principal supervisor Professor Cathi Best and co-supervisors Professor Denis Burnham and Dr Heather Winskel.

My friends and colleagues from MARCS who helped in various stages of my work: Arman Abrahanyan, Dogu Erdener Mark Antoniou, Tim Byron, Jemma Harris, Benjawan Kasisopa, Bettina Keresztesi, Christa Lam, Nathan Perry, Gail Charlton, Mel Gallagher, Karen McConachie, Darlene Williams, Laurline Beard, Colin Schoknecht, Johnson Chen.

My friends and colleagues from the School of Humanities and Languages: Bruno di Biase, Satomi Kawaguchi, Ruying Qi, George Sunders, Geoffrey Hull.


My best friend Nan Xu, who always knows when something is wrong, even if I have the biggest smile on my face. She knows me and loves me anyway.

Goran, who believed in me even when I didn’t. The prize of him at the end of the day is what made my work possible.

This thesis is dedicated to Maja, the bravest girl in the world. She keeps reminding me what really matters in life.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................................................ III
LIST OF TABLES .................................................................................................................................................... VII
LIST OF FIGURES ................................................................................................................................................ IX
ABSTRACT .......................................................................................................................................................... XI

CHAPTER 1 LEARNING SECOND LANGUAGE PHONETIC SEGMENTS ............................................................ 2
1.1 Speech Learning Model (SLM) .................................................................................................................. 3
1.2 Perceptual Assimilation Model (PAM) ...................................................................................................... 5

CHAPTER 2 PRODUCTION OF SPEECH SEGMENTS ....................................................................................... 9
2.1 An Overview of the Speech Production Process ..................................................................................... 9
2.1.1 The Role of Vocal Organs in Speech Production ............................................................................... 9
2.1.2 Respiration, Phonation (Voicing) and Articulation ........................................................................... 10
2.1.3 Articulation and the Source-Filter Theory of Speech Production .................................................. 12
2.2 Vowels as Speech Segments: Articulation and Acoustics of Vowel Production .................................. 13
2.2.1 Monophthongs .................................................................................................................................. 14
2.3 Diphthongs ............................................................................................................................................... 16
2.4 The Acoustic Characteristics of Vowels .................................................................................................. 16
2.4.1 Vowel Formants ................................................................................................................................ 16
2.4.2 Vowel Duration .................................................................................................................................. 17
2.4.3 Distinctions Between Consonants and Vowels ................................................................................. 18
2.5 Phonetic Variability: Coarticulation in Speech Production .................................................................. 18
2.6 Articulatory Phonology: Gesture as a Basic Phonological Unit ............................................................. 20

CHAPTER 3 PERCEPTION OF SPEECH SEGMENTS ....................................................................................... 23
3.1 Vowel Perception ................................................................................................................................... 23
3.1.1 Categorical Perception ...................................................................................................................... 23
3.1.2 Continuous Perception ...................................................................................................................... 25
3.1.3 Formants (F1, F2 and F3) in Perception of Vowels ........................................................................... 26
3.1.4 Dynamic Properties in Vowel Perception ....................................................................................... 27
3.2 Perception of Articulatory Gestures ...................................................................................................... 28
3.2.1 The Motor Theory of Speech Perception ......................................................................................... 28
3.2.2 The Direct Realist Theory of Speech Perception ............................................................................. 30

CHAPTER 4 SPEECH PRODUCTION IN A SECOND LANGUAGE .................................................................. 32
4.1 Effects of Age on Second Language Learning ....................................................................................... 32
4.1.1 Critical Period Hypothesis (CPH) .................................................................................................... 33
4.1.2 Age of L2 Learning and L2 Segmental Production: Towards an Alternative Explanation ........... 34
4.2 Production of L2 Phonetic Segments and Language Interaction ......................................................... 36
4.2.1 Interaction Between First and Second Language Phonetic Systems ............................................. 36
4.2.2 The SLM and the Production of Vowels ......................................................................................... 37
4.3 Perceived Distance Between First and Second Language Phonetic Segments ................................ 38
4.4 Other Factors Affecting Second Language Segmental Production ..................................................... 41

CHAPTER 5 CROSS-LANGUAGE SPEECH PERCEPTION ............................................................................ 43
5.1 Adults’ Cross-Language Speech Perception ............................................................................................ 43
5.1.1 Evidence from the Perception of Non-Native Vowel Contrasts .................................................... 44
5.1.2 Effects of Duration on Vowel on L2 Vowel Perception .................................................................. 46
5.2 From Language Universal to Language Specific in Speech Perception ................................................ 47

CHAPTER 6 THE PRESENT RESEARCH PROJECT .................................................................................... 49
6.1 The Serbian Language ............................................................................................................................. 50
6.1.1 Cultural and Linguistic Background ............................................................................................... 50
CHAPTER 7 STUDY 1. PERCEPTION AND PRODUCTION OF AUSTRALIAN ENGLISH /æ/ - /e/ AND /ɛ/ - /ɪ/ BY SERBIAN-ENGLISH BILINGUALS: IMPLICATIONS FOR THE SPEECH LEARNING MODEL

7.1 Predictions based on the SLM .................................................................................................................. 63
7.2 General method ......................................................................................................................................... 64
7.2.1 Participants ............................................................................................................................................. 64
7.2.2 Serbian speakers’ experience with English prior to migration to Australia ............................................ 65
7.2.3 The Serbian speakers’ native dialect ..................................................................................................... 65
7.2.4 Age of learning, length of exposure and chronological age of participants .......................................... 66
7.2.5 Stimuli ................................................................................................................................................... 66
7.2.6 Recording equipment ............................................................................................................................ 67
7.2.7 Experiments .......................................................................................................................................... 68
7.3 Experiment 1: Production of /æ/ - /e/ and /ɛ/ - /ɪ/ contrasts ..................................................................... 68
7.3.1 Design .................................................................................................................................................. 68
7.3.2 Procedure ............................................................................................................................................. 69
7.3.3 Results .................................................................................................................................................. 70
7.3.4 Results: Experiment 1 ............................................................................................................................ 73
7.4 Experiment 2: Perceptual assimilation and goodness-of-fit rating .................................................... 74
7.4.1 Design .................................................................................................................................................. 74
7.4.2 Procedure ............................................................................................................................................. 74
7.4.3 Results: Perceptual assimilation of the /æ/ - /e/ contrast .................................................................... 75
7.4.4 Results: Perceptual assimilation of the /ɛ/ - /ɪ/ contrast .................................................................... 78
7.4.5 Results: Goodness-of-fit rating for the /æ/ - /e/ contrast ................................................................. 79
7.4.6 Results: Goodness-of-fit ratings for the /ɛ/ - /ɪ/ contrast ................................................................. 80
7.4.7 Summary of the results for perceptual assimilation and goodness-of-fit ratings ................................ 81

7.5 Experiment 3: Perceptual discrimination ................................................................................................. 82
7.5.1 Design .................................................................................................................................................. 82
7.5.2 Participants ........................................................................................................................................... 82
7.5.3 Procedure ............................................................................................................................................. 82
7.5.4 Analysis .............................................................................................................................................. 83
7.5.5 Results ................................................................................................................................................ 84
7.5.6 Summary of the results of Experiment 3 .............................................................................................. 85
7.6 General discussion for Study 1 ................................................................................................................ 85

CHAPTER 8 STUDY 2. PERCEPTION AND PRODUCTION OF AUSTRALIAN-ENGLISH /æ/ - /e/ AND /ɛ/ - /ɪ/ BY SERBIAN-ENGLISH BILINGUALS: IMPLICATIONS FOR THE PERCEPTUAL ASSIMILATION MODEL

8.1 Predictions based on PAM ......................................................................................................................... 92
8.2 General method ......................................................................................................................................... 94
8.2.1 Participants ........................................................................................................................................... 94
8.2.2 Experience with English prior to migration to Australia ........................................................................ 94
8.2.3 Age of learning, length of exposure and chronological age .............................................................. 94
8.2.4 Stimuli .................................................................................................................................................. 95
8.2.5 Recording equipment ............................................................................................................................ 95
8.2.6 Experiments ......................................................................................................................................... 95
8.3 Experiment 4: Perceptual assimilation and goodness-of-fit rating .................................................... 96
8.3.1 Design ................................................................................................................................................ 96
CHAPTER 9  PHONOLOGICAL AND PHONETIC PRIMING EFFECTS ON PERCEPTION
OF ENGLISH /æ/ - /e/ AND /iː/ - /i/ BY SERBIAN-ENGLISH BILINGUALS .............. 118

9.1 PRIMING PARADIGM ........................................................................................................... 119
9.1.1 PRIMING AND THE ROLE OF PHONOLOGY IN AUDITORY WORD RECOGNITION .......... 120
9.1.2 TASKS IN PHONOLOGICAL PRIMING EXPERIMENTS ................................................ 121
9.1.3 AMOUNT AND LOCATION OF PHONOLOGICAL OVERLAP BETWEEN THE PRIME AND
THE TARGET ............................................................................................................................ 121
9.1.4 INTERSTIMULUS INTERVAL (ISI) AND STIMULUS ONSET ASYNCHRONY (SOA) .............. 123
9.1.5 THE ROLE OF ISI IN CROSS-LANGUAGE SPEECH PERCEPTION .................................... 122
9.1.6 PREDICTIONS ............................................................................................................... 124
9.1.7 PARTICIPANTS .............................................................................................................. 127
9.1.8 STIMULI ....................................................................................................................... 128
9.2 EXPERIMENTS 7 AND 8 ................................................................................................ 128
9.2.1 PROCEDURES ............................................................................................................ 128
9.2.2 DESIGN .................................................................................................................... 133
9.3 RESULTS: EXPERIMENT 1 - PRIMED PERCEPTUAL ASSIMILATION AND GOODNESS-OF-FIT
RATING .................................................................................................................................... 133
9.3.1 1000 MS ISI CONDITION .............................................................................................. 133
9.3.2 500 MS ISI CONDITION .............................................................................................. 140
9.4 RESULTS: EXPERIMENT 2 - PRIMED DISCRIMINATION OF /æ/ - /e/ AND /iː/ - /i/ .................................................. 145
9.4.1 1000 MS ISI CONDITION .............................................................................................. 145
9.4.2 500 MS ISI CONDITION .............................................................................................. 148
9.5 SUMMARY OF THE RESULTS AND DISCUSSION OF STUDY 3 ........................................ 150

CHAPTER 10  GENERAL DISCUSSION ....................................................................................... 154

10.1 SUMMARY OF THE FINDINGS .......................................................................................... 155
10.1.1 STUDY 1: IMPLICATIONS FOR THE SPEECH LEARNING MODEL .............................. 155
10.1.2 STUDY 2: IMPLICATIONS FOR THE PERCEPTUAL ASSIMILATION MODEL .......... 157
10.1.3 STUDY 3: PHONOLOGICAL AND PHONETIC PRIMING EFFECTS ON L2 VOWEL PERCEPTION .......... 158
10.1.4 SUMMARY OF FINDINGS IN THE LIGHT OF SLM AND PAM ...................................... 157
10.2 THEORETICAL IMPLICATIONS AND FUTURE RESEARCH: AOL AND L2 VOWEL PERCEPTION ...... 161

REFERENCES .......................................................................................................................... 166
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-1</td>
<td>Serbian vowels</td>
<td>52</td>
</tr>
<tr>
<td>6-3</td>
<td>Ause vowels</td>
<td>55</td>
</tr>
<tr>
<td>7-1</td>
<td>Analysis of variance for production of the /e/-/æ/ contrast</td>
<td>71</td>
</tr>
<tr>
<td>7-2</td>
<td>Analysis of variance for production of the /ɪ/-/u/ contrast</td>
<td>73</td>
</tr>
<tr>
<td>7-3</td>
<td>Analysis of variance for perceptual assimilation of the /e/-/æ/ contrast</td>
<td>77</td>
</tr>
<tr>
<td>7-4</td>
<td>Analysis of variance for perceptual assimilation of the /ɪ/-/u/ contrast</td>
<td>79</td>
</tr>
<tr>
<td>7-5</td>
<td>Mean ratings for HAD and HEAD</td>
<td>80</td>
</tr>
<tr>
<td>7-6</td>
<td>Mean ratings for HEED and HID</td>
<td>80</td>
</tr>
<tr>
<td>7-7</td>
<td>Analysis of variance for discrimination</td>
<td>84</td>
</tr>
<tr>
<td>8-1</td>
<td>Serbian keywords with short and long vowels</td>
<td>96</td>
</tr>
<tr>
<td>8-2</td>
<td>Analysis of variance for perceptual assimilation of th /æ/-/e/ contrast</td>
<td>100</td>
</tr>
<tr>
<td>8-3</td>
<td>Analysis of variance for perceptual assimilation of the /ɪ/-/u/ contrast</td>
<td>102</td>
</tr>
<tr>
<td>8-4</td>
<td>Mean goodness-of-fit ratings ratings for ause /æ/ and /e/ by early (el) and late (ll) serbian-english bilinguals</td>
<td>102</td>
</tr>
<tr>
<td>8-5</td>
<td>Mean goodness-of-fit ratings ratings for ause /ɪ/ and /u/ by early (el) and late (ll) serbian-english bilinguals</td>
<td>103</td>
</tr>
<tr>
<td>8-6</td>
<td>Analysis of variance for perceptual discrimination of the /æ/-/e/ and /ɪ/-/u/ contrasts</td>
<td>108</td>
</tr>
<tr>
<td>8-7</td>
<td>Analysis of variance for production of the /æ/-/e/ contrast</td>
<td>112</td>
</tr>
<tr>
<td>8-8</td>
<td>Analysis of variance for production of the /ɪ/-/u/ contrast</td>
<td>114</td>
</tr>
</tbody>
</table>
TABLE 9-1. THE STIMULUS TRIADS FOR TRIALS IN THE PHONOLOGICAL, PHONETIC, AND UNRELATED PRIME-TARGET CONDITIONS .................................................. 130

TABLE 9-2. THE TOKENS USED IN THE PRACTICE FOR THE PRIMED AX TASK ................. 131

TABLE 9-3. THE PRIME AND THE TARGET RELATIONSHIPS IN THE PRIMED PERCEPTUAL ASSIMILATION EXPERIMENTS ......................................................... 131

TABLE 9-4. PERCEPTUAL ASSIMILATION AT 1000 MS ISI EVC1 ....................................... 134

TABLE 9-5. ANALYSIS OF VARIANCE FOR PERCEPTUAL ASSIMILATION OF THE /æ/ - /e/ CONTRAST (1000 MS ISI) ........................................................................................................... 135

TABLE 9-6. PERCEPTUAL ASSIMILATION AT 1000 MS ISI EVC2 ....................................... 136

TABLE 9-7. ANALYSIS OF VARIANCE FOR PERCEPTUAL ASSIMILATION OF THE /ɛː/ - /ɪ/ CONTRAST 1000 MS ........................................................................................................ 137

TABLE 9-8. GOODNESS-OF-FIT RATING DIFFERENCE SCORES IN 1000 MS ISI .......... 139

TABLE 9-9. GOODNESS-OF-FIT RATING DIFFERENCE SCORES IN 1000 MS ISI .......... 139

TABLE 9-10. PERCEPTUAL ASSIMILATION AT 500 MS ISI EVC ......................................... 140

TABLE 9-11. ANALYSIS OF VARIANCE FOR PERCEPTUAL ASSIMILATION OF THE /æ/ - /e/ CONTRAST IN 500 MS ISI CONDITION ................................................................. 141

TABLE 9-12. PERCEPTUAL ASSIMILATION AT 500 MS ISI EVC2 ....................................... 142

TABLE 9-13. ANALYSIS OF VARIANCE FOR PERCEPTUAL ASSIMILATION OF THE /ɛː/ - /ɪ/ CONTRAST IN 500 MS ISI CONDITION ................................................................. 143

TABLE 9-14. GOODNESS-OF-FIT RATING DIFFERENCE SCORES IN 500 MS ISI EVC1 ...... 144

TABLE 9-15. GOODNESS-OF-FIT RATING DIFFERENCE SCORES IN 500 MS ISI EVC2 ...... 145

TABLE 9-16. DISCRIMINATION FOR THE 1000 MS ISI ......................................................... 146

TABLE 9-17. ANALYSIS OF VARIANCE FOR PERCEPTUAL DISCRIMINATION IN THE 1000 MS ISI CONDITION ............................................................................................ 147

TABLE 9-18. DISCRIMINATION 500 MS ISI ........................................................................... 148

TABLE 9-19. ANALYSIS OF VARIANCE FOR PERCEPTUAL DISCRIMINATION IN THE 500 MS ISI CONDITION ....................................................................................... 149
LIST OF FIGURES

FIGURE 2-1. A SIMPLIFIED DIAGRAM OF THE ORGANS OF SPEECH
(CLARK, YALLOP, & FLETCHER, 2007).................................................................. 10

FIGURE 2-2. FUNCTIONAL MODEL OF THE VOCAL TRACT
(CLARK ET AL., 2007)....................................................................................... 11

FIGURE 2-3. A SOURCE AND FILTER VIEW OF SPEECH PRODUCTION
(LADEFOGED, 1996).............................................................................................. 13

FIGURE 2-4. THE MONOPHTHONGS OF AUSTRALIAN ENGLISH.
(BERNARD, 1970, COX, 1996)........................................................................... 15

FIGURE 2-5. TRACT VARIABLES AND ASSOCIATED ARTICULATORS
(BROWMAN & GOLDSTEIN, 1992A).................................................................. 21

FIGURE 3-1. IDEALIZED FORM OF CATEGORICAL PERCEPTION SHOWING
IDENTIFICATION PERFORMANCE FOR TWO CATEGORIES AND
DISCRIMINATION PERFORMANCE BETWEEN CATEGORIES
(PISONI, 1978)......................................................................................................... 24

FIGURE 4-1. MEAN PRONUNCIATION ACCURACY RATINGS
(FLEG, MUNRO & MACKAY, 1995).................................................................. 34

FIGURE 6-1. SERBO-CROATIAN DIALECTS (KORDIC, 1997).................................. 51

FIGURE 6-2. DISTRIBUTION OF SERBIAN /i/, /e/, /a/ AND AUSE /æ:/, /ɛ:/, /æ/, AND /æ/
VOWELS IN F1-F2 VOWEL SPACE
(COX, 1996 ; SIMIC & OSTOJIC , 1987).................................................................. 57

FIGURE 7-1. PRODUCTION OF ENGLISH HAD AND HEAD BY
EARLY LEARNERS (EL) AND LATE LEARNERS (LL) SERBIAN-ENGLISH
BILINGUALS AND ENGLISH MONOLINGUALS (M)............................................ 70

FIGURE 7-2. PRODUCTION OF ENGLISH HEED, AND HID BY
EARLY LEARNER (EL) AND LATE LEARNER (LL) SERBIAN-ENGLISH
BILINGUALS AND ENGLISH MONOLINGUALS (M)............................................ 72

FIGURE 7-3. PERCEPTUAL ASSIMILATION OF ENGLISH HAD, AND HEAD BY
EARLY (EL) AND LATE (LL) SERBIAN-ENGLISH BILINGUALS ....................... 76

FIGURE 7-4. PERCEPTUAL ASSIMILATION OF ENGLISH HEED AND HID BY
EARLY (EL) AND LATE (LL) SERBIAN-ENGLISH BILINGUALS ....................... 78

FIGURE 7-5. MEAN D-PRIME (D') SCORES OBTAINED FOR THE EARLY (EL)
AND LATE (LL) SERBIAN-ENGLISH BILINGUALS AND FOR THE AUSE
FIGURE 8-1. PERCEPTUAL ASSIMILATION OF AUSE /æ/ AND /e/ BY EARLY (EL) AND LATE (LL) SERBIAN L2 LEARNERS.................................99

FIGURE 8-2. PERCEPTUAL ASSIMILATION OF AUSE /ɜ:/ AND /u/ BY EARLY (EL) AND LATE (LL) SERBIAN L2 LEARNERS.............................101

FIGURE 8-3. DISCRIMINATION OF AUSE /æ/ - /e/ AND /ɜ:/ - /u/ CONTRASTS BY EARLY AND LATE SERBIAN BILINGUALS AND AUSE MONOLINGUALS..........................................................107

FIGURE 8-4. PRODUCTION OF ENGLISH HAD, AND HEAD BY EARLY (EL) AND LATE (LL) SERBIAN-ENGLISH BILINGUALS AND ENGLISH MONOLINGUALS (M).................................................................111

FIGURE 8-5. PRODUCTION OF ENGLISH HEED, AND HID BY EARLY (EL) AND LATE (LL) SERBIAN-ENGLISH BILINGUALS AND ENGLISH MONOLINGUALS.................................................................113
ABSTRACT

Adults learning a second language (L2) (“late learners”) have difficulty achieving a native speaker’s level of accuracy in both perception and production of L2 phonetic segments. This difficulty often results in deviant production of L2 segments that is perceived as accented speech by native speakers of that language. It is generally agreed that this failure in nonnative segmental production and perception is caused by previous linguistic experience with the first (L1) language. Late learners are expected to show stronger L1 effects than learners who learnt their L2 in early childhood (“early learners”).

However, not all L2 phonetic segments are equally difficult for late learners. The learnability of L2 phonetic segments is thought to be perceptual in nature and depends on the perceived phonetic distance between them and the acoustically, phonetically and/or articulatorily most similar segment(s) in the learner’s L1 phonetic inventory. It is generally assumed that specific L2 segments will be perceptually related or assimilated to the most similar L1 segment(s) even if there is a detectable acoustic difference between them.

The studies reported in this thesis examined Serbian-English bilinguals’ perception and production difficulties with two Australian English vowel contrasts that are not contrastive in Serbian: /e/ - /æ/ and /i:/ - /u/. We compared participants who began learning English before 5 years (“early”) versus those who began after 15 years (“late”). In Study 1 and Study 2 early learners discriminated and produced both contrasts equally well, whereas late learners had greater difficulties perceiving and producing /e/ - /æ/. In Study 3 a priming paradigm was applied to discrimination and perceptual assimilation tasks in which the prime and target contain phonologically identical, phonetically similar or phonologically and phonetically unrelated vowels under two interstimulus intervals (ISI) that tap phonological versus phonetic levels of processing, according to prior research. Early versus late group differences suggest that discrimination and production accuracy reflect how listeners assimilate Australian English vowels to native Serbian vowels. “Early” and “late” learners related L2 vowels to L1 differently, which reflects differences in establishment of the L1 phonetic system at the time of L2 onset.
Adults learning a second language (L2) have difficulty achieving a level of accuracy equivalent to native speakers of that language. It seems that this is particularly true for the acquisition of phonetic segments (vowels and consonants). Bilinguals who have learned their second language in childhood generally produce second language phonetic segments more accurately (more like native speakers of the L2) than bilinguals who have learned their second language in adolescence or adulthood. The pronunciation of bilinguals from the latter group is often perceived by native speakers of the target L2 as divergent from their phonetic norms. This divergence is known as foreign accent in pronunciation or accented speech. Research in this area is mainly motivated by the aim of understanding why L2 learners, especially those whose L2 acquisition commences beyond childhood, differ in their pronunciation from native speakers of that language.

Various theories have been offered in support of the common belief that children acquire L2 pronunciation much more easily and with greater success than adults. While there is agreement on the advantage of early second language learning for nativelike L2 pronunciation, the theories offered to account for this phenomenon are quite distinct. One explanation is based on the assumptions of the Critical Period Hypothesis (CPH) that neurobiological development on a strict maturational timetable creates limits on language learning capacity. This assumption takes age as the main predictor of success in language learning. The influence of age in second language learning will be discussed in Chapter 4.

Another explanation for successes and failures by children and adults learning an L2 is the interactionist approach. This approach to L2 learning takes into account interactions between a learner’s first (L1) and second languages, and age of L2 learning as an index of development of the first language. This interaction is the basic assumption of the Speech Learning Model (SLM, Flege, 1987, 1992a, 1992b, 1995) which has been devised to account for difficulties in production of non-native vowels and consonants. Research within the SLM framework has focused on bilinguals who are very proficient in their L2, with an aim of explaining why some L2 segments are easier
to acquire than others. SLM forms one portion of the theoretical background for the research project reported in this thesis, and is most relevant to the L2 production research reported in Chapter 7. Another relevant theoretical framework, especially for the L2 perception research reported in Chapter 8, is the Perceptual Assimilation Model (PAM) (Best, 1994; 1995; Best, McRoberts & Goodell, 2001). According to these two models, the phonetic segments of an adult’s native language influence the perception of foreign segments and can strongly interfere with recognizing or distinguishing one foreign speech segment from another, thus limiting a listener’s ability to discriminate some non-native contrasts. The occurrence, nature and extent of such interference are dependent on how the foreign segments are assimilated to the native segments according to the similarities and differences from L1 pronunciation. PAM focuses on the way non-native categories are filtered through native language phonetic properties, while SLM is concerned with cross-language speech perception as the basis for difficulties in second language speech production.

1.1 Speech Learning Model (SLM)

The Speech Learning Model (Flege, 1987, 1992a, 1992b, 1995) is concerned primarily with ultimate attainment in second language pronunciation, so work carried out within this framework focuses primarily on bilinguals highly experienced in their L2. It accounts for age-related limits on the ability to produce L2 vowels and consonants in the way monolinguals do. The model aims to explain why some segments of L2 will be more resistant to acquisition while others will be learnt quite easily.

The hypotheses within the model are based on the assumption that phonetic systems used in the production and perception of vowels and consonants remain adaptive over the lifespan, and reorganize in response to segments encountered in an L2 through the addition of new phonetic categories, or through the modification of old ones. During L1 acquisition, speech perception becomes attuned to the contrastive phonetic elements of the L1. Learners of an L2 may fail to discern the phonetic differences between pairs of segments in the L2, or between L2 and L1 because the L1 phonology filters out features (or properties) of L2 segments that are not phonologically relevant in L1. Second language learners need accurate perceptual targets to guide their L2 learning, and since experience with their L1s distorts the perception of L2 segments,
the production of L2 segments may be inaccurate. This is especially the case for late learners whose L1 system has been firmly established, i.e., late learners.

According to the SLM, the establishment of a new phonetic category is possible for an L2 segment that differs phonetically from the closest L1 segments if bilinguals discern at least some of the phonetic differences between those L1 and L2 segments. This possibility increases as perceived dissimilarity between L1 and L2 segments increases. However, as the age of learning increases, the likelihood of this difference being discerned decreases. In the case of strong perceived similarity between the L2 segments and the closest L1 segments, an L1 phonetic category will be used to process perceptually linked L1 and L2 segments and these two segments will resemble one another in production. In the case when a bilingual establishes a separate category representation for an L2 segment that is more dissimilar to the closest L1 segment, it may differ from that of monolinguals as the consequence of a bilingual’s maintenance of phonetic contrast between categories in a common L1 and L2 phonetic space. Alternatively, the bilingual’s representation may be based on different features or feature weights than that of monolinguals.

The effects of age of L2 learning on the production of the L2, combined with effects imposed by the characteristics of the L1 phonetic segments, have been revealed for both consonants and vowels (see Chapter 4 for an overview of previous research).

While the SLM takes into account the perceptual relationship between L1 and L2 segments, it fails to specify the exact principles for establishing how the L2 and closest L1 phones are perceptually related. Therefore, in order to explain the relative failures and successes in L2 segmental production, a better understanding of cross-language speech perception is needed. The relevant theoretical background on cross-language speech perception research for this thesis is the PAM, as hypotheses tested in Study 2 (Chapter 8) are derived from PAM’s predictions. In contrast to the SLM, which does not specifically predict which L2 segments are perceived as similar or different to the L1 segments, PAM gives specific cross-language perceptual assimilation patterns for non-native contrasts that are predictions of the discrimination accuracy of these contrasts.
1.2 Perceptual Assimilation Model (PAM)

The Perceptual Assimilation Model (PAM) is based on the principles of Articulatory Phonology (Browman & Goldstein, 1990a, 1990b, 1992a, 1992b) (see Chapter 2 for detailed overview of Articulatory Phonology) and the assumption that detection of native gestural patterns in speech guides and constrains listeners’ perception of gestural information in non-native phonetic categories. According to Articulatory Phonology, units of speech production are articulatory actions (gestures) which are defined as dynamical units specified with characteristic sets of formation and release of constrictions in the vocal tract (such as lips, tongue tip, tongue body, glottis, velum, etc.). Hence, the primary principle of the Perceptual Assimilation Model is that listeners perceptually relate non-native speech segments to the native segmental constellations that are in closest proximity in the native phonological space.

According to the PAM, this perceptual relationship between non-native speech segments and the native segmental constellations is based on the gestural similarity (or discrepancies) between non-native and native segments. When non-native segments have gestural organization that is similar to the gestural organization of one (or more) native categories, the listener will detect native gestural organization in the non-native segment. Thus, the non-native segment will be perceptually assimilated to the closest (most similar) native category (or categories). Furthermore, these perceptual patterns are reliable predictors of the discrimination difficulties that listeners may encounter when perceiving non-native contrasts. However, in contrast with the SLM which focuses on the perception of single phonemes and does not specify the perception of non-native contrasts, the PAM makes explicit predictions about assimilation patterns (and discrimination accuracy) of non-native contrasts.

In order to explain possible ways in which listeners perceive non-native phonetic segments, Best (1995) describes the native phonological inventory in terms of a range of speechlike gestural properties (native phonetic domain). Outside of this range of speechlike gestures are vocal-tract generated non-speech gestures (e.g. choks, laughs, whistles and so on). Listeners may perceive non-native segments with respect to the native phonetic domain as more or less similar to a native phonetic category (or categories). Furthermore, the non-native segment may be perceived as globally speechlike, but with a gestural organization that falls in between specific native
phonetic categories, making it an uncategorizable speech sound. Finally, the non-native segment may not be assimilated as a speech sound at all, falling in the non-speech space beyond the boundaries of the native phonetic domain. In the third case, the non-native segment is non-assimilable as speech and is perceived as a non-speech event.

Discrimination of non-native segments is predicted to vary systematically among six possible patterns of assimilation for the paired members of a non-native contrast. Three contrast assimilation patterns concern non-native segments that are consistently identified with native categories, and the remaining three concern contrasts in which either one or both segments are not identified within any single native category.

1. **Two-Category (TC Type)** assimilation occurs when the members of the non-native contrast are similar to two different native segments. In this case, each member of the non-native contrast is assimilated to a different native segment. According to PAM, listeners will easily discriminate between non-native categories that are distinguishable for their difference in gestural constellations that specify a difference between two native categories as well. Thus, discrimination between these two non-native categories is expected to be excellent.

2. **Category-Goodness Difference (CG Type)** assimilation occurs when both segments from the non-native contrast are assimilated to the same native category, but one is more similar than the other to the native category. The discrimination is aided by the listener’s recognition that one non-native phoneme is more discrepant from the standard native exemplar than is the other. CG contrasts may vary in the degree of discriminability. In such cases intermediate discrimination (between poor and good discrimination) is expected.

3. **Single Category (SC Type)** assimilation occurs when both segments from the non-native sound contrast are assimilated equally well or equally poorly to a single native category. In this instance, poor discrimination is expected, because the two phones are equally similar or discrepant from the standard exemplar of the category and thus difficult to discriminate.

4. **Uncategorizable (UU Type)** assimilation occurs when both segments from the non-native contrast fall within native phonological space, but in between specific native categories. Discrimination is expected to be poor to moderate, depending on whether these two categories bear any remote similarity to any native category (or categories)
and the extent to which these similarities overlap between the two non-native phonemes.

5. *Uncategorized versus Categorized (UC Type)* discrimination occurs when one non-native sound is assimilated to a native category, but the other constituent of the non-native contrast remains uncategorized to any native phonetic categories. Discrimination is expected to be very good.

6. *Nonassimilable (NA Type)* discrimination occurs when both segments of the non-native contrast are so discrepant from the native language properties, that they are perceived as non-speech sounds. They are perceived to fall outside of the speech domain. Discrimination is expected to be moderate to very good.

Hypotheses based on SLM are generated to account for the accuracy of L2 segmental production and perception by highly experienced L2 learners, whereas PAM was originally devised to account for non-native segmental perception. However, PAM’s assumption can also be applied to account for L2 segmental perception by highly experienced bilinguals (Best & Tyler, 2007). Both SLM and PAM agree that basic speech mechanisms remain intact across the life span. According to the SLM, language-specific aspects of speech segments are specified in long-term memory representations (e.g., phonetic categories), while according to PAM, the listener directly perceives the articulatory gestures of the speaker without forming long-term memory representations. Instead, when perceiving the speakers’ articulatory gestures, the listener detects higher-order invariants in speech stimuli through perceptual learning. Regarding the relationship between the L1 and L2, the SLM posits that L1 and L2 phonological categories exist in a common phonological space, but it does not include the dynamics of changes to this common space at the phonetic level (as PAM does). The SLM predicts that L2 learners can establish a new L2 phonetic category if they detect phonetic differences between an L2 category and the nearest L1 category. The greater the perceived phonetic distance between the L2 the L1 category, the greater chance that the phonetic difference between the sounds will be perceived and a new L2 category will be established. Therefore, according to the SLM, the L1 phonetic categories limit the possibility for establishing an L2 category. SLM, therefore, distinguishes only between new and similar L1 and L2 categories, while PAM proposes several assimilation processes (as described above).
For both PAM and SLM, the perceived distance between L1 and L2 is crucial. According to PAM, the perceived distance between L1 and L2 phonemes is based on the resemblance of articulatory gestures used in production of these two phonemes. The SLM does not offer a specific criterion for measuring cross-language phonetic distance. Both models agree that this phonetic distance must be tested in cross-language mapping experiments in which the L2 segment is perceptually related and categorized as instance of an L1 category, and then their similarity is rated.

The focus of this thesis is the influence of the first language and age of second language learning on perception and production of L2 vowels. However, in order to gain greater understanding of L2 vowel perception and production, it is important to understand how vowels are produced and perceived in the context of considering the differences between vowels and consonants. Therefore, overviews of production of speech segments and perception of speech segments are presented in Chapter 2 and Chapter 3. The research related to cross-language segmental production and perception, and the relevance of the SLM and PAM for the studies reported in this thesis, are presented in Chapter 4 and Chapter 5.
CHAPTER 2
PRODUCTION OF SPEECH SEGMENTS

The process of articulation and the vocal organs involved in both the articulation of vowels and consonants are discussed in this chapter. These articulatory descriptions are a convenient way to characterize the differences between speech segments. Understanding speech physiology sets the background for understanding the acoustic properties of speech production that are discussed in the final section of the current chapter. The standard model of the acoustics of speech production, the source-filter model (Fant, 1960), and a more recent perspective, Articulatory Phonology (Browman & Goldstein, 1992a), will be discussed. This discussion will lead into the chapter on segmental perception, where a primary goal is to offer a framework for how experience with one’s native language influences perception (and production) of non-native segments.

2.1 An overview of the speech production process

2.1.1 The role of vocal organs in speech production

The organs of speech (the lungs, trachea, larynx, the pharyngeal and oral cavities, and the nasal passages) form a vocal tract, which is employed in the production of speech. A general overview of the organs of speech is presented in Figure 2-1.
During the production of speech segments, the airstream from the lungs is exhaled through one or more speech resonators. The presence or absence of obstructions of the airstream determines the nature of the speech segments produced. This process requires that the vocal tract articulators work in a coordinated manner.

2.1.2 Respiration, phonation (voicing) and articulation

Speech production is often described in terms of the three processes: respiration, phonation (voicing), and articulation. First, during respiration, the lungs provide the energy source. In the phonation phase, the vocal folds convert this energy into audible sound. In the final stage, the articulators transform this sound into intelligible speech. A functional overview of the vocal tract model is presented in the Figure 2-2.

In the context of speech production, during the respiration process when the lungs are compressed and their volume is decreased, the air pressure in the lungs increases relative to external air pressure and the air is exhaled in order to equalise the pressure with the outside air. This airflow expelled from the lungs is used as an energy source in speech production. The level of respiratory activity during speech production is greater than in normal breathing, but varies with the degree of overall vocal effort.
used. The general tendency in speech is for the respiratory system musculature to maintain a relatively consistent level of pressure below the glottis, known as subglottal pressure. The volume and pressure of the air supply in the process of respiration determines the duration and loudness of the sound produced.

In the phonation phase, the air exhaled from the lungs flows out through the larynx where the vocal folds are situated. If the vocal folds are closed, the airflow is blocked, causing a build-up of air pressure below them. Eventually, the air pressure builds up and forces the vocal folds apart. As the air flows through the vocal folds, the pressure is reduced, pulling the vocal folds back together. When the vocal folds return to the closed position, the accumulation of the subglottal air pressure starts again and eventually forces the vocal folds apart. During these repeating cycles, there is a point of reduced air pressure when the vocal folds are pulled close together. At this point the air is released in short, periodic bursts, causing vibration of the vocal folds. This process of vocal fold vibration is known as phonation or voicing. All vowels and vowel-like

Figure 2-2. Functional model of the vocal tract (from Clark et al., 2007).
sounds are normally produced when these air puffs flowing through the vocal folds set them into motion, causing them to vibrate (except in, e.g., whispering).

The actual sound produced during phonation is created by the periodic chain of air puffs emitted through the vibrating folds, not by the vibrations themselves. However, these streams of air puffs are not identifiable speech segments: in order to produce intelligible segments, the articulators in the cavities above the vocal folds must be activated. In particular, the sound emanating from the larynx is shaped by the changing position of the articulators in the oral cavity (tongue, lips, teeth, velum or soft palate) relative to each other. The movement of these structures is known as articulation.

2.1.3 Articulation and the source-filter theory of speech production

The coordinated actions of articulatory gestures result in the acoustic signal of speech, which listeners use to recover the linguistic message contained in the signal. The acoustic behaviour and properties of the human vocal tract in speech production are traditionally considered in terms of the source and filter model (Fant, 1960) presented in Figure 2-3.

According to Fant’s model, the speech signal can be viewed acoustically as the result of the properties of the sound source, modified by the properties of the vocal tract functioning as a frequency-selective filter. The properties of the filter are determined by the geometry of the vocal tract, which itself changes shape as the articulators change positions. Both the properties of the source and those of the vocal tract can be varied, and are varied continuously during speech. Speech, as well as any other type of sound, consists essentially of variations in air pressure due to movement in the event causing the sound. This motion will cause a disturbance in the surrounding air, which will spread outward in the form of sound waves. For every sound produced there must be a corresponding motion of the sound source. The vocal tract, open at the mouth and nostril at one end, and closed at the other (terminated by the larynx), acts like a complex resonator. The incoming airflow sets the column of air within the resonator into vibrating motion. This vibration causes the sound wave we hear. In other words, the vibrating air in the vocal tract causes backward and forward movements of the air particles at the opening between the lips. It is these movements that radiate variations in air pressure, causing vibrations of air outside the lips. The waveform produced beyond the lips will be repeated the same number of times in a second as the number of times
the vocal folds are blown apart, for voiced sounds (i.e., those involving vocal fold vibration). The vocal fold source, the actions of the vocal tract, and the further modifications by the radiation factor determine the output produced at the lips. The filtering action of the vocal tract will be different for each position of the articulators in the oral cavity.

Figure 2-3. A source and filter view of speech production (Ladefoged, 1996).

2.2 Vowels as speech segments: Articulation and acoustics of vowel production

When vowels and vowel-like segments are produced, the articulators in the oral cavity change their position relative to each other, but the airflow is uninterrupted. On the other hand, in the production of consonants the airflow is radically constricted or temporarily blocked by articulators in the oral cavity. Vowels in English are normally produced with an oral airstream: the velum is raised, blocking the nasal cavity and allowing the airflow to pass only through the oral cavity. If the velum is lowered, a vowel is nasalised as the air also flows through the nasal cavities. In English vowel nasalisation is not contrastive as it is in French, rather it is a consequence of the vowel occurring in a nasal consonant (m, n) context; these sorts of coarticulatory effects are discussed in section 2.3. The most significant modification in the production of vowels is the change of shape and position of the tongue body (raising, lowering, fronting or retracting), and the shape (spread or rounded) and degree of protrusion of the lips.
For example, when the vowel in the word heed ([iː]) is produced, the air puffs flowing through the vibrating vocal folds are directed through the oral cavity while the velum is raised blocking the nasal cavity. However, the articulators in the oral cavity also move, further altering the perceived quality of the vowel: the tongue body moves forward and upward towards the hard palate, and the lips are spread. During this process, there is no obstruction or impediment to the oral cavity airflow.

Traditionally, vowels are classified as monophthongs (single target vowels) and diphthongs (dynamic vowels that shift from one target vowel to a second one).

2.2.1 Monophthongs

Vowel articulation as previously discussed applies to vowels that have steady state articulation in which the tongue, lips and jaw (their acoustic correlates are explained in section 3.2.3) achieve one relatively stable target configuration. This is the characteristic of monophthongs. However, in running speech there is always some kind of articulatory movement at the beginning or end of a vowel when the articulators are moving from or towards the next articulatory target. The movements at the beginning of the vowel may be caused by the lips or tongue moving away from the configuration of the previous consonant, whereas the movement at the end of the vowel can be due to anticipation of the articulatory positions required for the production of the following consonant. This coarticulatory movement plays an important role in speech perception even when the vowel is identified as having a stable articulation as in monophthongs.

In production of monophthongs the two most important articulatory properties are the shape and the position of the tongue, and the shape and degree of protrusion of the lips.

*Tongue positions (vowel height and fronting)*

Traditionally vowels are plotted in a two-dimensional diagram representing the articulatory space with respect to the tongue position (front or back) and height (low or high). Figure 2-4 represents Australian English vowels plotted according to the tongue position: the first formant (F1) correlates negatively with tongue, or vowel, height and the second formant (F2) correlates with tongue (vowel) fronting. The data were obtained in the acoustic studies by Bernard (1970) and Cox (1996).

In this diagram, the vowel [i:] (as in heed) is plotted in the highest position near the y axis. This defines the articulatory position of the [i:] as a high and as a front vowel. In contrast to [i:], in the production of the vowel /ɔ/ (as in hod) the tongue height is lower and the tongue body is retracted toward the back of the oral cavity.

Lip postures (rounding)

In the production of vowels, the lips can have a spread, neutral or rounded shape. The differences between spread and neutral lip positions are related to the vowel height: high front vowels tend to have spread lips (as the Australian English vowel [i:] in heed), lower vowels tend to have more neutral posture (as the vowel [ɔ] in hod). The spread and neutral lip posture are united as unrounded, given that spread and neutral positions are not normally a distinctive difference in languages. Contrasting with the unrounded lip position is a rounded lip position (which sometimes includes a degree of protrusion). Lip rounding is not a distinctive difference for any given tongue position in English, but some languages exploit this distinction and contrast rounded and unrounded vowels of the same tongue height and front-back position in their vowel inventories (e.g., French and German distinguish between front rounded and unrounded
vowels of the same height whereas Vietnamese distinguishes back rounded and unrounded vowels of the same height).

### 2.3 Diphthongs

During the articulation of some vowels the tongue (and/or lips) moves and the auditory quality of the vowel is changed within a single syllable (for example, /ɔi/ as in *boy*). This change in vocal tract configuration transitions during the production of diphthongs results in two identifiable targets (rather than one as in monophthongs). Therefore, the articulation of diphthongs is explained in terms of tongue movement rather than a single tongue position. In production of diphthongs, the tongue moves either from a mid or low to a high position (*closing or outgliding diphthongs*), as in the vowel [ai] in *high*, or from peripheral to central position (*centring or ingliding diphthongs*), as in the vowel [ɪə] in *hear*.

### 2.4 The acoustic characteristics of vowels

The specific characteristics of each vowel sound are a direct consequence of the momentary attributes of vocal tract size and shape. This process can be described in terms of the source filter theory of vowel production (see Figure 1-3).

#### 2.4.1 Vowel formants

Vowels are most frequently described with reference to their formant characteristics, which are related to the articulatory shape of the vowel and the location of the corresponding resonances within the spectrum (frequency domain).

For each position of the vocal organs there is a corresponding characteristic mode of vibration (resonance) of the column of air in the vocal tract. These modes of vibration of the vocal tract are known as formants. The information-bearing formants of the speech spectrum are conventionally numbered upwards from the lowest in frequency (F1, F2, etc.). Theoretically, there is an infinite number of formants, but the three lowest are the essential parameters in the description of vowel quality. Thus, the formants of a vowel are directly dependent on the shape of the vocal tract and are largely responsible for its characteristic quality.
Each formant can be described by two characteristics: centre frequency (formant frequency) and bandwidth (a measure of the breadth of energy across the frequency domain, or a measure of the rate of damping in the time domain).

Formant frequencies depend on three factors: backward and forward movement of the tongue (the front-back position of the point of maximum constriction in the vocal tract), distance of the tongue body from the roof of the mouth (openness) and from the back of the throat, and the position of the lips.

2.4.2 Vowel duration

Vowels are also distinguished in terms of duration. The difference between long and short vowels is relative rather than absolute, as contextual and prosodic factors affect the ultimate length of the vowel. Openness (Lisker, 1974), vowel height (Lehiste, 1976), syllable stress, speaking rate, voicing and places of articulation of following and preceding consonants (Klatt, 1976) are positively correlated with vowel length. For example, if greater articulatory movement for a vowel articulation is needed (as in the case of low vowels in English), the vowel tends to be longer. Similarly, if the consonant following the vowel involves tongue movement, more time will be needed to establish the consonantal articulation and the adjacent vowel will be longer. Vowel length differences in English are associated with the difference between tense and lax vowels. In general, tense vowels have higher muscular tension (Trask, 1996), more extreme movements of the articulators, longer duration and higher subglottal pressure than lax vowels. In Australian English, for example, a difference between a tense and a lax vowel is the difference between /i:/ in heed and /i/ hid. The tense vowels are sometimes called advanced tongue root vowels (ATR) (Halle & Stevens, 1962) as they are produced when the tongue root is moved forward (advanced), causing the widening of the cross-sectional area of the pharynx. Vowel length is the acoustic correlate of this articulatory movement. However, tense and lax vowel pairs often differ in colour (height and/or fronting) as well as duration and tenseness. For example, in the tense-lax vowel pair [i: i] the lax vowel [i] is shorter, lower and slightly more centralized than the tense vowel [i:]. It is important to note that there has been much discussion about the tense/lax distinction and whether it is valid in the way front/back and high/low are considered (Ladefoged & Maddieson, 1996).
In the Serbian language, on the other hand, vowel length is exploited more systemically: every short vowel has a matching long counterpart. This length difference is not associated with differences in vowel tenseness or colour: the long-short pairs have essentially the same colour and tenseness. The distinction between them is known as a distinction between long and short vowels, although this is often neglected in the literature. The vowel duration difference in Serbian is morphologically conditioned. For example, short Serbian /e/ and /i/ in the genitive singular case become long /e:/ and /i:/ in the genitive plural case (e.g. ime [name] in singular becomes imena [names] in plural; /e/ in singular is short whereas /e/ in plural is long).

2.4.3 Distinctions between consonants and vowels

The main articulatory difference between vowels and consonants is the absence versus presence of a constriction that blocks or impedes the airflow. Consonants are grouped into different classes dependent on the place of constriction, the type of constriction, and the presence or absence of voicing during the consonant release. In production of stop consonants, for example, the airflow is temporarily blocked or obstructed by a complete closure of the vocal tract at some location by one or more articulators. In production of fricatives, on the other hand, a narrow constriction in the vocal tract creates turbulent airflow. By contrast, when vowels are produced, the airflow is uninterrupted by the motion of articulators in the vocal tract. In the case of approximants (such as /w/) the articulators form a constriction greater than in vowels, but the constriction is too wide to block the airflow or to cause any airflow turbulence at the point of constriction (as in production of stop, fricative and nasal consonants).

2.5 Phonetic variability: Coarticulation in speech production

Articulation of speech segments thus far has been discussed without reference to continuous speech. In normal continuous speech some articulatory processes occur very rapidly, interacting with each other. These interactions cause changes in the acoustic properties of the segments involved, due to coarticulation (overlapping production of adjacent phonetic segments). For example, at the beginning of production of a vowel, the lips and tongue may move away from the configuration of the previous consonant, whereas at the end, the movements of the articulators may be caused by the anticipation of the configuration of the following consonant. The following consonant may also
cause the lengthening of the preceding vowel. In English, vowels are longer when followed by voiced consonants; for example, the vowel [i:], as in the word ‘heed’, is longer than the same vowel preceding the final voiceless consonant in the word *heat*. This vowel lengthening before voiced consonants is not a language universal phenomenon, (Katz & Bharadwaj, 2001) although it is believed that when articulation of the consonant requires tongue movement, the vowel length is affected as the tongue moves toward the position required for the articulation of the consonant.

Not only vowel duration, but vowel quality is affected by adjacent consonants (Clark et. al. 2007) When the consonant following an oral vowel is nasal, the vowel is nasalised rather than being purely oral (e.g. *bet* vs. *bent*). The adjacent vowel also affects the articulation of the preceding consonant. For example, in the word *key* the [k] is articulated with the tongue placed in a more fronted position, anticipating the articulation of fronted vowel /i/, whereas in the word *cot* the initial consonant [k] is produced with the tongue body retracted in anticipation of the articulation of the low vowel [a]. This context-sensitive variability in speech production is known as coarticulation. It is caused both by the biomechanical properties of the vocal tract and by the nature of the neuromuscular control mechanisms operating the articulatory movements, which are to some extent determined by language-specific differences. In both cases, this interpretation of the causes of coarticulation is twofold: it may imply that the vocal tract and articulators have limitations in their speed of movement or it may imply that in speech production the articulators move just sufficiently to produce adequately intelligible phonetic distinctions.

Coarticulation is also observed as an overlap in time of articulatory movements. This is the case in the English words *burns* and *jaws*, where the fricative following the nasal is nasalised in the former word, and the alveolar fricative following the vowel in the latter word is produced with lips rounded because the preceding vowel is rounded.

Coarticulation is an important property of language. Anticipatory coarticulation occurs when articulation of a segment is affected by the articulation of the following segments, whereas perseveratory coarticulation occurs where the articulation of a segment affects the following segment. However, it is important to note that the phonetic differences in segments produced due to coarticulation are variations (*allophones*) of the contrastive speech segments (*phonemes*) and they have no contrastive meaning in the language. They are just different pronunciations of the same
segments. In some cases, however, due to coarticulation one speech segment can change an adjacent segment into another, contrastive speech segment (for example, in the word tissue the consonant [s] has been retracted in anticipation of the following [u] and became the palatal fricative /ʃ/).

To minimize the coarticulation effects in the studies reported in this thesis, the tested vowels /æ/, /e/, /i:/ and /ɪ/ always occurred in h_d context. /h/ is an aspiration that provides no spectral transitions from consonant to vowel, whereas final /d/ provides some dynamics and spectral transitions from vowel to consonant (the vowel to consonant transitions are less often the focus of studies on perceptual compensation for coarticulation). This h_d context was held constant for stimulus choices across the studies.

Coarticulation effects cannot be directly observed as the production of autonomous speech segments with certain invariant articulatory (and acoustic) characteristics and therefore a rigid model such as source-filter theory is insufficient to account for production of speech segments in connected speech. Furthermore, despite these variations in the production of the same phoneme, listeners do not perceive them as different phonemes. An explanation of the production of speech segments which accounts for the rapid articulatory movements and their overlap in speech production comes from the Articulatory Phonology perspective, which the Perceptual Assimilation Model (Chapter 1) is based on.

2.6 Articulatory phonology: Gesture as a basic phonological unit

Articulatory Phonology (Browman & Goldstein, 1990a, 1992a, 1992b) departs from the view that the units of speech production are phonemes, or acoustic cues. Instead, the basic units of speech production are articulatory actions, called gestures, which are defined as dynamical units specified with characteristic sets of formation and release of constrictions in the vocal tract (e.g., formed by the lips, tongue tip, tongue body, glottis, velum, etc.). The articulators that are involved in formation and release of these constrictions are organized in coordinated structures. The motion of these coordinated structures (tract variables) is necessary to achieve a functional goal for each individual gesture. Furthermore, there is a specific set of tract variables for each gesture. For example, the tract variable for lip protrusion is defined by the articulatory actions of
upper lip, lower lip and jaw, whereas the tract variable for tongue tip constriction location is defined by the articulatory movement of the tongue tip, tongue body and jaw. Furthermore, for each gesture two paired tract variable regimes are specified: one that controls the degree of constriction of a particular structure, and one that controls the location of the constriction. Tract variables and associated articulators that are hypothesized to be sufficient for characterizing most of the gestures in English are presented in Figure 2-5.

![Figure 2-5. Tract variables and associated articulators (from Browman & Goldstein, 1992a).](image_url)

Since constriction location and degree are two dimensions of the same constriction, they are considered related tract variables. For example, lip protrusion and lip aperture are two dimensions of the same constriction and are achieved by the movements of upper and lower lips and jaw; tongue tip constriction location and constriction degree are both achieved by the movement of tongue tip, tongue body and jaw.

A speech utterance is described as an ensemble of a small number of overlapping gestures, or gestural score. These gestures are phonological events and are the basic units of contrast between lexical items. For example, *pan* differs from *ban* in having a glottal opening for voicelessness of the initial gestural constellation, whereas velum, tongue tip, tongue body and lips are involved in the common gestures. *Bad* and
*dad*, on the other hand, are contrastive in that the initial gesture of *bad* involves lip closure whereas *dad* involves tongue tip closure.

Articulatory Phonology has served as the basis for an argument (Best, 1995) that in speech perception listeners directly perceive information about articulatory gestures and constellations. In non-native speech perception, listeners are influenced by gestural constellations in their native language. This is the main premise of the Perceptual Assimilation Model (Best, 1995) (Chapter 1). Before turning to non-native segmental perception and production, a brief overview of the main issues in segmental perception will be given in Chapter 3.
CHAPTER 3
PERCEPTION OF SPEECH SEGMENTS

In the preceding chapter, the basic properties of speech segments were introduced with an overview of the speech production process and the problem of coarticulation in speech production. This chapter will describe the perception of speech. When a listener is trying to identify a phonological element in speech perception, he faces a highly variable, context-dependent relationship between phonological structure and the acoustic signal. Despite these problems, native speakers of the same language are able to group speech signals into the same phonetic categories, even though there are dramatic variations within each category. The chapter starts with an overview of categorical and continuous perception, particularly emphasising perception of vowels, since the perception of vowels is of particular interest to the research reported in this thesis. That will be followed by a review of the relevant theories of speech perception.

3.1 Vowel perception

3.1.1 Categorical perception

In categorical speech perception, a continuum of speech segments varying along a physical dimension is perceived in terms of distinct categories. Defining characteristics of categorical speech perception are high predictability of discrimination from identification and a consequent discrimination peak at the category boundary.

In early studies on detailed phonetic aspects of speech perception (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Liberman, Harris, Hoffman, & Griffith, 1957) undertaken in Haskins Laboratories, synthetic computer-generated speech stimuli were used with the aim of investigating how the cues of perceived phonological input are encoded from the acoustic signal. To do this, a series of synthetic consonant-vowel (CV) tokens was generated by varying acoustic parameters in equal increments along a continuum from a value encoding one consonant category to the value encoding another contrasting consonant category. When exposed to these incremental changes in the parameter(s), listeners reach a point of sudden change in their perception from one consonant to the other. For example, in one experiment the acoustic unit of
measurement for voicing (voice onset time, or VOT) in consonants was manipulated. Voicing refers to vocal fold vibration (voiced) versus absence of vibration (voiceless) during the supra-laryngeal articulation of a consonant. VOT is defined as the period between the release of a complete obstruction of the airflow by active articulators (in production of stop consonants) and the beginning of vibrations of the vocal folds (Abramson & Lisker, 1967). When VOT is delayed in 10ms steps from 0 to 100 ms after the release of a stop consonant occlusion, English speaking listeners heard the stop as voiced up to 20 or 30 ms VOT and then as a voiceless at the 40 ms delay and beyond. Furthermore, around the VOT value where listeners changed their identification of stimuli from voiced to voiceless, they were also able to discriminate between voiced and voiceless stimuli without difficulties. On the other hand, within the stimuli that listeners identified as instances of the same consonant voicing, the discrimination between different stimuli was poor. That means that discrimination is reliable between two perceptual categories (identified as two different consonants) and poor within the boundaries of one perceptual category (stimuli identified as instances of the same consonant). The high predictability of discrimination from identification and a consequent peak in discrimination around an identification boundary was found in other studies investigating categorical perception. Categorical perception is presented in Figure 3-1.

![Figure 3-1](image.png)

**Figure 3-1.** Idealized form of categorical perception showing identification performance for two categories and discrimination performance between categories (From Pisoni,
The identification function (left y-axis) shows steep slopes at the category boundary and consistent labelling of the items within a category. The discrimination function indicates a peak in discrimination accuracy at the category boundary (right y-axis). Stimulus number along the continuum is presented on x-axis.

### 3.1.2 Continuous perception

Categorical perception is not so clear for vowel continua (Fry, Abramson, Eimas, & Liberman, 1962; Repp, Healy, & Crowder, 1979). When F1 and F2 of vowels were varied in equal steps (Fry et al., 1962) in similar manner to in the studies of VOT described above, with the same experimental techniques, cross-category and within-category instances were discriminated equally well. There was no categorical peak in discrimination. The subjects were able to identify three distinct vowel categories /e/, /æ/ and /u/, but their identification boundaries were shallow and inconsistent. This effect is known as continuous perception as opposed to categorical perception.

Therefore, vowels seem to differ from many consonants in being identified gradually along a continuum of values rather than categorically. This difference in perception may be related to the differences in articulation of consonants and vowels (Strange, 1998a, 1998b). Consonants are produced categorically: articulation of consonants requires a specific, definite place of constriction (e.g., closure at certain places for the stop consonants, or open glottis for articulation of voicing and closed glottis for articulation of voiceless consonants). Vowels are produced less categorically: for example, there is no one definite position of the tongue tip in production of front vowels, which may take different positions from very high and very fronted /i/ to slightly less high and/or less fronted /ɪ/, which may vary somewhat continuously from /u/.

However, one determining factor in discrimination along a vowel continuum is language experience. For example, American English and Swedish listeners showed clear between-category peaks in discrimination on /i/- /ɪ/- /e/ (unrounded vowels) and /i/- /y/- /u/ (rounded vowels) continua (Stevens, Ohman, Studdert-Kennedy, & Liberman, 1964), and did not differ in identification of category boundaries on the unrounded vowels continuum. On the rounded vowels continuum, on the other hand, American English listeners were less consistent in identifying category boundaries than Swedish listeners. This difference was interpreted to be due to the fact that front unrounded
vowels occur in the phonological systems of both Swedish and American English, whereas front rounded vowels exist in the phonological system of Swedish, but not in English.

Given the evidence for well-defined within-category discrimination of vowels (Beddor & Strange, 1982), it is generally agreed that vowels are perceived more continuously than consonants. Perceptual cues like rapid transitions in the character of the spectrum, periods of silence, and change in formant transitions into/out of adjacent vowels, may all be cues that perceivers use when identifying consonants. However, the remaining question is what type of information enables listeners to recognize members of each vowel category and distinguish them from one another, in natural speech?

Other factors that may affect the perception of vowels are the use of F1 vs. F2 as cues to vowel characteristics, differences in the step size of the varied F1/F2, the memory load in different discrimination tasks, as well as the consonantal context in which the tested vowels are presented (Rosner & Pickering, 1994).

3.1.3 Formants (F1, F2 and F3) in perception of vowels

According to the early research on vowel perception, the primary determiners of vowel quality are the frequencies of the first two formants, F1 and F2, which correspond to vowel height and vowel fronting, respectively. This view was supported by evidence that native English speakers were able to identify two-formant isolated vowels accurately (Delattre, Liberman, Cooper, & Gerstman, 1952). Furthermore, when the regions of F2 values were masked, the English speakers based their perceptual identification judgment on the remaining values of F1, thereby confusing front and back vowels. Masking that principally affected the F1 values resulted in identification errors based on reliance on the remaining F2 values: confusion occurred between differing vowel heights within the back vowels and within the front vowels. Considering the significant variability and overlaps in F1 and F2 values in naturally produced speech tokens (Cox, 1998), an obvious problem that arises from an attempt to explain vowel perception using only the F1/F2 dimension is how listeners determine the boundaries between different vowels whose F1 and F2 overlap. Two proposed methods for determining the boundaries between different vowels in F1/F2 space are establishing equivalence classes for vowels based on the acoustical measurements of F1 and F2 values (Broad, 1976), and establishing the boundaries using identification tests (Rosner...
It was proposed that in addition to the F1/F2 dimensions, F3 (Fujisaki & Kawashima, 1968; Shepard, 1972) and F0 (Fujisaki & Kawashima, 1968) might play a determining role in perception of vowel boundaries.

In all these early studies, the formant frequencies were extracted from the acoustic vowel nucleus (the core, most intense portion of the vowel assumed to be the least affected by the phonetic context). However, it has been shown that in identification of vowels listeners rely substantially on the dynamic coarticulatory information in the syllable (Nearey, 1989; Strange, 1998a; Strange, Jenkins, & Johnson, 1983). Vowel identification was accurate even when the steady state vowel target was removed and listeners were presented with the initial and final transitions only. We turn next to findings on perception of dynamic aspects of vowels.

### 3.1.4 Dynamic properties in vowel perception

The traditional approach to perceived vowel quality as represented in the categorical perception studies is based on representing and investigating vowels by a single spectral cross-section taken from the nucleus of vowel, which is then made into a steady-state isolated vowel for generation of synthetic vowel continua. The limitation of this approach is that vowel duration differences and dynamic change in frequencies and fundamental frequency (F0), which play a role in identification of natural vowels, are ignored – although listeners are able to identify vowels correctly when only brief onglides, brief offglides and “silent centres” (vowel centres gated out) are presented (Strange et al., 1983). Listeners’ ability to identify vowels lacking the target steady-state spectral characteristics of the nucleus was further supported in more recent studies (Hillenbrand & Gayvert, 1993a, 1993b; Hillenbrand & Nearey, 1999). Taken together, these results suggest that perception of vowels relies strongly on the dynamic articulatory movements embedded in the coarticulatory transitions between vowels and consonants, and certainly more so than on the steady-state “target” values of the vowel nucleus. This view is consistent with Articulatory Phonology (Browman & Goldstein, 1990a, 1992a, 1992b) as described in the preceding chapter, which posits that the units of speech production are not acoustic targets but dynamical articulatory actions, or gestures, with characteristic sets of formation and release of constrictions in the vocal tract.
3.2 Perception of articulatory gestures

Phonologically intended speech segments often become merged together due to coarticulation in production, resulting in the occurrence of context-sensitive phonetic segments. The acoustics of these segments, therefore, are a function of the position and movement of the articulators between adjacent (or sometimes even non-adjacent) segments/gestures. An important question in speech perception research is how listeners reverse this coarticulation process and map the acoustic signal to the corresponding phonological segments, regardless of variability in acoustic signals for the same segment. The main problems listeners face are the lack of linearity, invariance and segmentation. Linearity implies that for each perceived phoneme there is a discrete corresponding sound in the produced utterance. In speech, however, this is not possible: the acoustic features of the phonemes are carried and stretched over preceding and following phonemes. Furthermore, these acoustic features of phonemes are not constant, but vary with the phonetic context, the rate of speech, and speaker differences, all causing problems in segmentation of the speech signal into defined, context independent linguistic units (phonemes, syllables and words). These context-sensitive acoustics are evidence that articulatory gestures, rather than acoustic patterns, are the objects of speech perception. This is the main assumption of both the Motor Theory (Liberman, Cooper, Shankweler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985), and the Direct Realist Theory of Speech Perception (Fowler, 1986, 1989, 1990, 1991, 1996; Fowler & Brown, 2000), with the latter being of special importance for development of the Perceptual Assimilation Model (PAM, Best, 1995) (Chapter 1).

3.2.1 The Motor Theory of Speech Perception

The Motor Theory is a phonetically based model of speech perception proposed by Liberman and his colleagues (Liberman & Mattingly, 1985). It is based upon the finding that perceived phonemes are not firmly coupled to the acoustic properties of the speech signal.

The basic assumption of the Motor Theory is that we perceive the articulatory movements of the speaker conveyed by the speech signal, rather than acoustic or auditory events. These articulatory movements that listeners decode are the neuromotor commands, or intended gestures, to the articulators (e.g., tongue, lips and vocal folds). While the mapping between acoustic signal and linguistic unit is not straightforward,
the mapping between intended gestures and perception is. More importantly for speech perception, the intended gestures are constant even when their acoustic outcomes are variable.

The Motor Theory strongly links speech production and perception. The process of speech production, according to Liberman and his colleagues, is characterized by a series of links between phonemes, neuromotor commands, muscle contractions, vocal tract shapes and acoustic signals. Phonemes are assumed to stand in an approximately one-to-one relationship with neuromotor commands and muscle contractions. The relationship between muscle contractions and the shapes of the vocal tract is very complex because of the coarticulation of adjacent vowels and consonants. Because the relationship between vocal tract shapes and acoustic signal is not one-to-one, the complexity of mapping between phonemes and speech sounds is perceptually attributed to coarticulation.

According to the Motor Theory, speech perceivers also produce speech. Specifically, it claims that the listeners have knowledge of how to produce individual isolated speech segments and of the rules governing how the speech segments become coarticulated in normal running speech. This knowledge is then invoked in mapping a speech signal into phonological segments. The signal can be analysed this way because the listener can synthesize the syllable just heard using a special, speech-specific module in the brain. This module, specifically developed for this task, is a part of the larger innate biological specialization for language that is unique to humans.

The claim that speech is unique to humans has engendered the main challenges to the Motor Theory (Diehl, Lotto, & Holt, 2004). For example, studies in categorical perception have shown that within-category discrimination and the sharpness and position of between-category boundaries might be common to human and non-human listeners. Furthermore, another piece of evidence against the claim that speech is special comes from investigations of duplex perception of speech (Liberman & Mattingly, 1985) and non-speech stimuli. When the formant transition of the second formant of /da/ and /ga/ syllables was presented to one ear and the rest of the syllable to the other at the same time, listeners experience a dual percept: a nonspeech percept corresponding to the formant transition alone and a speech percept corresponding to the syllables /da/ and /ga/. This phenomenon was interpreted by Liberman and Mattingly (1985) as the result of the activation of both speech and nonspeech processing modules. However, the
same effect was achieved when the perception was tested with nonspeech sounds (Fowler & Rosenblum, 1991), indicating that duplex perception does not provide evidence for the Motor Theory. For example, the duplex perception was demonstrated for the sound of slamming door, with the high-frequency portion of the signal presented to one ear and the rest of the signal presented to the other ear.

### 3.2.2 The Direct Realist Theory of Speech Perception

The Direct Realist Theory of speech perception proposed by Fowler (Fowler, 1986, 1989, 1990, 1991, 1996; Fowler & Brown, 2000), like the Motor Theory of speech perception, states that the objects of speech perception are articulatory events or vocal tract gestures, rather than auditory or acoustic events. The Direct Realist view that listeners directly perceive information about articulatory gestures in speech is integrated in the PAM. The basic premise of PAM is that when perceiving non-native speech segments, listeners actually perceive similarities and dissimilarities between the native and non-native gestural constellations. Perceived gestural similarity between a non-native and a native category (or categories) leads to perceptual assimilation of the non-native categories to the most similar native category.

In contrast to the Motor Theory, however, the Direct Realist Theory of Speech Perception does not regard speech as a part of special module unique for humans; instead, rather similar perceptual strategies to those used in visual perception, for example, are deployed in the perception of speech. Therefore, the gestures in speech perception are not speech- or language- specific, but rather intrinsically non-linguistic. These actual vocal tract gestures shape the acoustic signal, not the intended gestures (neuromotor commands) as the Motor Theory claims. Thus, according to the Direct Realist theory, listeners do not refer to their own articulatory commands in order to perceive the speaker’s articulatory gestures. The acoustic signal is parsed into responsible gestures in the information medium that the listener uses to recover the gestures. The relationship between the signal and the percept is direct, which means that perception is an immediate experience of an object and/or event in the environment and it is not cognitively mediated. The gestures in the Direct Realist Theory are the same as defined in Articulatory Phonology (Browman & Goldstein, 1992a) and are the basis of PAM’s predictions about non-native segmental perception.
The explanation of how coarticulation is perceived according to the Direct Realist point of view is that adjacent gestures overlap in time and acoustic signals show influences of two or more overlapping gestures. Since each gesture lasts longer than a unique section of an acoustic signal itself, acoustic information for that single gesture becomes weaker during the duration of the gesture. At first it coexists with another stronger gesture, then it is the main information of the next acoustic segment that becomes dominant, and then it wanes in the following segment. Since these co-produced gestures constitute the acoustic signal in independent (but momentarily overlapping) manner, perceivers do not have difficulties recovering corresponding gestures from the acoustic signal. Characteristics of overlapping gestures, therefore, remain preserved in the signal.

In summary, when retrieving linguistic units from acoustic patterns of speech listeners face an amazingly complex process, starting from the manner in which speech is produced. Two relevant speech perception theories that account for this problem, the Motor Theory and Direct Realist Theory of speech perception, both posit articulatory gestures as the objects of speech perception rather than acoustic or auditory events. This view is in agreement with Articulatory Phonology which describes articulatory gestures as the basic units in speech production.
CHAPTER 4

SPEECH PRODUCTION IN A SECOND LANGUAGE

This chapter provides an overview of research on age effects in L2 learning, particularly emphasizing the language interaction perspective. Evidence from L2 segmental productions and their interpretation within the Speech Learning Model reveal possible limitations of the model that provide a preface to the following chapter on second language speech perception (Chapter 5).

4.1 Effects of age on second language learning

It is widely agreed that age of L2 learning (AOL) affects performance in that language. Children are able to learn a foreign language seemingly effortlessly with similar outcomes as native speakers of that language, which contrasts with adults, who generally do not reach similar outcomes. The linguistic domain which is most affected in L2 learning is production of phonological segments, specifically the pronunciation of vowels (Flege, 1992b; Flege, MacKay, & Meador, 1999; Munro, Flege, & Mackay, 1996; Piske, Flege, MacKay, & Meador, 2002; Walley & Flege, 1999) and consonants (Bohn & Flege, 1992; Flege, 1991; Flege, Munro, & Mackay, 1995a; Flege, Munro, & Skelton, 1992; Flege, Takagi, & Mann, 1995; Munro, Flege, & MacKay, 1995). Failure of adult learners to achieve pronunciation proficiency that matches native speakers of that language is known as foreign-accented speech. However, age effects on performance in L2 have been evident in other language domains: syntax, morphology, grammaticality judgments, lexical decisions, and acquisition of suprasegmental features. This evidence comes from a large body of behavioural and brain imaging studies.

Bilinguals who learned their L2 beyond childhood (“late” bilinguals) tend to produce longer sentences than native speakers (Guion, Flege, Liu, & Yeni-Komshian, 2000) and bilinguals who learned their L2 in childhood (“early” learners). Early learners perform better on grammaticality judgment tasks (Birdsong, 1992; Johnson & Newport, 1989; Johnson, Shenkman, Newport, & Medin, 1996; Juffs & Harrington, 1995; McDonald, 2006; White & Genesee, 1996), which may imply age differences in processes involved in learning grammar (DeKeyser, 2005). Young and older L2 learners also differ in morphological performance in reading comprehension tasks (Jiang, 2004),
lexical processing (Silverberg & Samuel, 2004), reading aloud (Zevin & Seidenberg, 2004) and access to semantic representations (Izura & Ellis, 2004). The difference between early and late learners was further confirmed in brain imaging research (Kim, Relkin, Kyoung-Min, & Hirsch, 1997; Weber-Fox & Neville, 1996, 1999). For example, evoked response potentials (ERPs) of Chinese-English bilinguals during grammaticality judgements on a variety of written English sentences (Weber-Fox & Neville, 1996) revealed that adult learners were distinguishable from native speakers in some aspects of the timing and brain distribution of waveforms. Furthermore, adult learners showed less left hemisphere lateralization in syntactic processing than the bilinguals who learnt English at an earlier age (Weber-Fox & Neville, 1999).

4.1.1 Critical Period Hypothesis (CPH)

Age constraints on learning have been shown to be important in development in other animal species (Petrinovic, 1998), human learning (Shore, 1997), and human motor and cognitive functions (Mei, 1994). Animals, for instance, have to acquire certain skills by a certain age (Mailer, 1991) otherwise they will not be acquired in an appropriate way. If animals have critical periods (CPs) for many of their most important and even life-dependent behaviours, similar biological programs regulating the timing of language acquisition might also be expected in humans. If this is the case for human language development, a person must be exposed to a language during a certain period of development in order to acquire that language, otherwise age-related changes in the language-learning mechanisms will prevent the achievement of native-like competence (Lenneberg, 1967). Language learning after that period should result in a sharp discontinuity in language learning outcome. The proposed innate language learning ability has to be exploited following a maturational schedule specific to language and determined by brain lateralization in puberty. According to the CPH, after puberty, relative brain plasticity is lost and consequently language development capacity is reduced (Krashen, 1973; Lamendela, 1977). Although originally formulated to account for L1 acquisition, CPH has been applied to explain difficulties (or differences) that adults experience when learning an L2 (Bialystok & Hakuta, 1999; Bialystok & Miller, 1999; Johnson, 1992; Johnson & Newport, 1989, 1991). For example, Johnson and Newport’s (1989) “maturational state hypothesis” states that early in life, humans have
a superior capacity for language acquisition, which declines with age and prevents late learners from achieving native-like proficiency in an L2.

### 4.1.2 Age of L2 learning and L2 segmental production: towards an alternative explanation

If neurological maturation reduces neural plasticity, leading to diminished ability to learn an L2, it is expected that the age at which the language learning mechanism operates less effectively is similar for all members of the species. However, there is considerable disagreement about the age when language-learning abilities deteriorate. According to some researchers, language performance declines if the age of learning is greater than 6 or 7 years (Long, 1990; Pinker, 1994), while others claim the decline occurs at 15 years (Johnson & Newport, 1989), or between 6-7 and 16-17 years (DeKeyser, 2000) or puberty (Scovel, 1988), or even as early as 3 years (Weber-Fox & Neville, 1996). Also, there is no evidence of a sharp decline in performance after a certain age of learning. For example, native American English speakers’ ratings of perceived foreign accent in sentences produced by Italian-English bilinguals whose age of learning English ranged from 3 to 21 years (Flege, Munro, & Mackay, 1995b) revealed a strong negative relationship between the age of learning and accuracy in production (Figure 4-1).

![Figure 4-1. Mean pronunciation accuracy ratings (y-axis) for English sentences produced by native Italian speakers of English who differed in their age of learning English (x-axis) (from Flege, 1995).](image)
Another important implication of CPH is that the loss of language acquisition abilities is universal and underlies all language domains. The fact that adults learn the syntax of a new language more easily than children but retain a foreign accent in speaking suggests that a CP may exist for pronunciation but not for syntax (Eubank & Gregg, 1999). If this is the case, it may be possible that different language domains are governed by different processing systems (Bialystok, 1997; Neville & Bruer, 2001). Moreover, the influence of the L1 is evident in the acquisition of L2 stress (Dupoux, Peperkamp, & Sebastian-Galles, 2001; Peperkamp & Dupoux, 2002). This implies that structural similarity between the learner’s L1 and L2, rather than age alone, is a predictor of success in the L2 (Bialystok, 1997; Guion, Flege, & Loftin, 2000).

Similarly, some L2 phonetic segments are easier to acquire than others (Flege, 1987). If there were a limit set by the age of learning, all L2 phonetic segments would be equally difficult to learn for late learners. In addition, experience with an L2 may eventually change accuracy in production of L1 segments (Bohn & Flege, 1992; Flege, 1997; Flege, Bohn, & Jang, 1997; Yamada, Strange, Magnuson, Pruitt, & Clark, 1994). This finding implies that what appears to be an affect of age of L2 learning on pronunciation accuracy is actually a consequence of interaction between bilinguals’ L1 and L2 phonetic systems, which is consistent with SLM and PAM. Both models posit that the mechanisms for language learning remain intact over the lifespan, therefore being equally available to adults and children. What distinguishes adults from children learning an L2 is the extent of their experience with L1: adult learners have had greater exposure to their L1, and therefore L1 phonetic segments impose a stronger influence on the perceptual learning of L2 phonetic segments. PAM emphasizes that previous L1 experience (Best & Tyler, 2007) affects perceivers’ extraction of invariants regarding articulatory gestures from the speech signal. According to SLM, this perceptual influence of L1 determines L2 category formation in adult L2 learners.

Finally, cases of successful, highly motivated adult L2 learners who achieved similar levels of proficiency to those attained by early and native speakers (Bongaerts, 1999) indicate that some individuals can learn an L2 phonology to nativelike levels even in adulthood. That observation, and evidence of late L2 learners’ improvements in segmental production after L2 perceptual training (Akahane-Yamada, Tohkura, Bradlow, & Pisoni, 1996; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999) indicate that language learning abilities remain intact over the lifespan.
4.2 Production of L2 phonetic segments and language interaction

A more plausible explanation than the CPH for foreign accent in late L2 learners is the explanation based on SLM and PAM: that the two languages of a bilingual interact. According to this account, language learning ability does not change with maturation. Older learners retain the ability to achieve native-like proficiency in L2 speech production. What prevents them from native-like success in L2 pronunciation is their extensive exposure to and experience with their L1.

Studies investigating the nature and extent of mastery in L2 pronunciation have also focused on the segmental inventories of the L1 as a factor that may influence the L2 mastery of early as well as late learners. The L2 speech production data now available suggest that previous L1 learning affects subsequent L2 learning through the intermediary of central L1-biased cognitive-linguistic and phonetic knowledge. That is, interaction between L1 and L2 phonetic systems, rather than age itself, causes L2 segmental production inaccuracy in adult learners.

4.2.1 Interaction between first and second language phonetic systems

According to the interactionist perspective (Flege, 1999a), integrated in SLM and PAM, bilinguals are not able to fully separate phonetic elements of their L1 and L2, which exist in a common phonological space and mutually influence each other (Flege, Schirru, & MacKay, 2003). Because of shared phonological space, neither of the two languages can be fully deactivated and the phonetic elements of the L1 and L2 will influence one another in production (Grosjean, 1999). Those L2 segments that are perceived as similar to the closest L1 segments will be more difficult to learn to produce accurately in the L2 than perceptually different L2 segments. This interaction of L1 and L2 is reflected in the linear relationship between increasing inaccuracy in L2 production and increasing age of L2 learning: with increasing age of learning, the L1 imposes a stronger influence on L2. In other words, the linear decline in production accuracy may be analysed as an effect of L1 on L2 rather than as an effect of age (Flege, 1995). The more established L1 is at the time that L2 learning commences, the greater the influence it will have on the L2. As L1 categories are established and elaborated through childhood and into adolescence, the more likely it is that learners will assimilate L2 vowels and consonants to the L1 (Best & Tyler, 2007).
The unavoidable mutual influence between L1 and L2 (Flege, 1995; Jia & Aaronson, 1999) is bi-directional. That means that very young L2 learners may develop an authentic accent in their L2 but at the cost of not being able to develop/maintain a native accent in their L1. In the case of older learners, as the L1 phonology continues to be refined with age, its influence on L2 phonological acquisition continually increases.

This language interaction approach to L2 segmental production emphasizes the way second language learners perceive their L2 under the influence of L1 phonetic segments or vice versa. Flege (1992a, 1992b, 1995) has concluded that one of the main causes of foreign accent in L2 speech is learners’ tendency to perceive L2 segments in terms of their native categories, particularly in the case of those L2 segments that are perceptually close to L1 segments (see also Rochet, 1995). This is supported by evidence of improvement in production (Akahane-Yamada et al., 1996) after participants have received training in perception of L2 segments.

4.2.2 Production of L2 vowels

The process of SLM’s “equivalence classification” (Flege, 1987, 1995), when learners classify separate L2 phonemes as variants of a single L1 segment (Rochet, 1995), may be evident in the early stage of L2 learning. When perceived as variants of the L1 category, the L2 phones are produced according to the phonetic realization rules of the L1. For example, French /y/ for Portuguese speakers is categorized as Portuguese /i/ and their production of French /y/ is perceived by French listeners as belonging to the French /i/ or /y/ category (Rochet, 1995). The production of French /y/ by English speakers is perceived by native speakers of French as varying between French /u/ or /y/. This production pattern reflects English listeners’ perceptual categorization of French /y/ as English /u/. Furthermore, English learners of French often produce French /u/ as /y/, suggesting that they assign both French /y/ and /u/ to their /u/ category, which tends to be rather fronted.

According to the SLM, an L2 category may be established at any time across the lifespan, but with less success for late learners. This hypothesis was confirmed in several studies investigating production of English vowels by native speakers of Italian (Flege, Mackay, & Meador, 1999; Piske et al., 2002). For example, native speakers of Italian replaced English /ʌ/ with the nearest Italian vowel (/a/) (Flege, MacKay &
Meador, 1999; Piske et.al. 2002). This is mainly the case with older learners, although some older learners were able to pronounce English /ʌ/ accurately. However, even some early learners substituted Italian /a/ for English /ʌ/. In contrast to English /ʌ/ which is relatively close to the Italian /a/, English /e/ is unlike any Italian vowel and therefore, according to the SLM, expected to be produced accurately by both late and early learners. Contrary to the hypothesis, late learners’ production was significantly less accurate (Flege et al., 2003) than that of early learners. Early learners, on the other hand, did not differ from the native speakers.

4.3 Perceived distance between first and second language phonetic segments

The SLM and research within its framework emphasize the interaction between L1 and L2 in terms of perceived distance between L1 and L2 segments, or degree of similarity between them. In particular, L2 segments that are “similar” to L1 segments are difficult to acquire (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2003; Flege, 1987) because a speaker classifies or perceives them to be equivalent to those in his or her L1, whereas “new,” dissimilar or different segments are easier to acquire because the speaker readily notices the L1-L2 differences. The establishment of these “similar” and “new” segments is accomplished by two mechanisms: a category assimilation mechanism and a dissimilation mechanism. The category assimilation mechanism operates when a new category fails to be established, despite an audible difference between it and the closest L1 speech segments, because instances of the L2 category continue to be identified as instances of the L1 category. In that case a “merged” category will develop, resembling the properties of both the L1 and L2 speech segments. The category dissimilation mechanism, on the other hand, is thought to operate when a new category is established for an L2 segment. The new established category and the closest L1 category will shift away from one another in the common phonetic space as bilinguals strive to maintain phonetic contrast among all of the elements in their combined L1-L2 phonetic space.

Japanese learners of English (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004) can perceptually differentiate English /ʌ/ but not /l/ from Japanese /ɾ/. This suggests that the ability to perceive cross-language differences between English /ʌ/
and Japanese /t/ is what enables learners produce English /a/ more accurately than English /l/. That is, interaction of L1 and L2 segments is actually determined by their perceived similarity. This perceived similarity may change as experience with the L2 increases, causing beginner learners and experienced bilinguals to perceive and produce them differently (MacKay, Flege, Piske, & Schirru, 2001). However, late learners may alter their L1 categories to accommodate similar L2 segments whereas early bilinguals may do the opposite, creating merged L1-based categories for similar L1 and L2 segments.

It is not clear when two segments are close enough (or distant enough) to be difficult (or easy) to learn. Some phones, for example, have a direct counterpart in L1 (Flege & Hillenbrand, 1984) with which they can be identified. The first approximation is to include L1 and L2 phones that are transcribed using the same IPA symbol. According to this criterion instances of /t/ occurring in French and English words are expected to be regarded by the L2 learner as being different realizations of the same category. The interlingual identification of such pairs might be expected to occur even when acoustic differences that may distinguish the L1 and L2 phones are auditorily detectable. These categories are regarded as “similar”.

Other L2 phones bear less obvious resemblance to phones in the L1 and may therefore not be regarded as the realization of any L1 category. This type of L2 phone is “new.” For example, using the IPA symbol criterion, it would be expected that for English learners of French /y/ is a new vowel, since the English vowel inventory has two high tense vowels, front-unrounded /i/ and central-rounded /u/, but not front-rounded /y/. According to this prediction, French /y/ is expected to be produced more accurately than French vowels which have a clearer counterpart in English (such as French /u/). Acoustical analysis of the production of French /y/ and /u/ by English-French bilinguals (Major & Kim, 1999) revealed that bilinguals did not differ from French speakers in producing the new vowel /y/, whereas they produced /u/ with significantly higher F2 values (tongue positioned more forward) than the French speakers. Experienced learners produced a greater contrast between French /u/ and /y/ than inexperienced learners. On the other hand, in the study by Rochet (1995), native speakers of French identified the production of /y/ by English speakers as belonging to both French /u/ and /y/. This result suggests that English speakers may not actually perceive – or produce—French /y/ as a new category.
In addition to the phonetic transcription criterion as a means of identifying “new” L2 segments for L2 learners, Flege (1989) proposed the use of acoustic criteria. For example, using acoustic measures, Bohn and Flege (1997) tried to establish whether English /æ/ is a new vowel for German learners of English. According to the F0, F1, and F2 values at the acoustic midpoints of the German vowels /e/, /a/, /ɛ:/ and /e/, there is no vowel in German that occupies the spectral region of vowel space taken up by English /æ/. Therefore, it was hypothesized that experienced German learners of English would closely match native speakers of English in production and perception of this “new” vowel. In a perception task, the listeners identified the vowels from a /bet/-/bæt/ continuum. The vowel portion for each stimulus had nominal durations of 150, 200 and 250 ms. Native English speakers’ responses were not influenced by the duration of the stimuli: they showed a clear crossover from /bet/ to /bæt/ as vowel spectrum varied between the endpoint stimuli, which were unambiguously identified. German speakers, on the other hand, were influenced by duration when identifying the stimuli. They identified short stimuli mostly as /bet/ and long stimuli mostly as /bæt/. Stimuli of medium duration were perceived as ambiguous. Also, the experienced German speakers were more influenced by the spectral differences than inexperienced speakers. This suggests that English language experience affects how native German listeners perceived English /e/ and /æ/. These results were taken as support for the hypothesis that a new vowel category /æ/ would be established for experienced learners of an L2, who eventually learn to perceive and produce it in a way that is similar to native speakers of that language, although not exactly as native speakers. Furthermore, this finding was taken as consistent with the hypothesis that L2 segments without any obvious counterpart in the L1 are eventually established as experience with the L2 grows. However, the study does not provide data on how German learners perceive the targeted English vowel relative to their native vowels. Without the L1 perceptual equivalence data, the results do not relate to SLM’s basic assumptions that production inaccuracies are due to the perceptual relationship between L1 and L2 segments, and that certain L2 segments will be identified with L1 segments even when detectable acoustic differences between them exist.
PAM also deals with perceived distance between L1 and L2 phonetic segments. According to PAM, success in L2 segmental perception depends on how close the L2 and L1 segments are in terms of articulation. That is, listeners will detect similarities and dissimilarities between non-native contrasts and native segments based on perceived articulatory properties (e.g., constriction locations, active articulators, constriction degree, and phasing). Although it was not designed to account for L2 segmental production, PAM’s predictions may be extended to generate predictions concerning accuracy in L2 production, which will be attempted in Chapter 8.

It must be pointed out that the terms “new” and “similar” L2 phones are indicators of how L2 learners perceive the targeted L2 segments. Therefore, the types of perceptual tasks suited to testing potentially “new” and “similar” L2 phones should also play a role in establishing similarity/dissimilarity criteria (see Rochet, 1995). In addition to perceived distance between L1 and L2 phonetic segments, there are other factors affecting second language segmental production, which will be reviewed in the following section.

4.4 Other factors affecting second language segmental production

Research within the SLM framework has focused on bilinguals who are very proficient in their L2, with the aim of explaining why some segments are easier to acquire than others. In this research, age effects are actually just a summary reflection of the effects of environmental factors that are generally confounded with maturational factors: education (Flege, Frieda, & Nozawa, 1997; Hakuta, Bialystok, & Wiley, 2003), amount of L1 and L2 use (Flege, Frieda et al., 1997), motivation (Bongaerts, 1999), and input (Choi, 2000; Yeni-Komshian, Flege, & Liu, 2000; Yeni-Komshian, Robbins, & Flege, 2001).

Older learners in the studies on L2 acquisition have often been immigrants who may make behavioural choices that bring them into frequent contact with other native speakers of their L1 and limit their contact with speakers of their L2. This is the case with adult immigrants who limit their contacts with the L2 in order to avoid social isolation and preserve cultural identity. Young immigrants, on the other hand, receive their formal education in the language of their new environment and acquire L2 literacy skills in formal and structured contexts. This presence or absence of general L2 literacy skills (Bialystok & Hakuta, 1999) and language instruction in school (Bialystok, 1997)
may be linked to what appears as age-related differences in performance. Younger learners’ linguistic and cultural identity is less likely to be as fully formed as that of their elders, and so the motivation to maintain their native language may be weaker. For example, immigrants arriving at ages older than 10 tend to maintain their L1, whereas immigrants arriving before age 10 seem to switch their dominant language from the home language to the language of the host country (Aoyama et al., 2003; Jia & Aaronson, 1999).

Older learners and younger learners, for all these reasons, often differ in the amount of exposure to their second language (Aoyama et al., 2003), the time spent in the country where the target language is in use (Riney & Flege, 1998), time spent in the company of native speakers (Flege, Frieda et al., 1997), amount of L1 use (Flege, Yeni-Komshian, & Liu, 1999; Guion, Flege, & Loftin, 2000; Piske & MacKay, 1999), number of years of education in L2 (Yeni-Komshian et al., 2000), experience with their L2 (Izura & Ellis, 2004; Silverberg & Samuel, 2004), and type of input (Flege & Liu, 2001).

In summary, interpretations of the constraints in second language segmental production must take into account the state of development of the learner’s L1 at the time of acquisition. Learning an L2 in adulthood means that the new linguistic experience is being built on an already existing linguistic competence in a native language. The nature of language learning in childhood and adulthood differs according to the degree of establishment of the L1. This difference has a determining role in L2 learning: with age the L1 becomes better established, imposing an increasing influence on the acquisition of the L2. This is the factor that has to be taken into account when interpreting inaccuracy in L2 segmental production: the outcome of the L2 learning cannot be interpreted separately from the state of the development of the learner’s L1. The interaction between bilingual’s L1 and L2 is a dynamic process, changing over the course of development under the influence of other environmental as well as biological factors.
CHAPTER 5
CROSS-LANGUAGE SPEECH PERCEPTION

In this chapter, adults’ cross-language speech perception, and specifically the perceptual constraints that may influence inaccurate production of L2, will be reviewed. The chapter focuses predominantly on the notion of “language specific” speech perception, which has been posited to influence speech production in the way described in the previous chapter. This is of particular importance for the experiments reported in this thesis, as the hypotheses tested in Study 2 (Chapter 8) are derived from PAM’s predictions. Finally, the development of speech perception in infancy and the implication of these changes for second language learning beyond childhood will be briefly discussed. These changes may contribute to differences in L2 segmental production between young versus adult learners of an L2.

5.1 Adults’ cross-language speech perception

Adults are “language specific” perceivers in that they differentiate phonological distinctions contrastive in their native language easily and effortlessly. This ability reflects the selective attuning to the articulatory-acoustic properties of their L1, which appears to take place within the first year of life (Best, 1995; Kuhl, 1995). As a result, learning a foreign language in adulthood requires perceptual reorganization of well-established phonological categories and necessitates learning to differentiate phonetic categories that do not occur in the native language. Furthermore, adult learners of an L2 must learn to differentiate speech segments that have become functionally equivalent (that occur as allophones or free variations within a single phonological category in the native language). Some non-native contrasts (e.g. non-English voicing contrasts among initial stop consonants for English listeners) are relatively easy to perceive (Tees & Werker, 1984). On the other hand, perception of non-native place of articulation distinctions is more difficult for non-native speakers (e.g. perception of English /r/ - /l/ by native speakers of Japanese) (Badlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lively, Pisoni, & Logan, 1992; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Morrison, 2002; Yamada & Tohkura, 1995). Japanese speakers identify both English /r/ and /l/ as instances of a
single Japanese liquid /r/ (Sekiyama & Tohkura, 1993). Similarly, L2 vowels are perceived in terms of native vowels and the principles governing the L1 vowel system (Beddor & Strange, 1982; Flege, Munro, & Fox, 1994).

The well documented influence of the L1 phonetic system on L2 segmental perception plays a crucial role for both SLM and PAM. In contrast to the SLM, which does not specifically predict which L2 segments are perceived as similar or different to the L1 segments, PAM posits specific cross-language perceptual assimilation patterns for non-native contrasts that are predictors of the discrimination accuracy of these contrasts (see Chapter 1).

5.1.1 Evidence from the perception of non-native vowel contrasts

Research on cross-language vowel perception supports the view that L2 vowels are perceived in terms of native vowels and according to the principles governing the L1 vowel system (Beddor & Strange, 1982; Fox, Flege, & Munro, 1995). Although PAM’s discrimination predictions based on the assimilation types in research on vowels are not consistently reported (Best, Halle, Bohn, & Faber, 2003; Kingston, 2003; Lengeris & Hazan, 2007), the findings largely provide evidence for PAM’s predictions. There is also some evidence of multiple-category assimilation of non-native vowels to more than one native vowel (Morrison, 2003).

The effect of phonological and phonetic properties of native language on non-native segmental perception has been observed in American English, French and Danish speakers’ categorization of Norwegian /i/-/y/, /y/-/u/, /y/-/u/ and /u/ - /u/ contrasts (Best et al., 2003). English, French and Danish have large but different vowel systems (e.g. Norwegian has four high vowels, front unrounded /i/, out-rounded /y/, in-rounded /u/ and back rounded /u/; French and Danish have /i/, /u/ and /y/; English has only /i/ and /u/). The distribution of vowels in these languages also differs: Danish has a rich upper half of the vowel space (e.g. high vowels) compared to English and French, where vowels are more evenly distributed. Also, the shared vowels differ in their phonetic realizations (e.g. Danish and Norwegian /y/ is more fronted and articulated with less lip protrusion than French /y/; English /u/ is more fronted than in other three languages. Also, /i/ and /u/ are diphthongized in American English while Danish, French, and Norwegian /i/ and /u/ are monophtongs). It was hypothesized, based on
PAM, that English listeners would assimilate Norwegian /i/ and /u/ to native /i/ and /u/, but as poor examples, because of the diphthongized nature of /i/ and /u/ in American English, and more fronted position of /u/ in American English than in Norwegian. Furthermore, it was hypothesized that American English listeners would hear Norwegian strongly fronted /y/ as English /i/ (rather than /u/) and /u/ as /u/. Based on this assimilation patterns, discrimination was expected to be poor for /i/-/y/ contrast, better for /u/- /u/ contrast and good for /y/- /u/ and /y/- /u/. Danish listeners were expected to assimilate Norwegian /y/ and /u/ to Danish /y/, but not as equally good matches to Danish /y/. Therefore, Danish listeners were expected to discriminate /i/-/y/, /y/-/u/ an /u/- /u/ excellently and /y/- /u/ less well. French listeners were expected to assimilate /u/ to French /y/, and Norwegian /y/ as either poor French /y/ or as French /i/. They were expected to discriminate /y/- /u/, /y/- /u/ excellently and /i/- /y/ relatively well. While the discrimination patterns across speaker groups largely confirmed predictions based on the model, Danish speakers’ good discrimination of the /y/- /u/ contrast was above the predicted performance levels. Both /y/ and /u/ were mainly assimilated to Danish long /y/, less to short /u/, but were assigned better ratings when assimilated to /y/ (see also Polka (1995) for similar English speakers’ perception of German lax /u/- /v/ contrast differing in degree of fit to English /u/ or /u/ vowels).

Discrimination predictions for TC (good discrimination), UC (very good discrimination), CG (moderate to very good discrimination) and SC (poor discrimination) contrasts were mostly confirmed in Greek and Japanese speakers’ perception of Southern British English contrasts (/i:/-i:, /u/-/u:/, /æ/-/æ/, /æ/-/æ/, /æ/-/æ/ in b_p and b_b contexts (Lengeris & Hazan, 2007). Both Greek and Japanese have five-vowel inventories, but only Japanese has short and long versions of these five vowels. By comparing perception of the same vowels by two groups of listeners with (phonologically) the same distinctions (/i, e, a, o, u/) that differ in duration only in one group’s L1, the degree to which each group used duration was tested.

As expected, the groups’ perception of non-native contrasts differed according to their first language. Discrimination accuracy was highly predictable from the
assimilation types. Japanese listeners used the experience with duration in their native language in the discrimination task. Furthermore, Greek listeners also used duration to some extent (e.g. to discriminate (/i:/-/t/ and /u:/-/u:/ contrasts), although not so effectively as Japanese listeners. Japanese listeners clearly assigned English tense and lax vowels to Japanese long and short vowels.

The role of vowel duration is of a particular importance for this study and will be discussed separately in the following section.

5.1.2 Effects of vowel duration on L2 vowel perception

In the chapter on L2 production (Chapter 4), empirical evidence was reviewed showing that in production of vowels, non-native speakers often exploit temporal differences in their L1 to compensate for the difficulty they have in producing L2 vowels differing in spectral characteristics. L2 learners may also use this temporal difference when perceptually categorizing L2 vowels even if vowel duration is not a contrastive feature in their L1 (Bohn, 1995). For example, native French and native American English listeners categorized a synthetic /kVt/ continuum using the French words /kot/ and /kOt/ (cote, cotte) (Gottfried & Beddor, 1988) when spectral and temporal properties of the vowel portion were varied orthogonally. Duration does not reliably differentiate vowels in French, whereas American English listeners have a distinction between tense and lax vowels. The French were not influenced by the duration differences in the identification task, relying only on the spectral differences that cue the distinctions between /o/ and /ɔ/. Unlike French listeners, American listeners were influenced by the duration manipulations. In light of these findings, Gottfried and Beddor concluded that the differing status of vowel duration in French and English determined the categorization performance of these two listener groups in the study. Being familiar with duration differences that are featured to some extent in their own language, Americans used duration to distinguish the non-native vowel contrast in this study.

Furthermore, perception of duration differences of non-native vowel contrasts may depend on the voicing of the following consonant (Hazan et al., 2006; Morrison, 2002). Japanese listeners, for example, identify English /i/ as /i/ in front of a voiceless and as /i/ in front of a voiced consonant. Duration differences are easier to perceive
than spectral differences and will be preferred if the listener’s L1 has not exposed them to spectral contrasts in a particular part of the vowel space (Bohn, 1995).

5.2 From language universal to language specific in speech perception

An important question in research on adult cross-language segmental perception is how native language effects on non-native speech perception emerge developmentally. Research examining infants’ speech perception suggests that the shift from language-general to language-specific perception occurs during the first year of life. Although a lengthy review of early speech perception is beyond the scope of this thesis, a brief overview of relevant findings and theoretical implications of developmental changes is important for understanding AOA effects on adults’ perceptual attunement to the contrastive properties of their native language.

Initially, infants are able to perceive language-universal properties of speech. This ability declines for some consonants by 8-10 months (Best, McRoberts, LaFleur, & Silver-Isendstadt, 1995; Werker & Tees, 1983, 1984) and for vowels around 6 months of age (Polka & Werker, 1994). Studies of infants’ perception of non-native vowel contrasts show an overall age-related decline of perceptual ability in both monolingual and bilingual infants. For example, English-learning 4 to 6 months old infants (Polka & Werker, 1994) are able to discriminate German /r/ - /u/ (a lax high front-rounded vs. a lax high back-rounded vowel) and /y/ - /u/ (a tense high front-rounded vs. a tense high back-rounded vowel). By the age of 6 to 8 months, on the other hand, their discrimination of this contrast undergoes a decline. Older English-learning infants (10 to 12 months) show no ability to discriminate these two contrasts. A similar pattern was observed for bilingual Spanish-Catalan infants (Bosch & Sebastian-Galles, 2003) tested on perception of the /e/ - /e/ contrast. This contrast is present only in Catalan. In Spanish, on the other hand, there is a single mid-front vowel /e/. The central questions in this study focused on how infants in bilingual environments would manage this distinction. As expected, younger Spanish-Catalan (Spanish L1) infants (e.g. 4 months) were able to perceive this distinction, but by the age of eight months this ability diminished; however, Catalan-Spanish (Catalan L1) were able to discriminate the /e/ - /e/ contrast after the age of 4 months.
The shift from language-universal to language-specific phonetic patterns during the first year of life is considered to reflect the development from perception of nonlinguistic information to language-specific phonetic information before the acquisition of a lexicon. With the acquisition of lexicon, infants finally attune to the contrastive phonological classes necessary for distinguishing words. A possible reason for the change in sensitivity occurring earlier for vowels than consonants may be related to vowels’ prominence in speech: vowels tend to be longer and higher in intensity than consonants and they carry phonetic and prosodic information. These properties of vowels may be more likely to attract infants’ attention than consonants. However, it is important to note that this decline in infants’ perceptual ability does not occur for all non-native contrasts (Best et al. 1988) and may be regained late in childhood (Polka, Colantonio, & Sundara, 2001). These findings clearly suggest that simple perceptual tuning to the phonological properties of the native language does not provide an explanation for why infants’ decline in perceptual ability is not found for all non-native contrasts. A more likely explanation for this decline in perception of certain non-native contrasts is related to the specific way the non-native contrasts map onto the kinds of phonemic distinctions present in the native language (Best, 1995). This specific perceptual mapping in infancy may cause differences in the learnability of non-native contrasts when a second language is learned later in life. Shorter exposure to the native language (e.g., acquisition of a second language in early childhood) may only mean that the influence of these early established native and non-native contrasts are not strong enough to impede perception of the second language contrasts. However, they may be still robust enough to cause perceptual difficulties for nonnative contrasts, even if the second language is learned early in life (Bosh & Sebastian-Galles, 2003). This argument will be further discussed in interpretations of the results of this thesis research. The next chapter will outline the rationale for the studies conducted.
CHAPTER 6
THE PRESENT RESEARCH PROJECT

The present studies examined Serbian-English bilinguals’ (L2 learners) perception and production of two Australian English (AusE) vowel contrasts: /æ/ - /e/ and /i:/ - /i/. The bilinguals differed in their age of learning (AOL) of Australian English (AusE) as a second language (L2). The findings based on the Speech Learning Model (SLM) (Study 1) were extended to generate and test predictions based on the Perceptual Assimilation Model (PAM) (Study 2) in order to assess the effect of AOL on assimilation patterns of L2 vowel contrasts (Best & Tyler, 2007). In addition, the effect of AOL on sublexical perception of these contrasts was assessed in a priming experiment (Study 3).

The language choices for the studies reported in this thesis were motivated by the differences in the complexity of their vowel inventories and the role of vowel duration in Serbian. The AusE vowel inventory represents one of the richer systems in the world, containing 12 monophthongs and 7 diphthongs (Cox, 2006). The Serbian language, on the other hand, has five short vowels and the vowel length is exploited in such way that each short vowel is matched by a long vowel (Kordic, 1997; Simic & Ostojic, 1989). According to previous research, the larger vowel inventory may be an advantage in learning non-native vowels. For example, Spanish learners have difficulties learning to discern the difference between English /i:/ and /i/ (Escudero & Boersma, 2004) because they are both perceptually assimilated to Spanish vowel /i/, whereas German learners are better at discriminating the same contrast (Bohn & Flege, 1990; Escudero & Boersma, 2004) because they perceptually assimilate the two English vowels to two different German categories. However, the speakers of languages with smaller vowel inventories may rely on vowel duration when learning English /i:/ - /i/ contrast. For example, like Serbian, Japanese has a vowel duration contrast and Japanese learners of English rely on duration difference to discriminate between English /i:/ (tense) and /i/ (lax) vowels (Morrison, 2002).
Given the evidence that L2 learners use their existing L1 categories when perceiving L2 segments (Best, 1995; Flege, 1995), this study is concerned with how Serbian-AusE English bilinguals’ experience with their native vowel inventory influences perception of the two AusE vowel contrasts. In particular, it investigates how experience with Serbian vowels influenced perception and production of the tested contrasts at different ages of learning (AOL) of AusE.

First some background information about Serbian and AusE English will be discussed with emphasis on their vowel inventories. Second, the predictions based on the models will be presented.

6.1 The Serbian language

6.1.1 Cultural and linguistic background

The Serbian language belongs to the southern group of Slavonic languages. The history of the Serbian language (and other Slavonic languages) began in the sixth and seventh centuries during a large migration of various Slavonic tribes (one of which was to become the Serbian nation) from the north of Russia, Byelorussia and Ukraine to the Balkan peninsula and the region of Pannonia. Although the Slavonic languages remained similar, separation of the Slavonic tribes was reflected in the development of distinctive characteristics in the languages spoken by the southern, eastern and western Slavs (hence Slavonic languages are divided into three groups: southern, eastern and western).

The Serbian language is spoken by approximately 17 million people (Kordic, 1997) in four of the former Yugoslavian republics: Croatia, Serbia, Montenegro and Bosnia and Herzegovina. After the disintegration of Yugoslavia, each of the three countries that have gained international recognition as independent states, calls this language according to its ethnic identity: Croatian, Serbian and Bosnian. Although standard Serbian and Croatian are still the same language (with several distinct dialects within Serbian and especially within Croatian) (Browne, 1993), broken contacts between the countries have led to divergence in the language’s norms. Both Serbia and Croatia have independent language planning policies with special focus on the development of new and distinctive vocabularies.
Of three dialects (Stokavian, Kajkavian and Cakavian) spoken in the area of the present day Serbian and Croatian languages, the Stokavian was chosen as the basis of standard Serbo-Croatian because it was the most widespread dialect (Kordic, 1997). It is spoken in all of Serbia, Bosnia and Herzegovina and a large part of Croatia, east and south-east of the Cakavian and Kajkavian dialect (figure 6-1).

Figure 6-1. Serbo-Croatian dialects (from Kordic, 1997).

The Stokavian dialect has three subdialects that developed from three different pronunciations of the Old Slavic \textit{jat} (long vowel \(\acute{c}\)): \(\text{ikavski (Ikavian, } \acute{c} > i)\), \(\text{ijekavski (Ijekavian, } \acute{c} > ije)\), and \(\text{ekavski (Ekavian, } \acute{c} > e)\).

In the early 1800s, Serbian and Croatian intellectuals seeking the unity of all south Slavs accepted the Stokavian dialect with Ijekavian and Ekavian pronunciations as the standard language. That was the beginning of the Serbian and Croatian linguistic history during which two standard languages developed into a single linguistic variety, Serbo-Croatian. This linguistic unity was broken in the early 1990s after destructive wars that separated the Yugoslavian republics.

Within the Ekavian dialect, which is spoken in Serbia, there are several regional dialects that differ in some aspects of grammar (Browne, 1993) (number of cases and verb tenses), consonant inventory and the realization of vowel length in unaccented
syllables (e.g., northern dialects maintain the standard vowel length distinction in unaccented syllables, which is often lost in southern dialects).

The native dialect of the participants in these studies is the Sumadijsko-Vojvodjanski dialect spoken in northern Serbia. Thus, these participants maintain the vowel length distinctions even in unstressed syllables. This dialect is a standard Serbian language.

6.1.2 Serbian vowel inventory

The Serbian vowels /a, e, i, o, u/ (table 6-2) can be long and short and may occur in any position in a word. In addition, /r/ can act as a vowel (long and short). This vocalic (syllabic) /r/ is not specifically marked in writing, but it is always syllabic when between two consonants (e.g. prsten [ring])

Vowel duration in Serbian appears as a morpho-phonemic contrast rather than on the level of minimal pair open class stem morphemes (see section 6.1.3).

Table 6-1
Serbian vowels

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>i</td>
<td>u</td>
<td></td>
</tr>
<tr>
<td>Mid</td>
<td>e</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

6.1.3 Prosodic phenomena: Stress and vowel length

The vowels in Serbian words can be accented (stressed). The term stress accent refers to the involvement of both pitch and length changes between stressed and unstressed syllables (Hamond, 2005). The accent may have a rising or a falling pitch. Both short and long vowels may carry rising and falling pitch. Hence, there are four different accents in Serbian: short falling, long falling, short rising, and long rising. The names of the accents suggest a pitch change on a given syllable (Lehiste & Ivic, 1986). Pitch ascends within long rising accented vowels, and drops during long falling
accented vowels. Short accented vowels have no such obvious pitch rise or fall. The difference between the short accents is the relationship with the following syllable: the syllable after a short rising accent begins equal to or higher in pitch than the accented syllable itself, then declines. The syllable after a short falling accent begins distinctly lower.

Accents can alternate the position or contour within the paradigm of a word (e.g., nouns and adjectives in different cases). The most frequent shift occurs in Genitive sg and Genitive pl: the short accented vowel in Genitive sg (and Nominative sg) becomes long accented in Genitive pl. For example, short e in the noun ime ‘name’ (Nominative sg and Genitive sg) becomes long in imena (Genitive pl). Furthermore, the change of vowel length in the accented syllable is accompanied by lengthening of the vowel in the syllable following the accented syllable. This phenomenon is known as post-accentual length. Post-accentual lengths are always associated with specific suffixes or grammatical forms (such as Genitive plural in nouns). Accent and post-accentual length are not indicated in writing.

6.2 Australian English (AusE)

6.2.1 Cultural and linguistic background

Australian English (AusE) is regional dialect of English spoken in Australia which includes many varieties of English (Cox, 2006), including Standard AusE, varieties of Aboriginal English and ethnocultural AusE dialects. English was brought to Australia in 1788 by speakers from England, Scotland, Wales and Ireland. The origins of AusE pronunciation are seen as an amalgam of regional varieties belonging predominantly to Irish English (Mitchell & Delbridge, 1965) and British English (Bernard, 1981) influences. Another view is that mixing took place somewhere in the southeast Midlands (Collins, 1975) and the complex mixture that came to Australia was then “piginized” through a reduction of phonological rules and variants. The most recent view is that different pronunciations of the mutually intelligible English varieties of the first English-speaking settlers formed the basis for modern standard AusE (Trudgill, 2004). Migration from these countries continued in the following two centuries. In the 20th century, these migrants were joined by many others from southern Europe, the Middle East and Asia. Those new language contacts defined a new stage in
the development of AusE, with possible disintegration of social class distinctions reflected in the pronunciations (Horvath, 1985).

The range of different pronunciations of the standard AusE include Broad, General and Cultivated Australian English (Mitchell, 1946). According to Mitchell, cultivated pronunciation is related to the highest prestige and is spoken by about 10% of Australians. Broad is spoken by a third of Australians and has the least prestige. General AusE pronunciation falls in between Broad and Cultivated pronunciations and is spoken by majority of the people. These three varieties of standard AusE are more similar to each other than to other English dialects (Cox, 2006; Cox & Palethorpe, 2007), sharing the same phonemic system and varying in the realizations of some vowels. However, each variety consists of a mix of Broad, General and Cultivated vowel pronunciations (Horvath, 1985)

The consonant system of AusE doesn’t differ to any great extent from other dialects of English.

6.2.2 AusE vowel inventory

The AusE vowel inventory is represented in the Table 6-2 using the revised phonemic transcription system proposed by Harrington, Cox and Evans (1997) and traditional symbols used by Mitchell (1946).
Table 6.2. AusE vowels: traditional (Mitchell, 1946) and revised phonemic transcription (Cox & Evans, 1997) (from Cox, 2006).

<table>
<thead>
<tr>
<th>Revised (Cox &amp; Evans, 1997)</th>
<th>Traditional (Mitchell, 1946)</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>/t/</td>
<td>/i/</td>
<td>heed</td>
</tr>
<tr>
<td>/t/</td>
<td>/i/</td>
<td>hid</td>
</tr>
<tr>
<td>/eɪ/</td>
<td>/ɛi/</td>
<td>hair</td>
</tr>
<tr>
<td>/e/</td>
<td>/ɛ/</td>
<td>head</td>
</tr>
<tr>
<td>/æ/</td>
<td>/æ/</td>
<td>had</td>
</tr>
<tr>
<td>/ʊ/</td>
<td>/ʊ/</td>
<td>hard</td>
</tr>
<tr>
<td>/əʊ/</td>
<td>/əʊ/</td>
<td>mood</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>/ɔ/</td>
<td>pod</td>
</tr>
<tr>
<td>/ɔɪ/</td>
<td>/ɔɪ/</td>
<td>heard, saw</td>
</tr>
<tr>
<td>/u/</td>
<td>/u/</td>
<td>hood</td>
</tr>
<tr>
<td>/u/</td>
<td>/u/</td>
<td>who'd</td>
</tr>
<tr>
<td>/i/</td>
<td>/i/</td>
<td>heard</td>
</tr>
<tr>
<td>/i/</td>
<td>/i/</td>
<td>the</td>
</tr>
<tr>
<td>/ɛɪ/</td>
<td>/ɛi/</td>
<td>hay</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>/ɛ/</td>
<td>hoe</td>
</tr>
<tr>
<td>/æʊ/</td>
<td>/æʊ/</td>
<td>high</td>
</tr>
<tr>
<td>/æ/</td>
<td>/æ/</td>
<td>bow</td>
</tr>
<tr>
<td>/ɔɪ/</td>
<td>/ɔɪ/</td>
<td>toy</td>
</tr>
<tr>
<td>/ʊ/</td>
<td>/ʊ/</td>
<td>here</td>
</tr>
</tbody>
</table>

The symbols in the left hand column of Table 6-2 (Harrington, Cox & Evans, 1997) represent the revised set of symbols for Standard Australian English, which were developed to more accurately indicate phonetic properties of the vowels than the earlier system by Mitchell (1946) given in the middle column. It is generally agreed that the traditional Mitchell transcription system for AusE vowels more strongly reflects a British rather than Australian standard. Furthermore, in Harrington, Cox and Evans’ (1997) transcription system, length is explicitly indicated as it has phonemic status and is the differentiating quality between, for example /e/ and /eː/. Length is also a major cue for the /i/- /iː/ contrast.

A recent report on acoustic analyses of AusE vowels (Cox, 1996, 2006; Harrington, Cox & Evans, 1997) revealed that there was not much variation across the Broad, General and Cultivated varieties in the targets for the monophthongs. There were no significant differences between General and Cultivated pronunciation. However, there were some differences for Broad pronunciation. For example, vowel /iː/ has a longer onglide in Broad pronunciation than in General or Cultivated, and there is more
tongue fronting for /æ:/, /ʊ:/ and /i/ in Broad than in General and Cultivated pronunciation. The possible implication of the use of two different transcription systems for comparison between Serbian and AusE vowels used as stimuli in the experiments reported in this thesis will be discussed in Section 6.3.

6.2.3 The two vowel systems

With respect to L2 vowel perception and production, the comparison between the two vowel systems predicts the difficulties L2 learners may have in L2 acquisition. This comparison regards the difference in size of vowel systems, the presence and absence of vowels in different inventories, and the different feature and spectral characteristics that signify vowel contrasts.

Regarding the size of the vowel inventory, Serbian, with 5 vowels (and each short vowel matched by a long vowel) belongs to those languages with rather small vowel systems. In such systems it is believed that vowels are evenly distributed in the vowel “space”, with one vowel occupying each position (e.g. front, central, back; high, mid, low) (see Table 6-1). However, there are suggestions (Lindlom, 1986) that vowels are not necessarily evenly dispersed over the available space in small vowel systems, but actually follow the principle of ‘sufficient contrast’ and therefore tend to occupy a more compact space than those of languages with large vowel contrasts, such as AusE. Regarding the distribution of vowels in Serbian and the AusE target vowels, the two vowel systems compare in the following way:

Both AusE /i:/ and /u/ are high front vowels, /u/ is shorter (lax) and more retracted than tense /i:/ and /u/. There is only one Serbian vowel (/i/) which occupies high frontal position. Regarding the AusE /æ/-/e/ (marked as /ea/-/e/ contrast in the figure) contrast, /æ/ occupies a lower position than /e/. There are two Serbian vowels /a/ and /e/, which are relatively close to AusE /æ/ and /e/ in terms of their spectral characteristics (Figure 6-2).
Figure 6-2. Distribution of Serbian /i/, /e/, /a/ and AusE /i/, /iː/, /e/, and /æ/ vowels in the F1-F2 vowel space according to the values for male speakers reported in Cox (1996) and Simic and Ostojic (1987).

The acoustic comparison points to the possibility that Serbian speakers exploit L2 spectral differences by analogy to existing differences in their L1. It was expected that both English /iː/ and /i/ would be perceived relative to the Serbian /i/, whereas AusE /æ/ and /e/ will be perceived relative to the Serbian /a/ and /e/.

Unlike AusE, the Serbian vowel system recognizes the length difference, but only on a morphological level. Since previous research indicates a certain dependency on duration in L2 vowel acquisition, these two AusE vowel contrasts were chosen because of the presence of a length distinction in lax-tense /ʌ/-/iː/ contrast and the absence of this distinction in /æ/-/e/ contrast. Serbian-English bilinguals’ reliance on temporal distinction in the categorization of AusE vowels was addressed in Study 2 (Chapter 8).

6.3 Early and late Serbian-English bilinguals’ perception and production of two AusE /æ/-/e/ and /iː/-/i/ contrasts

The choice of vowels and the hypotheses tested were based on SLM’s predictions that some non-native phonetic contrasts should be relatively easy to produce and perceive (“new” categories ) and some should be difficult (“similar” categories). Regarding AOL, according to SLM, adult learners will be less successful in learning similar categories than young learners because the similar segments may become functionally equivalent (may occur as allophonic variants) of a category in their L1. In
order to successfully establish a new category for such an L2 segment, adult learners must learn to perceptually differentiate it from the L1 segment.

These experiments are concerned with the language-specific L1 (Serbian) nature of L2 (AusE) vowel perception and its implications for L2 learning, with particular interest in the relationship between AOL and the perception of difficult AusE vowel contrasts.

A valid and reliable method has yet to be developed for determining when L2 learners treat articulatory and acoustic differences between L1 and L2 vowels (and consonants) as phonetically relevant. Flege (1997) suggested three criteria for establishing whether a L2 segment is a new category for L2 speakers: 1) the use of IPA transcriptions of L1 and L2 segments, 2) acoustic measurements of the L1 and L2 segments and 3) listeners’ perceptual judgments of the similarity/dissimilarity between L1 and L2 segments.

Using these criteria, an L2 phonetic category is considered to be identical to an L1 category if: 1) both segments from the L1-L2 pair are represented with a single IPA symbol, 2) acoustic analysis of multiple productions of the L1 and the L2 categories by multiple native speakers of both languages does not reveal significant differences, and 3) native listeners of both languages cannot detect a difference between the L1 and L2 segments in a perceptual discrimination task.

To be classified as similar or new, the segments from the L1 - L2 pairs must be different acoustically and discriminated auditorily. The reliability of the use of IPA symbols in classification of similar segments is rather problematic, as in most of the cases the segments from different languages are represented with the same IPA symbol even if the statistical analysis of acoustical measurements show significant acoustic differences and there are audible differences between them. An additional problem with the IPA symbol criterion is the use of multiple symbols for the same segment. For example, the vowels in English *beat* and *bit* are sometimes represented as */i/* and */u/* respectively (old Mitchell’s transcription system), and sometimes as */i:/ and */u/* as in Harrington, Cox & Evans’ (1997) transcription system (if the durational difference between the tense-lax vowel pair *beat* and *bit* is taken into account). Clearly, these two transcription systems use the same symbol for two different vowels: according to Harrington, Cox & Evans, */u/* is a high, front, lax, unrounded vowel close to Serbian */i/,
whereas according to Mitchell, /i/ is a high, front, tense, unrounded vowel dissimilar to Serbian /i/. In both cases, based on the symbol used only, it would be expected that Serbian learners have no difficulties in perception (and production) of these two clearly distinctive AusE vowels, as they share the same symbol as Serbian /i/. Similarly, vowels in the /æ/ - /e/ contrast according to Mitchell’s transcription are new for Serbian speakers as they don’t share the same phonetic symbol with any of Serbian vowels (as Mishel’s transcription for AusE /e/ is /e/). According to the Harrington, Cox & Evans, only /æ/ is new for Serbian speakers, whereas /e/ is not. Therefore, the use of these two transcription systems might affect the choice of stimuli used in this study in following way:

/i:/ - /i/ contrast:

1) if Mitchell’s transcription is applied, /i/ (a tense vowel) is the same vowel as Serbian /i/ and native Serbian speakers will not have any difficulties perceiving and producing it;
2) if Harrington, Cox & Evans’ transcription is applied, both /i/ (high, front lax vowel) and /i:/ (high, front, tense vowel) are new for Serbian speakers as these phonetic symbols are not used for any of the vowels in the Serbian language.

/e/- /æ/ contrast:

1) if Mitchell’s transcription is used, both vowels are new for Serbian speakers;
2) if Harrington, Cox & Evans’ transcription system is used, /æ/ is new for Serbian speakers, but /e/ is not.

Furthermore, the symbolization of duration difference is often ignored when representing vowels in languages that do distinguish between long and short vowels. Usually, it is considered that the Serbian vowel system contains five short vowels: /a/, /e/, /i/, /o/, /u/. Australian English approximations often given to Serbian learners are the first vowel in father (Serbian /a/), the vowel in ten (Serbian /e/), sit (Serbian /i/), stop (Serbian /o/) and boom (Serbian /u/). This description excludes the five long vowels (/a:, /e:, /i:, /o:, /u:/) in Serbian, where vowel duration is morphologically and lexically conditioned (that is, short vowels become long in, for example, plural nouns and verbs).

In addition to these three criteria, Flege and Bohn (1992) proposed that realizations of a new L2 vowel should occupy a portion of the acoustic phonetic space
(i.e., according to their F1 and F2 values) that is not occupied by the realizations of the closest L1 vowel(s). According to this criterion, for speakers of the languages with rich vowel inventories few L2 vowels would be new as most of the vowel acoustic space is occupied by the realizations of their L1 vowels. By contrast, for Serbian speakers most of the AusE vowels are new, as the Serbian vowel system contains only five short and five long vowels differing in duration.

The AusE vowel contrasts investigated in these experiments were chosen according to their relationship to the articulatory closest Serbian vowels in the vowel inventory. For the /i:/ - /a/ contrast, the only Serbian vowel that occupies the front high position is vowel /i/. It was hypothesised, based on the SLM (Flege, 1995), that the members of the AusE /i:/ - /a/ contrast would be perceived as similar to Serbian /i/ and therefore difficult for (late) Serbian-English bilinguals to produce and perceive. Members of the /e/ - /æ/ contrast, on the other hand, were positioned in close proximity to two Serbian vowels: /e/ and /a/ and were expected to be new for Serbian speakers and hence relatively easy to acquire even for late learners.

Given the evidence that L1 background determines the way L2 vowels are perceived and pronounced, and that this influence is not the same when the L2 is learned early in life versus later in adulthood, these experiments sought to provide more evidence of the perceptual influence of the L1 on the formation of L2 vowel contrasts regarding the AOL. Since the amount of experience bilinguals have with their L1 also changes over time, one important implication for theories of L2 segmental perception and production has to be the listener’s level of L1 development. These studies seek evidence not only that bilinguals rely on their L1 segments when perceiving the L2 segments, but also that the ways bilinguals exploit their experience with L1 segments changes with increasing experience with their native language.

The results of three studies will be presented in Chapter 7 (Study 1), Chapter 8 (Study 2), and Chapter 9 (Study 3). In Study 1 predictions based on the SLM were tested. In Study 2 PAM’s predictions about the relationship between perceptual assimilation of L2 vowels to L1 vowels and discrimination of the same two vowel contrasts were tested. Therefore, Study 2 extended PAM’s predictions to bilingual L2 vowel production and perception and the question of how perceptual assimilation patterns of L2 contrasts differed according to the age of learning of L2. In Study 3,
Serbian-English bilinguals’ phonological and phonetic processing of two AusE vowel contrasts was investigated using task-appropriate adaptations of the phonological priming method.
CHAPTER 7

STUDY 1. PERCEPTION AND PRODUCTION OF AUSTRALIAN ENGLISH /æ/ - /e/ AND /iː/- /i/ BY SERBIAN-ENGLISH BILINGUALS:

IMPLICATIONS FOR THE SPEECH LEARNING MODEL

This study examined the perception and production of the Australian English /e/ - /æ/ and /i/- /i/ vowel contrasts by two groups of Serbian speakers who differed in terms of the age at which they learned AusE. Vowel production accuracy was tested in one experiment in which native monolingual speakers of AusE identified the target English vowels produced by Serbian-English bilinguals. Vowel perception was tested in two experiments: (a) Discrimination and (b) Perceptual assimilation with goodness-of-fit ratings. The questions addressed in the study were: (a) whether production inaccuracy of /e/-/æ/ and /i/-/i/ contrasts reflected a tendency on the part of the Serbian-English bilinguals to perceptually relate the targeted AusE vowels to their native vowels, (b) whether one of these two contrasts would be more “difficult” to produce and perceive than the other, and (c) whether early and late Serbian-English bilinguals would differ in production and perception of these two vowel contrasts.

Initial focus in the stimuli choice for this study was on purely phonemic contrast (vowel contrasts that appear in minimal pairs of mono-morphemic words) and therefore vowel length was not taken into account. This approach was consistent with usual description of Serbian vowel system in literature as a 5-short vowel system in which vowels differ in the colour alone. Furthermore, vowel length, which becomes relevant on morphological level, is not orthographically represented. Given that Serbian orthography is highly transparent, this lack of representations for vowel length suggests that the vowel length is not purely phonemic.

The predictions were based on the Speech Learning Model (SLM). According to this model, the likelihood of category formation for L2 phones is affected by both age of L2 learning (AOL) and the perceived phonetic distance of an L2 phone from the closest L1 phone. Furthermore, late bilinguals often produce vowels in a second language differently than do monolinguals and early speakers (Flege, 1992a; Flege,
Bohn et al., 1997; Munro et al., 1996) and that accuracy in production of L2 vowels is related to the degree of perceived cross-language phonetic similarity (Flege, 1995).

**7.1 Predictions based on the SLM**

The following hypotheses concerning production and perception of the vowels in this study were based on the SLM: (a) the vowels in the AusE /i:/ - /i/ contrast would both be perceived as similar to the Serbian /i/, (b) late learners’ failure to perceptually distinguish the AusE /i:/ - /i/ from the Serbian /i/ would lead to inaccurate production, (c) late learners would fail to perceptually separate these AusE vowels from the Serbian /i/, (4) both vowels in the AusE /e/ - /æ/ contrast would be perceived as distant from any of the Serbian vowels and would not cause production or perception difficulties for late bilinguals, and, finally, (d) early learners would not have difficulties with either of the AusE contrasts.

However, given the length distinction between Serbian /i/ and /i:/, hypotheses based on SLM might be problematic. That is, it might be expected that Serbian speakers would treat AusE /i/ and /i:/ as their long and short Serbian equivalents. If that is the case, the critical element in perception of AusE /i/ and /i:/ vowels would be the length distinction (rather than the quality distinction as in the case of AusE /a/ - /æa/ contrast). If AusE /i/ and /i:/ are perceived as two different Serbian categories, PAM, rather than SLM framework, provides reliable criteria for establishing the degree of dissimilarity between L2 and L1 phonetic segments.

The SLM is based on three key lines of evidence. The first concerns that fact that younger L2 learners demonstrate a greater ability to perceptually differentiate between an L2 vowel and the closest L1 vowel, relative to older L2 learners (Flege, Mackay et al., 1999; Piske et al., 2002). Secondly, accuracy in production of L2 vowels is limited by how accurately the vowels are perceived (Flege, 1999b). Finally, with increasing AOL, discrimination accuracy of L2 vowel contrasts decreases (Flege, Mackay et al., 1999). That is, with increasing AOL, learners may increasingly fail to detect the phonetic differences between L2 vowels (and consonants) that are not contrastive in their L1. This failure in perception may result in inaccurate production of these segments. On the other hand, individuals learning an L2 early in life will be more
likely to have established a category for a given L2 vowel that enables them to perceptually differentiate between an L2 vowel and the closest L1 vowel.

Furthermore, the formation of new phonetic categories (or functional equivalence classes) for L2 vowels becomes less likely with increasing AOL. Vowels of a given degree of phonetic similarity to the closest L1 vowel will, with increasing age, become more likely to be heard as instances of an L1 category. Vowels that are not perceived as similar to any L1 vowels will be easier to perceive and, as a consequence, even late L2 learners may establish a category for them. L2 vowels that are perceptually assimilated to specific L1 vowels (at least by late L2 learners) will be acoustically close to those L1 vowels.

7.2 General method

7.2.1 Participants

Thirty Serbian-English bilinguals were recruited and allocated to one of two groups according to their age of learning (AOL) of Australian English. The AOL of AusE requirement for the participants in the “early” group was before the age of five years. The AOL of AusE for the “late learners” was after the age of fifteen years. Participants in both groups were Serbian migrants who spoke Serbian as their first language (L1) and AusE as the second language (L2). All participants had to be fluent in Serbian and English. Thus, the late learners were required to have been using English at least 5 years before the time of testing, so that L2 English should have been relatively well-established. The Language Background Questionnaire (LBQ) and the testing paperwork (Information statement and Consent form) were in Serbian (see Appendices on the Supplementary CD for testing paperwork) in order to assure that the participants were fluent in Serbian. Participants’ consent was obtained prior to testing. In addition, the participants received all oral and written instructions in Serbian, in order to assure that the Serbian language mode throughout the tasks was maintained (Antoniou, Best, Tyler, & Kroos, 2007). In previous research no particular attention was paid to maintaining participants’ native language mode.

In addition to Serbian-English bilinguals, a group of 15 native AusE monolingual speakers participated as a control group in the discrimination and production experiments (Experiments 1 and 3). AusE monolinguals did not participate
in Experiment 2, which was designed to test how Serbian-English bilinguals perceptually assimilated the AusE vowel contrasts to native Serbian categories. A further 10 native AusE monolinguals participated as listeners judging the production accuracy of the Serbian bilinguals in the L2 vowel production experiment (Experiment 3). The native AusE monolinguals were undergraduate and graduate students in the School of Psychology and the School of Humanities and Languages at the University of Western Sydney.

7.2.2 Serbian speakers’ experience with English prior to migration to Australia

In the case of the “late learners” the age of learning (AOL) actually refers to the length of residence (LOR): all of them started acquisition of AusE upon migrating to Australia. In the case of the “early learners”, the AOL does not necessarily imply the LOR, as most of the participants in this group were exposed only to Serbian for some (variable) period of time after arrival in Australia, before beginning to learn AusE (in primary schools or child care centres). Therefore, in the case of late learners in this study, AOL will refer to the LOR. Although none of the participants in the late group reported being able to speak or understand English prior to migration to Australia, all of them reported some exposure to English. Some of them reported limited exposure to English in the classroom settings (with teachers who were not native speakers of English but rather Serbians teaching English). None of the participants who reported receiving classroom instruction in English were able to speak and understand English before coming to Australia. It is important to note that an absolute lack of exposure to spoken English is no longer possible in the modern world, due to access to English language films, music and media. For Serbian migrants in particular, exposure to American English is most likely, due to the high popularity of American films and TV shows that are always subtitled, never dubbed. Some degree of exposure to British English is also possible, again due to availability of spoken British media. Exposure to AusE prior to migration (i.e., in Serbia), on the other hand, is highly unlikely.

7.2.3 The Serbian speakers’ native dialect

The native dialect of the Serbian language was also taken into account when the participants were screened for participation in the experiment. The purpose of this increased amount of control compared to previous studies was to minimize Serbian dialect differences that might introduce variability in the speakers’ AusE perception and
production. All native speakers of Serbian were speakers of the Ekavian subdialect of the Sumadijsko-Vojvojanski region of Serbia, which belongs to the Stokavian dialect group. However, it is important to note that due to political factors from 1990-1996, numerous refugees from other Serbian dialect regions in the war zones migrated to Vojvodina and north Serbia, where the participants in this study were from. That is, it was impossible to find a Serbian speaker of any Serbian dialect who had not been frequently and heavily exposed to other Serbian dialects as well. Nonetheless, most dialects do share the same characteristics of the vowel system as the native dialect of the speakers in this study.

7.2.4 Age of learning, length of exposure and chronological age of participants

Thirty Serbian-English bilinguals participated in this study: 15 early learners and 15 late learners. For early bilinguals, mean age, AOL and length of AusE use were, respectively, 26.9 years (SD=3.56), 4.14 years (SD=1.02) and 22.64 years (SD=3.81). For late Serbian bilinguals, mean age, AOL and AusE use were, respectively, 30.7 years, (S=5.4), 21.57 years (SD=4.35) and 7.6 years (SD=4.37).

The mean age of the 15 AusE monolingual participants in the control group in Experiment 1 and 3 was 21.13 years (SD=1.55). The mean age of the 10 AusE monolingual listeners who judged the production accuracy in the production task (Experiment 3) was 21.4 years (SD=2.3).

7.2.5 Stimuli

The target stimuli for the experiments were the AusE vowels /æ/, /e/, /i:/, and /u/, produced by several native AusE speakers within an h_d frame (had, head, heed, hid) and embedded in a sentence frame. For example, “Hid is the next word to say. I say hid again. Hid.” The final tokens in citation form were used as stimuli. The same sentence frame was used for all stimuli.

The speakers were three male undergraduate students of Psychology at the University of Western Sydney. Prior to the recoding session, the speakers were screened for their language background via the Language Background Questionnaire. Only speakers who reported speaking no other language but AusE, and having no exposure to other languages on a regular daily basis, were recorded. Note that in a multilingual society such as Australia, it is not possible to find a speaker with absolutely no exposure
to other language(s). Nevertheless, an attempt was made to limit the extent of this exposure in those recruited for recording. Therefore, the speakers were required to be speakers of AusE and to have no family members who spoke any other language than Australian English or any other variety of English but AusE. Furthermore, they were excluded if they had experienced long exposure to other English varieties as well. Examples of such exclusions were an applicant who had been on a 6 month long working holiday in the United Kingdom, another who reported having a flatmate from the United States, and a third who worked in a bilingual environment.

Recorded speakers repeated the sentences with the target stimuli 10 times for each vowel in random order across the repetitions. Three tokens in citation form (that is, from the repeated word after the end of the carrier sentence) for each vowel were chosen for each speaker and used as stimuli for perception tasks for the Serbian participants. The use of multiple speakers forced the participants to focus on higher-order speech information rather than lower-level acoustic matches within a single speaker. The vowels in the chosen tokens were matched according to their duration, F0, F1, F2 and F3 at the vowel midpoint. Furthermore, the chosen tokens were also matched between vowel categories on duration and F0 (see supplementary CD for the Appendices). Thus, 36 stimuli were generated (4 vowels x 3 speakers x 3 tokens). The amplitude for each word was normalized to 90% of the peak intensity of all words and stored in separate files per word.

In addition, tokens of the AusE vowels /o/ and /u/ were recorded for the practice trials preceding the experimental tasks using the same speakers as before. These vowels were chosen because they were expected to be easy to discriminate for Serbian English bilinguals. The tokens were recorded in the same h_d and sentence context as the target stimuli (see Appendix 6.10).

Additionally, 3 tokens of each vowel per speaker were used to elicit production data (Experiment 1). Again, these were presented in an h_d frame embedded in the carrier sentence, “H_d is the next word to say”.

7.2.6 Recording equipment

Recording took place in a sound attenuated recording room in the MARCS Auditory Laboratories at the University of Western Sydney. The equipment used were an ACCER Travel mate 660 notebook computer, Edirol USB Audio Capture 24 bit 96
kHz UA-25 external sound card and Koss H-250 cardioid microphone. The speech recording and the stimulus editing were done using Cool Edit Pro 2000, with the sampling rate set to 44 kHz.

7.2.7 Experiments

Three experiments were conducted in the following order: (a) a Perceptual discrimination experiment (oddity discrimination task), (b) a Perceptual assimilation experiment with goodness-of-fit ratings, and (c) a Production experiment. Since the production accuracy of AusE /e/ - /æ/ and /i:/ - /ɪ/ contrasts was the main reason for further investigation of the perceptual assimilation of these contrasts to the native Serbian vowels, the production results are considered Experiment 1 and are presented first, followed by the perceptual assimilation results (Experiment 2) and, lastly, the discrimination results (Experiment 3). Only participants whose age of learning, years of use (at least five years), order of language acquisition (Serbian first, followed by AusE), and native language background (the same native dialect) met the requirements were tested. Testing was carried out at various sites, including a testing room in the MARCS Auditory Laboratories at the University of Western Sydney, a testing room at the School of Psychology at the University of New South Wales, or in a room reserved for individual use in the library of University of Sydney. Three participants were tested in their homes. These multiple locations were used in order to recruit as many Serbian participants as needed, by making the test sessions convenient to them. All care was taken to assure that room quietness was comparable across test locations. All room were quiet environments, and the background noise level was checked prior to each testing never exceeded 20 dB. The library room was booked before the experiments took place. All experiments were conducted in a single testing session, with 20 min breaks between the experiments.

7.3 Experiment 1: Production of /æ/ - /e/ and /i:/ - /ɪ/ contrasts

7.3.1 Design

The independent variables were age of learning (AOL) with two levels of measurement (early learners-EL and late learners-LL) and English vowel contrast (EVC) with two levels: /æ/-/e/ (EVC1) and /i:/-/ɪ/ (EVC2). AOL was a between-
subject factor and EVC was a within-subject factor. Dependent variables were the percentage of Serbian-English bilinguals’ accurate production scores and the identified vowel category (IDC) (IDC1: /æ/, /e/ and IDC2: /i:, /u/ by the native monolingual AusE listeners.

7.3.2 Procedure

The participants’ consent was obtained prior to the testing (see Appendices 6.1 and 6.2 for the consent form). The experiment consisted of two parts: speakers’ production and AusE monolingual listeners’ identification of the produced tokens. In the production part early and late Serbian-English bilinguals’ and native AusE monolinguals’ (control group) production of the AusE vowels /æ/, /e/, /i:/, and /u/ in h_d frame was recorded. All three groups of speakers previously participated in the discrimination experiment. Serbian-English bilinguals had also completed the perceptual assimilation experiment with goodness-of-fit ratings. The production accuracy of the target AusE vowels was assessed in the second part by determining the percentage of times a vowel was identified as intended by a panel of 10 monolingual native speakers of AusE listeners.

The technique used to elicit the production of AusE words with the target vowels was similar to that employed by Flege and colleagues (1995). The speakers were instructed to repeat the target word (that is, the target vowel in the h_d frame) after hearing it at the beginning of a carrier sentence “H_d is the next word to say.” The target vowels for productions were randomly presented. Each participant repeated each vowel 5 times. Three repetitions of each vowel were presented to the group of 10 AusE monolinguals for identification. This way 540 tokens were generated (45 speakers X 4 vowels X 3 repetitions).

The stimuli for imitative production were presented to the Serbian-English bilinguals and AusEs in the control group using the same equipment and recording setup as described previously. The software used for stimuli presentation and recording was DMDX. Recorded pronunciation of these words was digitised at 44.1 kHz using Cool Edit Pro 2000.

In the second part of the experiment 10 monolingual AusE listeners identified vowels produced by the 30 bilinguals and the 15 AusE monolinguals in the first part of
the experiment. The same equipment used for stimulus elicitation was used for identification by native AusE listeners.

The stimuli were randomly presented to participants one at a time in three counterbalanced sets of 180 trials in each set. This way one token of each vowel for each speaker was presented in each set. The listeners were instructed to identify each vowel using one of the following keywords presented on the computer screen: heed, hid, head, had, hard, hut, hot, horde, hood, heard. The interval between responses and the presentation of the next stimulus was 1000 ms.

7.3.3 Results

Percent of each target vowel correctly identified by native AusE listeners was averaged for each participant and then each group’s mean score was derived from these values. These results are presented in Figure 6-1 for the /e/- /æ/ contrast and in Figure 6-2 for the /i:/ - /ɪ/ contrast. Two separate ANOVAs were conducted, one for each contrast. The ANOVA results for the /e/- /æ/ contrast are presented in the Table 7-1, and the ANOVA results for the /i:/ - /ɪ/ contrast are presented in the Table 7-2.

Figure 7-1. Production of English had and head by early learners (EL) and late learners (LL) Serbian-English bilinguals and English monolinguals (M). On the x-axis AusE target words (bottom line) and the judgments on the Serbian-English bilinguals’
production (top line) are presented. The percentage of identified productions is presented on the y-axis.

An ANOVA revealed a significant main effect of AOL, significant main effect of IDC1, significant main effect of EVC1, significant interactions between AOL and EVC1, AOL and EVC1, and AOL, and IDC1 and EVC1. There was also a significant three-way interaction among AOL, EVC1 and IDC1 (see Table 7-1).

Table 7-1
Analysis of Variance for production of the /e/-/æ/ contrast

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>2</td>
<td>27.55**</td>
<td>0.00</td>
</tr>
<tr>
<td>IDC1</td>
<td>1</td>
<td>7.58*</td>
<td>0.02</td>
</tr>
<tr>
<td>EVC1</td>
<td>1</td>
<td>1358.23**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x IDC1</td>
<td>2</td>
<td>4.20*</td>
<td>0.03</td>
</tr>
<tr>
<td>AOL x EVC1</td>
<td>2</td>
<td>1987.48**</td>
<td>0.00</td>
</tr>
<tr>
<td>IDC1 x EVC1</td>
<td>2</td>
<td>9460.10**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x IDC1 x EVC1</td>
<td>2</td>
<td>897.61**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x IDC1 x EVC1 x S within group error</td>
<td>18</td>
<td>(25.28)</td>
<td></td>
</tr>
</tbody>
</table>

Value enclosed in parenthesis represent mean square errors. S = subjects.
* p < 0.05. **p <0.01

Group comparisons (EL vs. LL, EL vs. M and LL vs. M) in simple effect test of the AOL x IDC1 x EVC1 interaction revealed that the interaction was significant for the comparison of early versus late learners (AOL [EL vs. LL] x IDC1 x EVC1, F (1, 9) = 832.12, p < 0.01, late learners versus monolinguals (AOL [LL vs. M] x IDC1 x EVC1, F (1, 9) = 78.6 p < 0.01 and early learners versus monolinguals (AOL [EL vs. M] x IDC1 x EVC1, F (1, 9) = 8.94, p < 0.05. Early learners produced both vowels from the tested contrasts as intended. Late learners, on the other hand, pronounced both head and had as head. The difference between early learners and AusE monolinguals is due to the lack of variance in production of these two vowels by monolingual speakers: the
absence of within group variation for AusE monolinguals will render any difference among the groups’ means statistically significant.

Results on production of the /i: /-/t/ contrast is presented in Figure 7-2.

Figure 7-2. Production of English heed, and hid by early learner (EL) and late learner (LL) Serbian-English bilinguals and English monolinguals (M). On the x-axis AusE target words (bottom line) and the judgments on the Serbian-English bilinguals’ production (top line) are presented. The percentage of identified productions is presented on the y-axis.

An ANOVA revealed a significant main effect of AOL, a significant interaction between IDC2 and EVC2 and a significant interaction among AOL, IDC2 and EVC2 (see Table 7-3). Group comparisons (EL vs. LL, EL vs. M and LL vs. M) in simple effect test of the AOL x IDC2 xEVC2 interaction revealed significant effect of AOL when early and late learners were compared (AOL [EL vs. LL] x IDC2 x EVC2, $F(1,9) = 33.77$, $p < 0.01$ and when late learners and monolinguals were compared (AOL [LL vs. M] x IDC2 x EVC2, $F(1,9) = 36$, $p< 0.01$. There was no significant difference between early learners and monolinguals. The significant differences obtained for early learners versus late learners and late learners versus monolinguals are the consequences of lack of the variance for early learners and monolinguals owing to their almost perfectly identified productions.
Table 7-2
Analysis of Variance for for production of the /i:/ - /I/ contrast

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>2</td>
<td>33.86**</td>
<td>0.00</td>
</tr>
<tr>
<td>IDC2</td>
<td>1</td>
<td>1.76</td>
<td>0.22</td>
</tr>
<tr>
<td>EVC2</td>
<td>1</td>
<td>7.01</td>
<td>0.03</td>
</tr>
<tr>
<td>AOL x IDC2</td>
<td>2</td>
<td>0.12</td>
<td>0.88</td>
</tr>
<tr>
<td>AOL x EVC2</td>
<td>2</td>
<td>1.64</td>
<td>0.22</td>
</tr>
<tr>
<td>IDC2 x EVC2</td>
<td>2</td>
<td>65145.09**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x IDC2 x EVC2</td>
<td>2</td>
<td>19.98**</td>
<td>0.00</td>
</tr>
<tr>
<td>(AOL x PVC2 x EVC2) x S within group error</td>
<td>18</td>
<td>(6.31)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects.
* p < 0.05. ** p < 0.01.

7.3.4 Results: Experiment 1

According to the SML predictions (Flege, 1992a; Flege, Bohn et al., 1997; Munro et. al., 1996), age of L2 acquisition has an effect on production of L2 vowels. Late Serbian-English bilinguals failed to produce the /æ/-/e/ contrast accurately, producing head for both had and head targets. The late learners’ accurate production of the /i:/ - /I/ contrast also confirmed the hypothesis that one of the contrasts would be more difficult than the other for the late learners. However, based on the acoustical dissimilarities of AusE /æ/ and /e/ to Serbian /a/ and /e/, it was hypothesized that perceived similarity between the English /i:/ - /I/ contrast and Serbian /i/ would prevent late Serbian-English bilinguals from establishing a new category for either of these two vowels. On the other hand, neither early nor late bilinguals were expected to have no difficulties producing the English /æ/-/e/ contrast, as both members of this contrast should be perceived as dissimilar (“new”) from any Serbian vowel.

Production results revealed the opposite: late bilinguals failed to establish separate categories for the /æ/-/e/ contrast, substituting both /æ/ and /e/ with /e/. Conversely, the late learners did establish separate categories for the /i:/ - /I/ contrast...
producing /ɪ/ 96.4% and /iː/ 93.6% of the time as intended. Early learners produced both contrasts accurately. Early learners’ mean production accuracy for the /ɪ/ and /iː/ vowels was 100% and 98% respectively. Their mean production accuracy for both /æ/ and /e/ vowels was 97.3%.

In order to investigate how early and late Serbian-English bilinguals perceptually related AusE /æ/-/e/ and /iː/-/ɪ/ contrasts to the native Serbian categories, the Perceptual assimilation experiment with category goodness ratings was conducted. The aim of this experiment was to determine the perceptual relationship between AusE vowels in the /æ/-/e/ and /iː/-/ɪ/ contrasts and the closest Serbian vowels.

7.4 Experiment 2: Perceptual assimilation and goodness-of-fit rating

7.4.1 Design

The independent variables were age of learning (AOL) with two between-subject levels (early learners: EL, and late learners: LL), English stimulus vowel contrasts (EVC) with two within-subject levels /æ/-/e/ (EVC1) and /iː/-/ɪ/ (EVC2), Serbian response-choice vowels (SRCV) with two within-subject levels (Serbian vowel contrast SVC: /a/-/e/ and Serbian vowel SV: /i/). The dependent variable was the percentage of identification for the Serbian response-choice vowels.

7.4.2 Procedure

Each trial consisted of two presentations of a target stimulus token: in the first presentation of one of the four AusE vowels in the h_d frame the participants were asked to choose one of the five orthographic representations for Serbian vowels presented on the computer screen that matched the same vowel as the AusE word they just heard. The possible responses were A, E, I, U, and O (the orthographic representations for five short Serbian vowels: /a/, /e/, /i/, /u/ and /o/).

Before the experiment the participants were given oral and written instructions then completed 8 practice trials (two AusE vowels repeated 4 times) with the experimenter present. After the practice trials were completed, the participants competed the experiment.
After hearing an AusE word with one of the targeted vowels via loudspeaker, participants were required to click on the box on computer screen containing the corresponding Serbian best-matching vowel. Responses were given using the computer mouse. Once the response was given, after a 1000ms interval participants heard the same target word again. This time the numbers from 1 to 5 appeared on the bottom of the screen. The listeners had to give their goodness-of-fit rating between AusE and chosen Serbian vowel using a 5-point scale (“1” = very little similarity between the AusE and the Serbian vowel, “5” = exactly the same vowel). The next trial started 1000 ms after the rating response was given.

There were 108 randomly presented trials (4 vowels x 3 speakers x 3 tokens x 3 repetitions). The stimuli were presented using the same equipment as previously described.

7.4.3 Results: Perceptual assimilation of the /e/-/æ/ contrast

Percent of AusE vowels perceptually assimilated to the Serbian vowels by Early and Late Serbian-English bilinguals was averaged for each participant and then each group’s mean score was derived from these values. Two separate ANOVAs were conducted for each contrast. The results for the /e/-/æ/ contrast are presented in Figure 7-3. Early learners perceptually assimilated AusE /æ/ to Serbian /a/ 95.9% of the time, and AusE /e/ to Serbian /e/ 98.3%. Late learners, on the other hand, perceptually assimilated AusE /æ/ to Serbian /e/ 75.6% and AusE /e/ to Serbian /e/ 95.8% of the time.
Figure 7-3. Perceptual assimilation of English had, and head by early (EL) and late (LL) Serbian-English bilinguals. On the x-axis Serbian target vowels /a/ and /e/ are presented. The percentage of perceptually assimilated English vowels to Serbian vowels is presented on the y-axis. Error bars represent standard error of the mean.

The results of the ANOVA on the perceptual assimilation of EVC1 are presented in Table 7-3.
**Table 7-3**

Analysis of Variance for perceptual assimilation of the /e/-/æ/ contrast

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>1</td>
<td>0.00</td>
<td>0.99</td>
</tr>
<tr>
<td>S between-group error</td>
<td>28</td>
<td>(0.47)</td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVC1</td>
<td>1</td>
<td>2.02</td>
<td>0.17</td>
</tr>
<tr>
<td>SVC1</td>
<td>1</td>
<td>355.43**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC1</td>
<td>1</td>
<td>0.53</td>
<td>0.47</td>
</tr>
<tr>
<td>AOL x SVC1</td>
<td>1</td>
<td>290.50**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC1 x SVC1</td>
<td>1</td>
<td>1034.28**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC1 x SVC1</td>
<td>1</td>
<td>467.60**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC1 x SVC1 x S within group error</td>
<td>42</td>
<td>(94.30)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects.
* p < 0.05. **p < 0.01.

There was a significant main effect of SVC1, a significant interaction between AOL and SVC1, a significant interaction between EVC1 and SVC1 and a significant three-way interaction among AOL, EVC1 and SVC1. Group comparisons (EL vs. LL) in simple effects test of the AOL x EVC1 x SVC1 interaction revealed that the interaction was significant: AOL (EL vs LL) x EVC1 x SVC1, F (1,9) = 447.09, p < 0.01.

Early and late learners’ perceptual assimilation of the /æ/-/e/ contrast reflected the production pattern obtained in the Experiment 1. Early learners produced both /æ/ and /e/ accurately and perceptually related these two AusE categories to two different Serbian vowels, /a/ and /e/. Late learners, on the other hand, failed to distinguish vowels /æ/ and /e/, producing them both as /e/ and perceptually assimilating both /æ/ and /e/ to a single Serbian vowel, /e/.
7.4.4 Results: Perceptual assimilation of the /ɪ/-/ɪ:/ contrast

The results for the /ɪ/-/ɪ:/ contrast are presented in Figure 7-4. Both early and late Serbian-English bilinguals perceptually assimilated AusE /ɪ/ and /ɪ:/ to Serbian vowel /i/. However, two participants from the early learner group consistently choose Serbian /e/ category for AusE /i/ (2.6%).

![Figure 7-4](image-url)

*Figure 7-4. Perceptual assimilation of English heed and hid by early (EL) and late (LL) Serbian-English bilinguals. On the x-axis Serbian target vowels /i/ and /e/ are presented. The percentage of perceptually assimilated English vowels to Serbian vowels is presented on the y-axis. Error bars represent standard error of the mean.*

The result of the ANOVA on the perceptual assimilation of EVC2 are presented in Table 7-4. There was a significant main effect of SVC2 and a significant interaction between EVC2 and SVC2. There was no difference in perceptual assimilation of EVC2 between early and late learners.
Table 7-4
Analysis of Variance for perceptual assimilation of the /u/-/i:/ contrast

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>1</td>
<td>0.86</td>
<td>0.36</td>
</tr>
<tr>
<td>S between-group error</td>
<td>28</td>
<td>(0.07)</td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVC2</td>
<td>1</td>
<td>0.86</td>
<td>0.36</td>
</tr>
<tr>
<td>SVC2</td>
<td>1</td>
<td>36082.58**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC2</td>
<td>1</td>
<td>0.86</td>
<td>0.36</td>
</tr>
<tr>
<td>AOL x SVC2</td>
<td>1</td>
<td>0.77</td>
<td>0.39</td>
</tr>
<tr>
<td>EVC2 x SVC2</td>
<td>1</td>
<td>7.79**</td>
<td>0.01</td>
</tr>
<tr>
<td>AOL x EVC2 x SVC2</td>
<td>1</td>
<td>3.23</td>
<td>0.08</td>
</tr>
<tr>
<td>EVC2 x SVC2 x S within</td>
<td>28</td>
<td>(8.84)</td>
<td></td>
</tr>
<tr>
<td>group error</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects.
* p < 0.05. **p < 0.01.

7.4.5 Results: Goodness-of-fit rating for the /e/- /æ/ contrast

The rating scores for the /e/- /æ/ contrast are presented in the Table 6-5. Separate t-test analyses were conducted for each group. For both early and late Serbian-English bilinguals, AusE /e/ was a better instance of the chosen Serbian category than was AusE /æ/. Early bilinguals’ mean similarity score for AusE /e/ and Serbian /e/ was 3.33, for AusE /æ/ and Serbian /æ/ was 2.60. This difference was significant: t(14) = 34, p < 0.01. Late bilinguals’ mean similarity score for AusE /e/ and Serbian /e/ was 3.62, for AsE /æ/ and Serbian /e/ was 2.49. This difference was also significant: t(14) = 103.55, p < 0.01.
Table 7-5
Mean ratings for had and head

<table>
<thead>
<tr>
<th></th>
<th>/had/</th>
<th>/head/</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOL</td>
<td>/a/</td>
<td>/e/</td>
</tr>
<tr>
<td>EL</td>
<td>2.60</td>
<td>-</td>
</tr>
<tr>
<td>LL</td>
<td>-</td>
<td>2.49</td>
</tr>
</tbody>
</table>

7.4.6 Results: Goodness-of-fit ratings for the /i:/- /i/ contrast

The rating scores for the /i:/- /i/ contrast are presented in Table 7-6. Separate t

test analyses revealed that both early and late Serbian-English bilinguals perceived

AusE /i/ as a better instance of the Serbian /i/ than AusE /i:/.

Early learners’ mean rating score for AusE /i/ was 3.43, for AusE /i/ was 2.48. This difference was significant:

\[ t(14) = 28.40, p < 0.01 \]

Late learners’ mean goodness-of-fit rating score for AusE /i/ was 3.72, for AusE /i/ was 2.37. This difference was also significant:

\[ t(14) = 11, p < 0.01 \]

Table 7-6
Mean ratings for heed and hid

<table>
<thead>
<tr>
<th></th>
<th>/heed/</th>
<th>/hid/</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOL</td>
<td>/e/</td>
<td>/i/</td>
</tr>
<tr>
<td>EL</td>
<td>-</td>
<td>2.48</td>
</tr>
<tr>
<td>LL</td>
<td>-</td>
<td>2.37</td>
</tr>
</tbody>
</table>
7.4.7 Summary of the results for perceptual assimilation and goodness-of-fit ratings

The Perceptual Assimilation experiment was carried out with the aim to establish which Serbian vowel categories the Serbian-English bilinguals perceptually related AusE /e/-/æ/ and /i:/-/ı/ contrasts to. In addition, category goodness ratings were obtained in order to capture the perceptual distance between the AusE vowels and the chosen Serbian category.

For the /e/-/æ/ contrast early and late learners displayed different assimilation patterns. Early learners perceptually assimilated AusE /e/ to Serbian /e/, and AusE /æ/ to Serbian /a/. Late learners perceptually assimilated both /e/ and /æ/ to Serbian /e/. However, for both early and late learners, AusE /e/ was a better instance of the chosen Serbian vowel than AusE /æ/.

Both groups of Serbian-English bilinguals perceptually assimilated vowels from the /i:/-/ı/ contrast to Serbian /ı/. Furthermore, for both groups AusE /ı/ was a better instance of the Serbian /ı/ than AusE /i:/ /ı/. The assimilation pattern for the /i:/-/ı/ contrast is not surprising, given that the possible responses were Serbian /a/, /e/, /ı/, /o/, and /u/. Only Serbian /ı/ occupies the high front position in the vowel space close to the positions occupied by AusE /ı:/ and /ı/ vowels. Therefore, the Serbian vowel /e/ as a perceptual representation of AusE /ı/ chosen by two participants in the early group was rather unanticipated. A possible cause of this confusion may be in orthographic representations of the Serbian vowels offered to the participants on the computer screen. The orthographic representations for Serbian /e/ in both the Latin and the Cyrillic alphabet is the letter E (a common spelling for English /ı/). Given that the two bilinguals who identified AusE /ı:/ as Serbian /e/ belong to the group of early Serbian-English bilinguals, it was likely that spelling confusion owing to the predominant use of English spelling had governed their choices in perceptual assimilation task.

These results lead to Experiment 3. The aim of the Experiment 3 was to investigate the relationship between the perceptual assimilation and discrimination of the non-native vowel contrasts.
7.5 Experiment 3: Perceptual discrimination

7.5.1 Design

The independent variables were age of learning (AOL) with two between-subject levels (early learners: EL; and late learners: LL), and English stimulus vowel contrasts (EVC) with two within-subject levels /æ/-/e/ (EVC1) and /i:/-/u/ (EVC2). The dependent variable was the percentage of accurate discrimination.

7.5.2 Participants

In addition to the Serbian early and late learners of AusE, a group of 15 monolingual native speakers of Australian English participated in the experiment as a control group. The same Serbian-English bilinguals who participated in the perceptual assimilation task also participated in the discrimination task.

7.5.3 Procedure

Serbian-English bilinguals’ phonetic sensitivity to AusE /e/-/æ/ and /i:/-/u/ vowel contrasts was assessed in an oddity discrimination task (Flege, MacKay et al., 1999). In this task three stimuli were presented in each trial. The participants were told to identify the position of an odd item out if they heard the vowel in the stimuli as different from the other two. The subjects were offered four possible response alternatives on the computer screen: “1,” “2” and “3” if they identified an odd item out in the first, second or third position, respectively, or “none” if they heard three different examples of a single vowel (“catch trials”). The inter-stimulus interval (ISI) among the three stimuli in each trial was 1000 ms. Before the experiment, subjects were given oral and written instructions. In order to familiarize participants with the task twelve practice trials with the vowels /o:/ and /u/ (four trials for each vowel) were conducted in the experimenter’s presence before the participants started the experiment.

In the experiment, each vowel contrast was tested in 36 trials: 18 trials contained an odd item out and 18 trials contained the same vowel in all three positions. All trials were presented in one block in random order. There were nine trials for each of the two vowels in a contrast with no odd item out, and nine trials per contrast with an odd item out. The odd item occurred in each of three positions with equal frequency. For example, to test the /i:/-/u/ contrast, nine /u/-/i/-/u/ and nine /i:/-/i:/-/i:/trials were
presented to the participants. There were nine different trials in which /i:/ was an odd item and nine different trials in which /u/ was an odd item.

The tokens in each trial were always produced by three different speakers (physically different tokens). This encouraged the participants to respond to only to phonetically relevant differences, not only to any auditorily detectable difference. An odd item out occurred in each position with equal frequency. To successfully discriminate vowels the participants had to recognize the categorical identity of a set of physically different tokens of the same vowel category while ignoring acoustic/auditory differences among instances of the category, which were phonetically irrelevant to their categorical identity.

The discrimination experiment was actually conducted prior to the assimilation and rating task, in order to avoid biasing the discrimination results by prior experience categorizing the stimuli to Serbian vowels.

7.5.4 Analysis

First, d–prime (d’ scores) were calculated for each vowel contrast examined. These scores were derived from the proportion of “hits” (correct selection of an odd item in different trials) and “false alarms” (incorrect selections of an odd item out in the catch trials) obtained for each contrast. D’ was calculated as the difference between standardised scores of these two rates (d’ = z(H) - z(F)). The d’ scores provide an unbiased measure of perceptual sensitivity by taking into account the responses to the different trials and the catch trials. In order to avoid infinite values for standardized scores for hit rates of 100% or false alarm rates of 0% a constant of 0.01 was subtracted (from the former) or added (to the latter) prior to obtaining corresponding standardized scores. When H = F, d’= 0 (no sensitivity). The highest possible d’ (greatest sensitivity) was 6.93. 69% correct for both different and same trials corresponds to a d’ of 1.0. Note that it is possible in this way to obtain negative value for d’ which would then indicate participants’ strong preference to false alarm response Higher values of d’ indicate strong ability to discriminate contrasts, whereas values bellow 0.5 indicate lower discrimination ability. The groups d’ scores are presented on the Figure 7-5.
Figure 7-5. Mean d-prime (d’) scores obtained for the early (EL) and late (LL) Serbian-English bilinguals and for the AusE monolinguals (M) for the /e/-/æ/ (EVC1) and /i:/ - /u/ (EVC2) contrasts. Mean d’ scores are presented on the y-axis. The AusE vowel contrasts are presented on the x-axis.

7.5.5 Results

An ANOVA revealed a significant main effect of AOL, a significant main effect of EVC and a significant interaction between EVC and AOL (see Table 7-7).

Table 7-7
Analysis of Variance for discrimination

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>2</td>
<td>136.87**</td>
<td>0.00</td>
</tr>
<tr>
<td>S between-group error</td>
<td>42</td>
<td>(17.83)</td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVC</td>
<td>1</td>
<td>151.33**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC x AOL</td>
<td>2</td>
<td>193.30**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC x AOL x S within group</td>
<td>42</td>
<td>(12.75)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects.

* p < 0.05. **p < 0.01.
Overall, the /i:/-/u/ contrast was better discriminated the /æ/-/e/ contrast. Group comparisons (EL vs. LL, EL vs. M and LL vs. M) in simple effects tests of the EVC x AOL interaction revealed that the interaction was significant for the comparison of early versus late learners (AOL[LL, EL] x EVC, F (1,42) = 67.24, p < 0.01), as well as for the comparison of late learners versus monolinguals (AOL[LL, M] x EVC, F (1, 42)= 74.96  p < 0.05). This interaction was not observed when early bilinguals and AusE monolinguals were compared. Both early and late learners were able to discriminate the /i:/-/u/ contrast. Late learners had difficulties discriminating the /æ/-/e/ contrast.

7.5.6 Summary of the results of Experiment 3

The results showed that early Serbian-English bilinguals, but not the late Serbian-English bilinguals, accurately discriminated the AusE /e/-/æ/ vowel contrast. Both early and late Serbian-English bilinguals were able to discriminate the AusE /i:/-/u/ vowel contrast. However, early bilinguals’ discrimination scores were lower than the AusE monolinguals’ for both contrasts.

7.6 General discussion for Study 1

The purpose of this study was to evaluate the SLM’s hypotheses regarding the effect of age on perceived cross-language phonetic similarity (and dissimilarity) on the learning of L2 vowels. Production and perception accuracy of Australian English /e/-/æ/ and /i:/-/u/ vowel contrasts by early and late Serbian-English bilinguals was examined. The SLM predicts that those L2 vowels that are perceived as dissimilar from the closest L1 vowel will be produced more accurately than those L2 vowels that are perceived as similar to the closest L1 vowel. Moreover, the earlier in life that L2 learning begins, the more likely L2 learners will be to establish new phonetic categories for L2 vowels. This prediction is based on the hypothesis that as phonetic units making up the L1 vowel system develop, they become more powerful attractors of L2 vowels, thereby inhibiting the development of new phonetic categories for L2 vowels.

It was hypothesized that early and late Serbian-English bilinguals would differ in production and perception of AusE /e/-/æ/ and /i:/-/u/ vowel contrasts. Production accuracy of these vowels was expected to be reflected in the way Serbian-English
bilinguals perceptually related the targeted AusE vowels to their native Serbian vowels. Late learners were expected to perceive the vowels from /e/-/æ/ contrast as dissimilar to any of their native vowel categories and therefore to produce them more accurately. The vowels from the /i:/-/ɨ/ contrast were expected to be perceived as close to Serbian /ɨ/ and therefore difficult for late learners to produce.

Consistent with the previous findings on AOL effects in production of L2 vowels (Flege, Bohn et al., 1997; M. Munro et al., 1995), early and late Serbian-English bilinguals differed in production accuracy on the AusE /e/-/æ/ contrast. However, contrary to the hypothesis, late learners failed to produce the /e/-/æ/ contrast accurately but had no difficulties in production of the /i:/-/ɨ/ contrast. The difference in late learners’ production accuracy of these two AusE contrasts is consistent with the hypothesis that L2 vowels differ in their learnability (Flege, 1997) and that late learners may fail to establish a new category for a similar L2 vowel due to the equivalence classification mechanism. The surprising finding, on the other hand, was that the /e/-/æ/ vowel contrast was expected to be perceived as dissimilar from any other Serbian vowel and therefore relatively easy for late Serbian-English bilinguals to produce. Instead of establishing new categories for both /e/ and /æ/ vowels, late learners produced both /e/ and /æ/ as /e/. The /i:/-/ɨ/ contrast, conversely, was expected to be perceived as similar to Serbian /ɨ/ and the category formation for these vowels was expected to be blocked by the equivalence classification mechanism. Contrary to the hypothesis, the late bilinguals established separate categories for both AusE /i:/ and /ɨ/ vowels.

The production data obtained in Experiment 1 suggested that early and late Serbian-English bilinguals differed in perceptual assimilation of the /e/-/æ/ contrast to the closest vowel in Serbian phonetic inventory (Flege et al., 1994). Early learners perceptually related the /e/-/æ/ contrast to two different Serbian categories (/e/ and /a/). Late learners perceptually assimilated both AusE /e/ and /æ/ to a single Serbian category (/e/). For both groups AusE /e/ was rated as a better instance of the chosen Serbian category than AusE /æ/.

Perceptual assimilation and rating data obtained for the AusE /i:/-/ɨ/ contrast did not differ between the groups: both groups of Serbian-English bilinguals identified the
vowels from the /i:/-/ɪ/ contrast as instances of the Serbian /i/, with /ɪ/ being rated as closer match to the Serbian category than /i:/.

Finally, comparison of the perceptual assimilation data showed that early and late learners differed in discrimination of the /e/-/æ/ contrast: early bilinguals had no difficulties discriminating between /e/ and /æ/ vowels while late learners displayed poor discrimination between these two vowels. Both groups were able to discriminate the /i:/-/ɪ/ contrast accurately.

Furthermore, late Serbian-English bilinguals displayed the same perceptual assimilation patterns for both AusE /e/-/æ/ and /i:/-/ɪ/ contrasts: both contrasts were perceptually assimilated to a single Serbian category (/e/ and /ɪ/, respectively) but one member of the contrast was rated as a better instance of the chosen Serbian category. However, the late learners had greater difficulty discriminating the /e/-/æ/ contrast than the /i:/-/ɪ/ contrast.

The assimilation patterns might be interpreted within the Perceptual Assimilation Model framework (Best, 1995) which was proposed to account for variability in discrimination of non-native contrasts. According to the PAM, discrimination is expected to be excellent when each member of a non-native contrast is assimilated to a different native category (two-category assimilation type: TC), whereas poor discrimination is expected when both members of a non-native category are assimilated as equally good instances to a single native category (single category assimilation type: SC). Good to moderate discrimination is expected when both phones of a non-native contrast are assimilated to a single native category (category goodness difference assimilation type: CG) but differ in the goodness-of-fit to that category. The relative difficulty among category goodness contrasts depends on the degree of category goodness difference between the contrasting phones. According to the PAM’s predictions, early learners perceptually assimilated the /e/-/æ/ contrast as a TC type, hence the discrimination was good. Late learners perceptually assimilated the same contrast as a CG difference, but the category goodness between the contrasting phones seemed to be insufficient to allow good discrimination. However, this study did not provide enough evidence for the interpretation of the perceptual assimilation pattern of the /i:/-/ɪ/ contrast: only Serbian /i/ occupies the high front position in the vowel space.
close to the positions occupied by AusE /i:/ and /u/ vowels and therefore Serbian /i/ was the only one choice in the perceptual assimilation task. On the other hand, the different goodness-of-fit rating for /i:/ and /u/, and excellent discrimination, imply that /i:/ and /u/ were perceived as two different categories, but both close to only one of the Serbian vowels offered as possible responses in the perceptual assimilation task. It is possible that Serbian-English bilinguals relied on their experience with vowel duration in Serbian to distinguish between AusE tense-lax vowels /i:/ and /u/ (Bohn, 1995). Given the choice between Serbian /i:/ and /i/, the participants in the study would perceptually assimilate AusE /i:/ and /u/ to Serbian /i:/ and /i/, respectively.

Taken together, the results of this study provided support for certain predictions based on the SLM model. The production accuracy of L2 vowel contrasts seemed to be a reliable predictor of perceptual accuracy with these contrasts. L2 learners’ relative degree of accuracy in producing L2 speech segments varied as a function of their perceived relation to the segments in their L1 inventory (Flege, 1995). Furthermore, the perceived relation between L2 and L1 segments appears to change as the AOL increases.

On the other hand, predictions based on the SLM’s criteria for establishing the degree of difficulty in L2 vowel acquisition were not upheld in this study. The late learners had no difficulties with the vowels in the contrast that was predicted to be both “similar” to the vowel category in their native language, but they had great production and discrimination difficulties with the non-native vowel contrast which was expected to be “different” from any vowel in their L1.

Additionally, comparison of the perceptual assimilation and discrimination data addressed the question of AOL-related variability in discrimination of L2 vowel contrasts. Another important question addressed was the relationship between the extent to which L2 phonetic segments can be assimilated to native vowel contrasts and the discriminability of these L2 phonetic contrasts (Best, 1995) by early and late bilinguals. The SLM framework was inadequate to account for the relationship between the perceptual assimilation patterns and discriminability of non-native vowel contrasts.

Study 2, therefore, aimed to further investigate the effects observed in the Study 1 by applying predictions based on the PAM. In this study Serbian-English bilinguals’ experience with vowel duration distinctions in Serbian was not taken into account. The
responses in the perceptual assimilation experiment were limited to the short Serbian vowels only. In this limited pool of response choices, it was expected that both groups of bilinguals would perceptually assimilate AusE lax-tense vowel pair /i:/ and /ɪ/ to the Serbian /i/. However, their experience with native vowel duration contrasts might have played a role in discrimination task. Therefore it was decided that all Serbian vowels, long and short should be included as possible responses in a second perceptual assimilation the perceptual assimilation experiment in Study 2.
CHAPTER 8

STUDY 2. PERCEPTION AND PRODUCTION OF AUSTRALIAN-ENGLISH /æ/ - /e/ AND /i:/ - /u/ BY SERBIAN-ENGLISH BILINGUALS:

IMPLICATIONS FOR THE PERCEPTUAL ASSIMILATION MODEL

Cross-language speech perception research has shown that adults often have difficulty discriminating non-native segments (Polka, 1991, 1992) with discrimination accuracy varying from very poor to very accurate for different non-native contrasts (Best et al., 1988; Larson-Hall, 2004). The results obtained in the Study 1 (chapter 6) are in agreement with these findings: for late Serbian-English bilinguals the AusE /e/-/æ/ vowel contrast was more difficult to discriminate than the AusE /i:/-/u/ contrast. Early Serbian-English bilinguals, on the other hand, did not differ in discrimination accuracy on these two AusE contrasts. Furthermore, early and late learners differed in perceptual assimilation of the /e/-/æ/ vowel contrast to the closest vowel in the Serbian inventory. Early learners perceived /e/ and /æ/ vowels as instances of two different Serbian categories (/e/ and /a/), whereas late learners perceptually related both /e/ and /æ/ to a single native category (/e/). The groups did not differ in perceptual assimilation and discrimination of the /i:/- /u/ contrast. Both /i:/ and /u/ were perceived as instances of Serbian /i/ and were easy to discriminate regardless of the age of L2 learning. The production data suggested that late learners’ production difficulties were perceptually related: the contrast that was difficult to discriminate was produced inaccurately. Late learners substituted both AusE /e/ and /æ/ with /e/. The SLM framework offers a possible interpretation of these findings: when L2 phonological segments are perceived as phonetically similar to the closest segment in the bilinguals’ L1 segmental inventory, the equivalence classification mechanism prevents the establishment of a new category for the L2 segment. The more dissimilar the L2 and L1 phonetic segments are, the more likely that new L2 phonological categories will be developed.
The SML framework does not, however, provide reliable criteria for establishing the degree of dissimilarity between L2 and L1 phonetic segments that will allow a new L2 category to be established. Furthermore, the SLM addresses primarily the acquisition of production of individual L2 phonological segments by relatively experienced adult L2 speakers, and makes no explicit predictions about discrimination of non-native contrasts. Given evidence that speakers who discriminate well between vowel stimuli (Perkell, Guenther et al., 2004) and consonant stimuli (Perkell, 2004) with subtle acoustic difference produce relatively more clear-cut vowel and consonant contrasts than speakers who discriminate less well between the same pairs, the relative discriminability of non-native contrasts may be a predictor of production accuracy of these segments in L2 learning.

Variability in discrimination accuracy of non-native contrasts is addressed in the Perceptual Assimilation Model (PAM) (Best, 1994, 1995; Best & McRoberts, 2003). According to the PAM, discrimination of a non-native contrast depends on how each member of the contrast is assimilated to a native segment. Discrimination is expected to be excellent for two-category assimilation (TC), relatively good for category goodness difference assimilation (CG) and poor for single category assimilation (SC).

PAM was initially devised to account for monolinguals’ perception of unfamiliar non-native contrasts. When initially exposed to non-native contrasts, listeners perceptually assimilate unfamiliar speech segment to the articulatorily closest speech segment in their native language due to their native language experience. However, the model’s predictions are also relevant for adults with effective exposure and conversational experience with an L2 (Best & Tyler, 2007). Experienced listeners may perform better in a discrimination task than inexperienced listeners (Best & Strange, 1992; Larson-Hall, 2004; Mackain, Best, & Strange, 1981). Evidence from research in perceptual learning (Bradlow et al., 1997; Pisoni, Aslin, Perey, & Hennessy, 1982) suggest that it is possible to modify the perceptual mechanisms utilized by adults in categorizing non-native speech segments. The question of special interest for the experiments reported in this chapter is how these perceptual mechanisms are modified in bilinguals who differ in AOL.

It has been proposed that with increasing AOL, L2 learning is hindered by a loss of neuroplasticity that accompanies normal neural maturation (Scovel, 2000). Evidence from cross-language speech-perception research has suggested that AOL effects in
perception of non-native phonetic segments must be understood in the context of L1 perceptual attunement and its constraints on acquisition of L2 phonological system (Hojen & Flege, 2006) (Iverson et al., 2003; Kuhl, 2000; Pallier et al., 2003). For example, Spanish-Catalan bilinguals who began learning L2 Catalan before the age of 6 years performed significantly worse than native Catalan speakers in a lexical decision task involving discrimination between Catalan /e/ and /e/, due to perceptual tuning to their L1 (Sebastian-Galles, Echeverria, & Bosch, 2005).

The aim of the experiments reported in this chapter was to extend PAM’s non-native speech perception framework to account for bilingual speakers differing in AOL. This was done by examining the relationship between the discrimination and identification of AusE /e/-/æ/ and /i:/-/u/ contrasts in early and late Serbian-English bilinguals. Given that in Study 1 Serbian-English bilinguals experience with vowel duration was not taken into account, in Study 2 this omission was corrected by inclusion of both Serbian short vowels /i/, /e/, /a/, /o/, and /u/ (as in Study 1) and their long counterparts /i:/, /e:/, /a:/, /o:/ and /u:/ in perceptual assimilation and discrimination experiments. This vowel choice was motivated by the late learners’ accurate perception of expectedly “difficult” AusE /i:/-/u/ contrast. It was speculated that in discriminating between AusE /i/ and /i:/, late learners perceived temporal distinction between AusE lax (/i/) and tense (/i:/) vowels (rather than the spectral differences between them) using their experience with vowel duration in Serbian.

As in Study 1 (Chapter 7), production and perception of AusE /e/-/æ/ and /i:/-/u/ contrasts were tested in one production and two perception experiments in Study 2. In the production experiment of Study 2, native monolingual speakers of AusE judged L2 vowel production accuracy by identifying the vowels produced by Serbian-English bilinguals. Vowel perception was tested in two experiments in Study 2: a) perceptual discrimination and (b) perceptual assimilation with goodness-of-fit ratings.

8.1 Predictions based on PAM

According to the PAM, vowels from the English contrast /i:/-/u/ were expected to be perceptually assimilated to two distinct Serbian vowels, short /i/ and long /i:/ (TC assimilation). In Study 1 (chapter 6) it was found that both early and late Serbian-
English bilinguals perceptually assimilated AusE vowels from the /i:/-/u/ contrast to a single Serbian vowel /i/. Rating scores revealed that AusE /i/ was a closer match to Serbian /i/ than AusE /i:/ for both groups suggesting that the /i:/-/u/ contrast was assimilated as CG assimilation type. In addition, Serbian-English bilinguals had no difficulties discriminating between these two AusE vowels, regardless of the AOL. Thus, the hypotheses based on the SLM framework were not supported. The assimilation choice for the /i:/-/u/ vowels was not surprising, since only Serbian short /i/ from the five short Serbian vowels given to the participants (/a/, /e/, /i/, /o/, /u/) as possible responses, occupied the high front position close to AusE /i:/ and /u/. Given their familiarity with vowel length distinctions in Serbian, the Serbian speakers’ excellent performance in the /i:/-/u/ discrimination task was the consequence of their sensitivity to the quantitative acoustic differences between AusE /i:/ and /u/. In that case, the vowels from AusE /i:/-/u/ contrasts may actually have been assimilated to two different Serbian categories /i/ and /i:/ as a two category assimilation type (TC) but the Serbian vowel choices they had been given did not allow them to indicate this length difference. Hence discrimination was good although both /i:/ and /u/ appeared to have been assimilated to a single category /i/.

By comparison, for the /e/-/æ/ contrast in Study 1 early and late learners displayed different assimilation patterns. Early learners perceptually assimilated AusE /e/ to Serbian /e/, and AusE /æ/ to Serbian /a/. Furthermore, early learners had no difficulties discriminating between these two AusE vowels. Late learners, however, perceptually assimilated both /e/ and /æ/ to Serbian /e/ and performed poorly in discrimination task. However, for both early and late learners, AusE /e/ was a better instance of the chosen Serbian vowel than AusE /æ/.

These results raise two possibilities within the PAM framework. Regarding AOL effects on perception of the /e/-/æ/ contrast, early and late Serbian-English bilinguals would be expected, according to PAM, to demonstrate two different perceptual assimilation types: 1) early bilinguals would perceptually assimilate the /e/-/æ/ contrast to two Serbian categories (TC assimilation), hence discrimination should be
excellent; 2) late bilinguals should perceptually assimilate both vowels to a single Serbian category /e/. If both AusE /æ/ and /e/ are equally good matches to the chosen Serbian vowel category (SC assimilation) discrimination is expected to be poor. If, on the other hand one AusE vowel is a better instance of the Serbian /e/ than the other vowel from the contrast, (CG assimilation) discrimination is expected to be relatively good.

8.2 General method

8.2.1 Participants

As in the previous study, the participants were 30 Serbian-English bilinguals differing in age of learning (AOL) of Australian English. None of the participants had taken part in the previous experiments. The AOL AusE requirement was less than 5 years of age for the early learners, and late learners had to acquire AusE after the age of 18. An additional requirement for the late learners was to have been using English at least 5 years before the time of testing. Both early and late learners were born in Serbia and migrated to Australia and had to be fluent in Serbian, which had to be their first language (L1). The same language conditions were maintained throughout the testing as in Study 1 (Language Background Questionnaire in Serbian and in English are given in Appendices 7.1 and 7.2. Participants’ information statement and consent form are provided in Appendices 7.3 and 7.4)

Fifteen native monolingual speakers of AusE served as the control group in the production and discrimination experiments. In the production experiment an additional 10 native AusE monolinguals participated as listeners judging production accuracy. All AusE participants were recruited from the School of Psychology and the School of Humanities and Languages at the University of Western Sydney.

8.2.2 Experience with English prior to migration to Australia

All language requirements and participants’ experience with English prior to migration to Australia were the same as in Study 1.

8.2.3 Age of learning, length of exposure and chronological age

There were 15 early and 15 late learners in each group. The mean age, AOL, and length of AusE use for the early learners was 27.8 years (SD = 5.6), 3.7 years (SD 1.1)
and 24.2 years (SD = 5.2). The mean age, AOL and length of AusE use for late bilinguals was: 33.5 years (SD = 5.7), 22.4 years (SD=6.0) and 10.9 years (SD = 4.5). The mean age of AusE monolinguals in the control group was 21.8 years (SD = 2.4).

8.2.4 Stimuli

As in Study 1, AusE vowels /æ/, /e/, /i:/, and /u/ were produced in h_d frame (had, head, heed, hid) embedded in the sentence frame “Hid is the next word to say. I say hid again. Hid.” The speakers were ten male monolingual AusEs.

The recording procedure and stimulus editing did not differ from that reported in Chapter 7 for Study 1. The sentences with the target stimuli were repeated 10 times for each vowel in random order across the repetitions.

Three tokens in citation form for each vowel were chosen from three different speakers and used as stimuli for the elicitation of production data and as stimuli in perception experiments for the Serbian participants. The use of multiple speakers, as in Study 1, forced the participants to focus on higher-order speech information instead of lower-level acoustic match within a single speaker. Additionally, the larger pool of speakers enabled more careful matching between the chosen tokens.

8.2.5 Recording equipment

Recording equipment, procedure and venue were the same as in Study 1.

8.2.3 Experiments

All participants completed three experiments, in the following order: the discrimination experiment (AXB) (presented as Experiment 5) the perceptual assimilation and goodness-of-fit rating experiment (presented as experiment 4) and the production experiment (presented as Experiment 6). Since one of the PAM’s predictions is that the assimilation pattern of non-native segments to native segments is a predictor of discrimination accuracy, the perceptual assimilation experiment will be presented first, followed by the results of the discrimination experiment and production accuracy data.

The same testing contexts as in Study 1 were used.
8.3 Experiment 4: Perceptual assimilation and goodness-of-fit rating

8.3.1 Design

The same design as in the discrimination experiment in Study 1 was used.

8.3.2 Procedure

The aim of this experiment was to determine the perceptual relationship between AusE vowels in the /æ/-/e/ and /i:/-/u/ contrasts and the closest Serbian vowels. Each trial consisted of two presentations of a target stimulus token: in the first presentation of one of the four AusE vowels in the h_d frame the participants were asked to choose one of the Serbian keywords presented on the computer screen that contained the same vowel as the AusE word they just heard. The possible responses were the keywords containing all Serbian vowels, long and short, embedded in the b_ba frame (Table 8-1). There were 10 keywords: 5 for short Serbian vowels (/a/, /e/, /i/, /o/, /u/) and 5 for long Serbian vowels /a:/, /e:/ /i:/, /o:/, /u:/). The Serbian vowel were embedded in b_ba context (e.g. baba, beba, biba, etc.). Five of these keywords were nonwords, and five were words in the Serbian language. The words were: baba (grandmother), beba (baby), buba (ant), baaba (pl. grandmothers), buuba (pl. ants), beeba (babies)

<table>
<thead>
<tr>
<th>Keywords with short vowels</th>
<th>Keywords with long vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>baba</td>
<td>baaba</td>
</tr>
<tr>
<td>beba</td>
<td>beeba</td>
</tr>
<tr>
<td>biba</td>
<td>biiba</td>
</tr>
<tr>
<td>buba</td>
<td>buuba</td>
</tr>
<tr>
<td>boba</td>
<td>booba</td>
</tr>
</tbody>
</table>

Since vowel length is not orthographically represented in Serbian, a double orthographic representation of the short vowels was used to indicate the Serbian words.
with the same vowel but differing in length. This was carefully explained to the Serbian participants with examples from their language.

A keyword perceptual assimilation task allowed for assimilation patterns to be ascertained precisely. Furthermore, possible confusion between the orthographic representations as in Study 1 caused by different spellings in AusE and native Serbian was avoided. In the experiment participants heard a given h_d token with the targeted AusE vowel and were asked to indicate which Serbian vowel the test AusE vowel most closely resembled, by choosing one of the 10 Serbian words presented on computer screen. Once the response was given, after 1000 ms, the participants heard the same AusE target again. For the second presentation, participants were asked to rate the similarity between the English vowel and their chosen Serbian vowel using a 7-point scale (“1” = very little similarity between the AusE and the Serbian vowel, “7” = exactly the same vowel). In the perceptual assimilation experiment in Study 1 a 5-point scale was used. The reason we used larger rating scale here was in that the participants in the previous study were reluctant to use the extreme (1 and 5) vales on the scale. That means that a 5-point scale for the rating task was essentially used as a 3-point scale by the participants. With the inclusion of two more values, 1 and 7, the participants were given more choices to accurately rate the similarity of targeted AusE and closest Serbian vowel(s).

Analysis of the keyword responses and ratings for each stimulus category allowed for the determination of the assimilation pattern the participants used on each target contrast. To be classified as a TC assimilation, each stimulus of the contrast had to be assimilated to a different single native vowel at least 50% of the time. To be classified as an SC assimilation, both stimuli had to be assimilated to the same native vowel at least 50% of the responses and the mean goodness-of-fit ratings of the stimuli could not differ significantly. In a CG assimilation type both stimuli had to be assimilated to a same native vowel at least 50% of the time. In addition, the difference in the mean category goodness ratings between the AusE vowel and the chosen Serbian vowel had to be statically significant.

There were 108 trials presented in random order (4 vowels x 3 speakers x 3 tokens x 3 repetitions). The equipment for stimuli presentation was the same as in Study 1.
8.3.3 Instructions for participants

A problem that the participants often reported about the rating task in Study 1 was their lack of confidence when rating the similarity between AusE and Serbian vowels and their concern that they had not used the rating scale accurately. Since the rating is of a particular importance for the PAM, especially for the CG assimilation pattern, in this second study special attention was paid to the participant’s understanding of the task. They were given detailed descriptive oral and written instructions, then completed 8 practice trials with the experimenter present. The number of practice trials was limited to only two AusE vowels repeated 4 times, with the aim of providing an easy task, to familiarize participants with the task and the position of the Serbian keywords on computer screen.

Before receiving oral instructions, participants read the written instructions in Serbian. Next, the experimenter explained the differences between long and short vowels, giving an example of Serbian words with the long and short vowels given in appropriate contest. These examples were:

1. Na klupi sedi jedan deda. (An old man is sitting on the bench.) (a short vowel /e/ in first syllable)
2. Na klupi sedi mnogo deda. (The old men are sitting on the bench.) (a long vowel /e:/ in first syllable)

In the first example deda is a Serbian word with a short vowel /e/, meaning “old man” (singular). In the second sentence deda is an example for the long /e/, meaning “old men” (plural). Once when the participants were familiar with the difference between the long and short vowels in Serbian, they were instructed that the orthographic representations for each vowel would be used twice to indicate the long vowels. Next they were shown the list of the keywords that they were going to use in the task and asked to read them aloud. This was done in order to confirm that the participants reliably used the new orthographic code to correctly say the long and short vowels.

Once the listeners were familiar with the keywords they were to use in the perceptual assimilation task, they were instructed how to use the rating scale and encouraged to use all 7 points, including the extreme ones, in case the AusE vowel was
a very bad example of the Serbian vowel (1), or conversely, in case it was exactly the same as the Serbian vowel (7). After receiving the instructions, the listeners completed practice trials with the experimenter present. When the practice was completed, they proceeded to the experiment.

8.3.4 Results: Perceptual assimilation of AusE /æ/ and /e/

The percent choices of the Serbian keywords given in response to the English words are presented in Figure 8-1. Early learners’ strongest response for English had was baba (68% of the time); for head it was beba (91.85% of the time). Late learners’ strongest response to both had and head was beba (78.52% of the time for had and 88.89% of the time for head).

The results of the ANOVA on the perceptual assimilation of EVC 1 are presented in Table 8-2.

![Figure 8.1. Perceptual assimilation of AusE /æ/ and /e/ by early (EL) and late (LL) Serbian L2 learners. The bottom line of the x-axis labels presents the AusE keywords with the targeted vowels. The Serbian keywords with the vowels the listeners perceptually assimilated the AusE vowels are presented on the top line of the x-axis labels. On the y-axis the mean percentage of the choice responses is presented.](image)

ANOVA revealed a significant effect of SVC1, a significant interaction between AOL and SVC1 and a significant interaction of EVC1 and SVC1 (see Table 8-1).
A significant three-way interaction among AOL, this English L2 contrast, and the Serbian L1 assimilation vowels revealed that the early and late learners’ perceptual assimilation of the /æ/-/e/ contrast reflected the pattern obtained in Study 1. Early learners perceptually assimilated /æ/ to Serbian /a/, and /e/ to Serbian /e/, whereas late learners perceptually assimilated both /æ/ and /e/ to Serbian /e/.

**Table 8-2**
Analysis of Variance for perceptual assimilation of the /æ/-/e/ contrast.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>1</td>
<td>0.18</td>
<td>0.67</td>
</tr>
<tr>
<td>S between-group error</td>
<td>28</td>
<td>(50.34)</td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVC1</td>
<td>1</td>
<td>3.61</td>
<td>0.07</td>
</tr>
<tr>
<td>SVC1</td>
<td>1</td>
<td>183.60**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC1</td>
<td>1</td>
<td>0.77</td>
<td>0.39</td>
</tr>
<tr>
<td>AOL x SVC1</td>
<td>1</td>
<td>77.85**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC1 x SVC1</td>
<td>1</td>
<td>77.81**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC1 x SVC1</td>
<td>1</td>
<td>50.94**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC1 x SVC1 x S within group error</td>
<td>42</td>
<td>(639.93)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects.
*p < 0.05. ** p < 0.01.

**8.3.5 Perceptual assimilation of AusE /i:/ and /i/**

Both groups’ strongest response to English hid was the short Serbian vowel in biba (early learners 86.67% of the time and late learners 97.04% of the time) and to English heed was the long Serbian vowel in biiba (84.44% for early and 88.15% for late learners) (see Figure 8.2).

The results of the ANOVA on the perceptual assimilation of EVC 2 are presented in Table 8-3.
Figure 8.2. Perceptual assimilation of AusE /i:/ and /u/ by early (EL) and late Serbian learners (LL). The bottom line on the x-axis presents the AusE keywords with the targeted vowels. The Serbian keywords with the vowels the listeners perceptually assimilated the AusE vowels are presented on the top line. On the y-axis the mean percentage of the given responses is presented.

The ANOVA revealed a significant interaction between AOL and SVC2 and a significant interaction of EVC2 and SVC2 (see Table 8-3). Both early and late Serbian-English bilinguals perceptually assimilated AusE /i:/ to Serbian /i:/ and AusE /u/ to Serbian /u/.
Table 8-3.
Analysis of Variance for perceptual assimilation of the /ɪ:/-/ɪ/ contrast

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Between subjects</td>
<td></td>
</tr>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>1</td>
<td>0.31</td>
<td>0.58</td>
</tr>
<tr>
<td>S between-group error</td>
<td>28</td>
<td>(50.34)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVC2</td>
<td>1</td>
<td>3.41</td>
<td>0.08</td>
</tr>
<tr>
<td>SVC2</td>
<td>1</td>
<td>1.57</td>
<td>0.22</td>
</tr>
<tr>
<td>AOL x EVC2</td>
<td>1</td>
<td>0.11</td>
<td>0.75</td>
</tr>
<tr>
<td>AOL x SVC2</td>
<td>1</td>
<td>4.09*</td>
<td>0.05</td>
</tr>
<tr>
<td>EVC2 x SVC2</td>
<td>1</td>
<td>332.50*</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC2 x SVC2</td>
<td>1</td>
<td>1.83</td>
<td>0.19</td>
</tr>
<tr>
<td>EVC2 x SVC2 x S within group error</td>
<td>28</td>
<td>2.73)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects.
*p < 0.05. **p < 0.01.

8.3.6 Perceived similarity to Serbian vowels: Goodness-of-fit ratings for /æ/-/e/

Mean goodness-of-fit ratings for each AusE vowel were submitted to independent t-tests analysis with the aim of assessing how Serbian speakers of AusE perceptually assimilate the two vowel contrasts to their native categories. The mean ratings for vowels /æ/ and /e/ are presented in the Table 8.4.

Table 8-4.
Mean goodness-of-fit ratings ratings for AusE /æ/ and /e/ by early (EL) and late (LL) Serbian-English bilinguals.

<table>
<thead>
<tr>
<th></th>
<th>had</th>
<th>head</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>beba</td>
<td>baba</td>
</tr>
<tr>
<td>EL</td>
<td>3.29</td>
<td>2.97</td>
</tr>
<tr>
<td>LL</td>
<td>2.84</td>
<td>2.83</td>
</tr>
</tbody>
</table>
The goodness-of-fit data revealed that regardless of the different assimilation patterns for AusE /æ/ and /e/, (early learners assimilated AusE /æ/ to Serbian /a/ and AusE /e/ to Serbian /e/, whereas late learners perceptually assimilated both /æ/ and /e/ to Serbian /e/), AusE /e/ was a better instance of each group’s chosen Serbian category than AusE /æ/ was.

Early learners perceptually assimilated AusE /æ/ to Serbian /a/ with a mean goodness-of-fit rating 2.97 and English /e/ to Serbian /e/ with mean goodness-of-fit rating 4.4. The difference between the ratings was significant: t (14) = 20.574, p < 0.05. Late learners perceptually assimilated both /æ/ and /e/ to Serbian /e/ with mean goodness-of-fit ratings 2.85 and 4.29, respectively. The difference between these two ratings was also significant: t(14) = 55.443, p < 0.05.

8.3.7 Perceived similarities to Serbian vowels: Goodness-of-fit ratings for /i:/ and /i/.

Mean goodness-of-fit ratings for the /i:/- /i/ contrast are given in the table 8-5. The t-test analysis of the rating scores revealed that for both groups /i/ was a better match to Serbian /i/ than AusE /i:/ was to Serbian /i:/: t (14) = 18.95, p < 0.05 and t (14) = 18.49 p < 0.05,

<table>
<thead>
<tr>
<th></th>
<th>hid</th>
<th>heed</th>
</tr>
</thead>
<tbody>
<tr>
<td>biba</td>
<td>4.97</td>
<td>3.50</td>
</tr>
<tr>
<td>LL</td>
<td>4.92</td>
<td>3.00</td>
</tr>
</tbody>
</table>

8.3.8 Summary of the results for Experiment 4

Early learners perceptually assimilated each member of the AusE /æ/ - /e/ contrast to two different Serbian vowel categories (/a/ and /e/). This assimilation pattern meets the criteria for a Two Category Assimilation (TC). Late learners perceptually
assimilated both /æ/ and /e/ to a single Serbian category (/e/), but not as equally good examples of that category. Therefore, this assimilation pattern and rating is described as a Category Goodness difference assimilation (CG). Perceptual assimilation patterns for the /i:/-/' contrast were identical for both groups: the vowels were perceived as instances of Serbian /i/ and /i:/, respectively (TC assimilation).

Taken together, the results of the perceptual assimilation experiment are consistent with the view incorporated in both PAM (Best, 1995) and SLM (Flege, 1995) that experience with the native language shapes perceptual learning of L2 speech segments. Furthermore, perceptual assimilation of an L2 category to the native system is constrained with the AOL in that the nature of L1 influence on perceived similarities (and dissimilarities) between L2 and L1 categories changes with increasing experience with a bilingual’s L1. Therefore, with increasing AOL late Serbian-English perceptual learning departs from that of early bilinguals, resulting in different assimilations of the AusE /æ/-/e/ contrast to the native Serbian vowel(s). However, early and late bilinguals did not differ in the perceptual assimilation of the /i:/-/' contrast. It is possible that the same perceptual assimilation pattern for both groups of bilinguals concealed different strategies: for example, instead of perceiving spectral differences between AusE /i:/ and /a/ as early learners do, late learners may have solely relied on the quantitative differences between /i:/ and /a/.

According to the PAM, adults’ discrimination of a non-native contrast depends on how each of the contrasting stimuli is assimilated. Discrimination accuracy on the AusE /æ/-/e/ and /i:/-/'contrast was examined in Experiment 5.

8.4 Experiment 5: Discrimination (AXB)

In this experiment, Serbian-English bilinguals’ discrimination of the AusE /æ/-/e/ and /i:/-/' contrasts was tested. According to PAM, pairs of non-native segments that are perceptually related to native categories according to TC or CG assimilation would be very easy (TC) to relatively easy (CG) to discriminate. It was expected that both groups would demonstrate excellent discrimination of the /i:/-/' contrast as both groups assimilated AusE /i:/ and /a/ to two Serbian categories /i:/ and /i/, respectively (TC
assimilation). The perceptual assimilation data for the /i:/-/u/ contrast obtained in Study 1 did not match discrimination accuracy, we hypothesize, because the participants had not been given two native categories (a short /i/ and a long /i:/) as possible responses for these AusE vowels. Therefore, it was expected that the excellent discrimination performance obtained by both groups in Study 1 would be replicated in this experiment, but more clearly accounted for by our inclusion of Serbian vowel length differences in the response choices for this experiment. As for the /æ/-/e/ contrast, according to the assimilation patterns for early and late bilinguals (TC and CG assimilation, respectively), it was expected that early learners would demonstrate excellent discrimination, and late learners relatively good discrimination. However, the results obtained in the oddity discrimination task in the perceptual discrimination experiment reported in Study 1 suggested that late learners had great difficulties discriminating between the AusE /æ/ and /e/ contrast, which is not in agreement with the CG assimilation obtained for this contrast in the perceptual assimilation experiment in Study 2 (Experiment 4). In order to reduce the high memory demand of the oddity discrimination task, AXB discrimination procedure was used in this experiment.

8.4.1 Design

The independent variables were age of learning (AOL) with three levels (early learners-EL, late learners-LL and AusE monolinguals-M), and English vowel contrast with two levels (EVC) (/æ/-/e/ (EVC1) and /i:/-/u/ (EVC2)). In addition, a separate analysis was conducted in order to test the role of short term memory and the target item (X) in discrimination of the tested contrasts. The additional independent variables were trial type with two levels of measurements (X matches the first item of the trial [AAB and BBA trials] and X matches the third item of the trial [ABB and BAA trials]). Furthermore, the second new independent variable was the target (whether X was /i/ or /u/ for the /i:/-/u/ contrast or whether X was /e/ or /æ/ for the /e/-/æ/ contrast. The dependent variable was the percentage of accurate discrimination scores.

8.4.2 Participants

In addition to the Serbian early and late learners of AusE, a group of 15 monolingual native speakers of Australian English participated in this experiment as a
control group. The language background information for the monolingual group was obtained though the screening procedure using the Language Background Questionnaire. The same Serbian-English bilinguals who participated in the perceptual assimilation task also participated in the discrimination task. The discrimination experiment was actually conducted prior to the assimilation and rating task, in order to avoid biasing the discrimination results by prior experience categorizing the stimuli to Serbian vowels.

8.4.3 Procedure

In an AXB discrimination task (Gottfried, 1984), on each trial three stimuli are presented in a sequence. The participants’ task is to decide whether the stimulus in the middle (X) in each trial is the same as the first stimulus (A) or the third stimulus (B). Stimulus presented in A and B position always belong to two different categories. The AXB discrimination procedure was used in the previous experiments that tested PAM’s predictions (Best & Strange, 1992) and in other experiments that investigated vowel perception (Polka, 1995).

In this experiment, the inter-stimulus interval (ISI) in each trial was 1000 ms (Best et al., 2001; Best & Strange, 1992). The inter-trial interval was 3000 ms. A and B were always contrasting vowels from an AusE vowel contrast. The middle item (X) was either a categorical match to A or to B, but a physically different token. That way it was assured that the participants would not rely on simple acoustic differences (Polka, 1992) when discriminating between the stimuli in a trial. A disadvantage of the AXB task with 1000 ms could be the load imposed on memory (Beddor & Gottfried, 1995). For example, since the position of the target was fixed in the middle position of stimuli in each trial, better discrimination on trials when X matched B (the third item) could be interpreted as a consequence of a recency effect due to auditory short-memory constrains (Best et al., 2001). In order to assess these short-term memory effects, discrimination performance on different trial types (AAB, ABB, BBA and BAA) was also analysed.

Each vowel contrast was represented 144 times across the four trial types. Each of the 9 tokens per vowel occurred four times in each trial type, twice in A and twice in B position (Best et al., 2001), always paired with a different opposing token. All tokens in one trial were always produced by different speakers.
The same equipment as in previous experiment was used. When participants had heard all the stimuli in one trial, the numbers “1” and “3” appeared on the screen. The listeners were required to click on the box labelled “1” if the middle item (target stimulus) was the same as the first one, or “3” if the middle item was the same as the third one.

The Serbian participants were given oral and written instructions in Serbian. Before the experiment each participant completed a practice set which contained 12 trials (3 speakers X 2 vowels x 2 repetitions) with the AusE /o:/ and /u/ vowels.

8.4.4 Results: Discrimination of the /æ/-/e/ and /i:/-/ɛ/ contrasts

Data were scored and treated in the same way as in for the Experiment 4. Percent correct discrimination for each target contrast is presented in Figure 8-3.

An ANOVA revealed a significant main effect of AOL a significant effect of EVC, and a significant interaction between AOL and EVC (see Table 8-6). Overall, the /i:/-/ɛ/ contrast was better discriminated than the /æ/-/e/ contrast.
Table 8-6.
Analysis of Variance for perceptual discrimination of the /æ/-/e/ and /i/:/u/ contrasts

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>2</td>
<td>136.87**</td>
<td>0.00</td>
</tr>
<tr>
<td>S between-group error</td>
<td>42</td>
<td>(17.83)</td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVC</td>
<td>1</td>
<td>151.33**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC x AOL</td>
<td>2</td>
<td>193.30**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC x S within group error</td>
<td>42</td>
<td>(12.75)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors.
S = subjects.
* p < 0.05. **p < 0.01.

Group comparisons (EL vs. LL, EL vs. M and LL vs. M) in simple effects tests of the EVC x AOL interaction revealed that the interaction was significant for the comparison of early versus late learners (AOL[LL, EL] x EVC, F (1,42) = 285.78, p < 0.05), as well as for the comparison of late learners versus monolinguals (AOL[LL, M] x EVC, F (1, 42)= 295.04, p < 0.05), but it was nonsignificant for early learners versus monolinguals. Both early and late learners were able to discriminate between /i:/ and /u/ with a mean discrimination accuracy of 94.63% for early learners and 88.52% for late learners. Early learners had no difficulties discriminating between /æ/ and /e/ (mean accuracy = 95.56%) whereas late bilinguals performed just above chance on this contrast (mean accuracy = 58.33%). Similarly, late learners differed from AusE monolinguals in discrimination of the /æ/-/e/ contrast but not of the /i:/-/u/ contrast. There were no significant differences between early bilinguals’ and monolingual AusE participants’ discrimination of either of the tested contrasts (see Figure 8-3).

In order to assess the role of a short term memory effects, the percent correct responses were further analysed in a three way within subjects ANOVA on the factors AOL, trial type and target type. The analysis was conducted for each contrast separately.
There was no main effect of trial type and no main effect of the target for tested contrasts. For the /æ/-/e/ three group comparisons (EL vs. LL, EL vs. M; and LL vs. M) revealed a significant difference in discrimination of this contrast between late and early learners: AOL (LL vs. EL), $F(1,42) = 441.03, p < 0.05$, and between late learners and AusE monolinguals: AOL (LL vs. M), $F(1,42) = 415, p < 0.05$. Overall, late learners’ discrimination of the tested vowel contrast was significantly poorer than that of late learners and of monolinguals regardless of trial type and target. Only one significant three way interaction was found between late learners and AusE monolinguals: AOL (LL vs. M) x memory (AAB/BBA vs. ABB/BAA) x target (/æ/ vs. /e/), $F(1,42) = 6.38, p < 0.05$. Late learners’ discrimination accuracy on AAB trials when the target was /e/ was significantly poorer when compared to AusE monolinguals’ performance.

The significant three-way interaction among the AOL, trial type and target was found when late learners were compared with early learners on the /i:/-/u/ contrast: AOL (EL vs. LL) x memory (AAB/BBA vs. ABB/BAA) x target (/i:/ vs. /u/), $F(1,42) = 4.36, p < 0.05$, and when late learners were compared with AusE monolinguals: AOL (LL vs. M) x memory (AAB/BBA vs. BBA/AAB) x target (/i:/ vs. /u/), $F(1,42) = 2.06, p < 0.05$. Late learners’ discrimination accuracy on AAB trials when the target was /i:/ significantly lower than that of early learners and monolinguals.

8.4.5 Summary of the results for Experiment 5

Early and late Serbian-English bilinguals differed in discrimination of the /æ/-/e/ contrast. Early learners displayed nativelike discrimination accuracy whereas late learners had great difficulties performing just above chance in discrimination of this contrast. Both early and late learners were able to discriminate between /i:/ and /u/ vowels highly accurately.

Effects of auditory short term memory constraints were found in late learners’ discrimination on AAB trials when the targets were /e/ and /i/.
8.5 Experiment 6: Production of AusE /æ/-/e/ and /i:-/ɪ/ contrasts

8.5.1 Design

The independent and dependent variables were the same as in the production experiment described in Study 1, Experiment 1. Independent variables were age of learning (AOL) with two levels of measurement (early learners-EL and late learners-LL) and English vowel contrast with two levels of measurement (EVC) /æ/-/e/ (EVC1) and (/i:-/ɪ/) (EVC2). Dependent variables were the percentage of Serbian-English bilinguals’ accurate production scores and the identified vowel category (IDC) (IDC1: /æ/, /e/ and IDC 2: /i:, /ɪ/) by the native monolingual AusE listeners.

8.5.2 Procedure

The procedure was identical to the procedure used in Study 1, Experiment 1. A group of 10 native AusE monolingual listeners judged early and late Serbian-English bilinguals’ and native AusE monolinguals’ production of AusE vowels /æ/, /e/, /i:/ and /ɪ/). The order of experiments was the same as in Study 1: all three groups of speakers previously participated in the discrimination and perceptual assimilation experiments before proceeding to the production experiment. The technique used to elicit the production of AusE words with targeted vowels was also the same as in Study 1, Experiment 1: the participants were instructed to repeat the target word in a h_d frame after hearing it at the beginning of randomly presented carrier sentences (e.g. “H_d is the next word to say.”). Each participant repeated each vowel five times across this experiment. Three repetitions of each vowel by each Serbian-English participant and each AusE participant were presented to a different set of AusE monolinguals for identification. This way 540 tokens were generated (45 speakers X 4 vowels X 3 repetitions). The stimuli for imitative production were presented to the Serbian-English bilinguals and AusEs using the same equipment as in Study 1.

In the second part of the experiment 10 native AusE monolinguals identified the produced tokens by choosing one of the following keywords presented on the computer screen: heed, hid, head, had, hard, hut, hot, horde, hood, heard. The random presentation of the stimuli in three counterbalanced sets (as in Study 1) ensured that one token per vowel for each speaker was presented in one set. There were 180 trials in each
set. The break between sets occurred at 20 min. The interval between the participant’s response and the presentation of the next stimulus (ITI) was 1000 ms.

8.5.3 Results

The groups’ mean percentages of correctly identified vowels by native AusE listeners are presented in Figure 8.4 (/æ/-/e/ contrast) and Figure 8.5 (/i:/-/ɪ/ contrast). Two separate ANOVAs were conducted for each contrast (see Table 8-7 for the /æ/-/e/ contrast and Table 8-8 for the (/i:/-/ɪ/ contrast).

Production of the /æ/-/e/ contrast

Figure 8-4. Production of English had, and head by early (EL) and late (LL) Serbian-English bilinguals and English monolinguals (M). On the x-axis AusE target words (bottom line) and the judgments on the Serbian-English bilinguals’ production (top line) are presented. The percentage of identified productions is presented on the y-axis.

An ANOVA revealed a significant main effect of IDC1, a significant interaction between AOL and IDC1, a significant interaction between EVC1 and IDC1, and a significant three way interaction among AOL, EVC1 and IDC1 (see Table 8-7).
Table 8-7.
Analysis of Variance for production of the /æ/-/e/ contrast

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>2</td>
<td>3.16</td>
<td>0.07</td>
</tr>
<tr>
<td>EVC1</td>
<td>1</td>
<td>0.04</td>
<td>0.85</td>
</tr>
<tr>
<td>IDC1</td>
<td>1</td>
<td>619.94**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC1</td>
<td>2</td>
<td>2.19</td>
<td>0.14</td>
</tr>
<tr>
<td>AOL x IDC1</td>
<td>2</td>
<td>780.58**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC1 x IDC1</td>
<td>2</td>
<td>5847.74**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC1 x IDC1</td>
<td>2</td>
<td>674.89**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC1 x IDC1 x S within group error</td>
<td>18</td>
<td>(27.56)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects. $*p < 0.05$. $**p < 0.01$

Group comparisons (EL vs. LL, EL vs. M and LL vs. M) in simple effects test of the AOL x EVC1 x IDC1 interaction revealed that the interaction was significant for comparison of early versus late learners (AOL [EL vs. LL] x IDC1 x EVC1, $F (1,9) = 660.22$, $p < 0.05$) and late learners versus monolinguals (AOL [LL vs. M] x IDC1 x EVC1, $F (1, 9) = 1423.13$, $p < 0.001$). Early learners produced both vowels from the tested contrast as intended. Late learners, on the other hand, produced both *head* and *had* as *head*. Early learners did not differ from monolinguals in production of the tested contrast.
Production of the /i:/ - /ɪ/ contrast

An ANOVA revealed a significant main effect of AOL, a significant main effect of IDC2, a significant interaction between AOL and IDC2, a significant interaction between EVC2 and IDC2 and a significant three-way interaction of AOL, EVC2 and (see Table 8-8).
Table 8-8.
Analysis of Variance for production of the /i:/-/æ/ contrast

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>2</td>
<td>9.81**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC2</td>
<td>1</td>
<td>4.09</td>
<td>0.07</td>
</tr>
<tr>
<td>IDC2</td>
<td>1</td>
<td>45.80**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC2</td>
<td>2</td>
<td>2.98</td>
<td>0.08</td>
</tr>
<tr>
<td>AOL x IDC2</td>
<td>2</td>
<td>38.64**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC2 x IDC2</td>
<td>2</td>
<td>38831.30*</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC2 x IDC2</td>
<td>2</td>
<td>62.22*</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x EVC2 x PC2 x S within</td>
<td>18</td>
<td>(12.34)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects. * p < 0.05. ** p < 0.01.

Groups comparisons for simple effects of interaction among AOL, EVC2, PC2 and S revealed a significant AOL x EVC2 x IDC1 interaction when late learners and monolinguals were compared (AOL [LL vs. M], F (1,9) = 19.82, p < 0.05) and when early learners and monolinguals were compared (AOL [EL vs. M], F(1,9) = 11.00, p < 0.05). Late learners did not significantly differ from early learners. Both groups of bilinguals produced the tested vowels as intended. However, the significant differences between early learners vs. monolinguals and late learners vs. monolinguals are the result of ceiling effect in the monolingual group’s performance, resulting in extremely low variance for the monolingual group.

8.5.4 Summary of the results for production experiment

The production results obtained in Study 1, Experiment 1 were replicated in this experiment. Late Serbian-English bilinguals did not produce the /æ/-/e/ contrast accurately, producing /e/ for both for both /e/ and /æ/. Early learners, on the other hand, produced both /e/ and /æ/ accurately. The groups’ accurate production of the /i:/-/æ/ contrast was also in line with the previous results.
8.6 General discussion of Study 2

In this study, PAM’s non-native speech perception framework was applied, beyond the framework of SLM addressed in Study 1, to account for AOL-related differences between early and late Serbian-English bilinguals’ perception and production of AusE /æ/-/e/ and /i:-/i/ contrasts.

The differences between early and late bilinguals in perceptual assimilation suggest that AOL has an effect on perceptual assimilation of at least some non-native vowel contrasts to bilinguals’ native categories. Early bilinguals perceptually assimilated AusE /æ/-/e/ contrast to two different Serbian vowel categories: /æ/ was perceptually assimilated to Serbian /a/, and /e/ to Serbian /e/. Late bilinguals instead perceptually assimilated the same contrast as a CG assimilation: both /æ/ and /e/ were perceived as instances Serbian /e/ differing in degree of perceived similarity to /e/.

AOL did not have an effect on perceptual assimilation of the /i:-/i/ contrast: both groups of bilinguals perceptually assimilated AusE /i:/ to Serbian /i:/ and AusE /u/ to Serbian /i/ (TC assimilation). As predicted, the perceived quantitative distinction between AusE tense-lax vowels /i:/ and /u/ was reflected in their perceived similarities to Serbian long (/i:/) and short (/i/) vowels. As discussed in Study 1, Serbian-English bilinguals perceptually assimilated both /i:/ and /u/ vowels to a Serbian short /i/ when only short Serbian vowels were given as possible responses to the AusE stimuli. In this experiment, the participants were given a set of Serbian keywords containing both short and long vowels. As predicted, under these expanded conditions, the bilinguals perceptually assimilated /i:/ and /u/ vowels to two Serbian categories, long /i:/ and short /i/, respectively. This finding is consistent with the previous findings that L2 learners may rely on duration in perception of vowel distinctions (Bohn, 1995; McAllister, Flege, & Piske, 2002).

According to PAM, perceptual difficulties in discrimination of non-native vowel contrasts are predictable on the basis of assimilation types. Excellent discrimination is expected when the non-native contrasting phones are assimilated to two different non-native phones and boundaries between the two L1 categories serves as boundaries between contrasting non-native categories. The TC assimilation types obtained for the
tested AusE contrasts in these experiments were good predictors of accuracy in
discrimination. Early learners performed well on /æ/-/e/ and /i:/-/ɨ/ contrasts. In both
cases discrimination was aided by assimilation to native phonological contrasts (Best,
1995).

The CG categorized L2 vowel types, according to the PAM model, are expected
to be discriminated well, although significantly less well than TC types. The
discrimination scores for the CG assimilation type obtained in this study, the /æ/-/e/
contrast for late Serbian bilinguals (assimilated to Serbian /e/) did not support
predictions concerning the relationship between category goodness and discrimination.
Late learners had great difficulties in discrimination of this contrast. Their poor
discrimination performance on this AusE contrast characterized as a CG type might
result from confounded CG and SC assimilation. Although the differences in goodness-
of-fit ratings between each member of the contrast and Serbian /e/ were significant
(Ause /e/ was rated as a better match to Serbian /e/) none of the AusE vowels was rated
as an ideal match to Serbian vowel /e/. It is possible that this significant difference in
goodness-of-fit ratings concealed an actual SC assimilation type, i.e., that both AusE
vowels were poor exemplars of the closest Serbian vowel /e/. In the perceptual
assimilation and rating task participants’ responses were given for non-native segments
presented one at the time and the listeners were required to pay thorough attention to the
within-category differences. Thus, AusE /æ/ tokens were rated as more deviant from
Serbian /e/ than AusE /e/ tokens in this type of task. In the discrimination task, on the
other hand, the listeners’ responses were given for multiple stimuli presented
consecutively in one trial. The participants were required to decide whether these
stimuli belonged to the same vowel category or not, but the task did not include the
judgement on the degree of AusE vowels discrepancy from the closest Serbian match.
One way to overcome this methodological gap would be to propose a perceptual
assimilation task and goodness-of-fit rating task for multiple presented tokens to match
the stimuli presentation conditions in discrimination task.

An alternative explanation is related to the criterion for establishing when an L2
stimulus is to be considered perceptually assimilated to an L1 category. In this study,
the AusE stimulus had to be categorized as an exemplar of the chosen Serbian vowel at
least 50% of the time across a given listener group. This criterion allows for SC
assimilation to be confounded with CG assimilation (Lotto, 2000; Lotto, Kluender, & Holt, 1998). The use of a more stringent criterion (Harnsberger, 2001) would insure that category goodness as a predictor of discriminability would not be confounded with the discrimination of a SC type.

Early and late Serbian-English bilinguals differed in discrimination of the /æ/-/e/ contrast. Early learners displayed nativelike discrimination accuracy whereas late learners had great difficulties performing just above the chance in discrimination of this contrast. Both early and late learners were able to discriminate between /i:/ and /u/ vowels accurately.

The effects of auditory memory were found in late learners’ discrimination on AAB trials (when the target X matched the first item) for targets /e/ and /i/. Better discrimination of these vowels in X position where X matched the third item of the trial (recency-type trials) suggested that auditory memory influenced discrimination of /e/ and /i/, but not discrimination of /æ/ and /u/.

Finally, the production experiment results were in agreement with those obtained in Study 1 (Experiment 1). The contrast that caused perceptual difficulties for late learners (the /æ/-/e/ contrast) was not produced accurately by this group. They produced both vowels as /e/ according to the AusE judges.

Taken together, these results suggest that PAM’s assimilation types change over the course of L2 development. When an L2 is learned early in life (e.g. early childhood) the perceptual influence of the L1 phonetic system is less prominent than in late L2 acquisition (e.g. late adolescence and adulthood). However, these L1 effects are not equal for all L2 contrasts. It could be speculated that the perceptual influence of Serbian is less prominent in perception of high front /i:/ and /u/ vowels, since both groups of bilinguals had no perceptual difficulties with this contrast.
CHAPTER 9

PHONOLOGICAL AND PHONETIC PRIMING EFFECTS ON PERCEPTION OF ENGLISH /æ/-/e/ AND /iː/-/ɜ:/ BY SERBIAN-ENGLISH BILINGUALS

The results of the experiments reported in chapter 7 and chapter 8 suggested that age of learning affects Serbian-English bilinguals’ perception of the AusE /æ/-/e/ contrast but has no effect on perception of the /iː/-/ɜ:/ contrast. The questions addressed in the experiments reported in this chapter are concerned with the changes in discrimination accuracy of the /æ/-/e/ contrast when phonemic and phonetic processing are activated by phonological and phonetic priming and manipulation of interstimulus intervals (ISI). Whilst various studies have tried to explain the dynamics of the activation and deactivation of phonological segments in lexical units during the perception of spoken words, the priming paradigm has not yet been used to investigate the processes involved in cross-language segmental discrimination and categorization. Phonological (and/or phonetic) priming methodology has been extensively used to investigate processes involved in spoken word recognition and the underlying (sublexical) mechanisms that regulate mapping of sensory information from the acoustic input to the representations stored in listeners’ native-language mental lexicon.

Previous research investigating the influence of neighbourhood density and phonotactic probability suggests that both lexical representations (whole word forms) and sublexical representations (parts of words, such as phonological segments and sequences of segments, morphological and orthographical features) are involved in both spoken (Luce & Pisoni, 1998; Vitevitch, 2003; Vitevitch & Luce, 1998) and visual word recognition (Frost, Ahissar, Gotesman, & Tayeb, 2003). For example, it has been shown that words with sparse neighbourhoods (i.e. words that have few words similar to them) were responded to more quickly and accurately than words with dense neighbourhoods (words that have many words that resemble them) (Vitevitch & Luce, 1998). Similarly, nonwords with common segments and sequences of segments
(nonwords with high phonotactic probability, e.g. initial /s/ in English) were rated more like words in English than nonwords with less common segments and sequences of segments (nonwords with low phonotactic probability, e.g. initial /j/ in English). As words with dense neighbourhoods tend to be comprised of segments and sequences of segments with high phonotactic probability and words with sparse neighbourhoods tend to be comprised of segments with low phonotactic probability, it is possible that in the absence of lexical representations listeners relied on sublexical (phonotactic) information in the processing of spoken words. We posit here that sublexical information may be used when second language learners acquire new words: initially, a novel sequence of segments with no associated semantic information should be retained through sublexical representations. Subsequently, a lexical representation for this sequence of segments can be established.

In this study the priming paradigm utilised in the investigation of the spoken word recognition was adapted to examine the sublexical (phonological) level in speech processing. In particular, the perceptual assimilation and discrimination of two AusE vowel contrasts (/æ/-/e/ and /i/-/i/) by early and late Serbian learners of Australian English were investigated using task-appropriate adaptations of the phonological priming method.

The logic of the priming paradigm is that the processing of a stimulus (the prime) should affect the processing of a subsequent stimulus (the target) when the prime and the target share some formal information that exists in the phonological representations involved in lexical access. In phonological priming the phonological structure of the prime can facilitate or inhibit the recognition of the target. The magnitude and direction of this effect depends on the number of phonemes that the prime and the target have in common, and/or the number of sub-phonemic features they have in common, as well as the time gap between the prime and the target and different strategies the participants may use depending on the task.

9.1 Priming paradigm

The priming paradigm, originally used in the field of spoken word recognition and investigations of semantic memory (Meyer & Schvaneveldt, 1971) involves presentation of two stimuli in a sequence: a prime (to which no response is required)
and a target (to which a response is required). If the prime and the target share some formal information (e.g. semantic relationship between the prime and target as in the prime-target pair *bread-butter*) the processing of the target is expected to be affected by the automatic or obligatory processing of the prime, even if the instruction is to ignore the prime. The participants in the experiments that use priming are required to say the target aloud (Sereno, 1991), to decide if the target is a word or a nonword (Radeau, Morais, & Dewier, 1989; Segui & Grainger, 1990), or to identify a briefly presented target (Slowiaczek, Nusbaum, & Pisoni, 1987). Reaction time (the time from the onset of the target until the participant’s response), error rate, number of correct responses, or event-related electrocortical potentials are used as indicators of the prime’s effect on processing of the target. For example, in experiments that examined the reaction time relationship between the prime and the target, a faster reaction time when the prime and the target are related is interpreted as facilitation of the target by the prime. A slower reaction time is interpreted as inhibition. The independent variables that are typically manipulated are the nature and/or degree of relatedness between the prime and the target.

In an early experiment by Meyer and Schvaneveldt (1971) it was found that when the prime and the target were semantically related the subjects were faster to classify the target as a word than in cases when this semantic relationship was absent. Since that study, semantic priming has been consistently used in research on semantic memory and lexical decision (Foss, 1982) word identification (Jackson & Morton, 1984) and naming tasks (Jackson & Morton, 1984; Keefe & Neely, 1990).

9.1.1 Priming and the role of phonology in auditory word recognition

The early findings from semantic priming experiments were extended to examine facilitation effects of phonemically and orthographically related primes and targets. In phonological priming experiments, the prime and the target (presented auditorily or visually) share one or more phonemes. Similarly as in semantic priming, the task involves a prime followed by the target to which the participant’s response is required. A facilitation effect is found when the targets and the primes are related phonotactically (Pisoni, Nusbaum, Luce, & Slowiaczek, 1985), phonologically (Meyer & Schriefers, 1991; Slowiaczek et al., 1987), prosodically (Grosjean & Gee, 1987), or orthographically (Jakimik, Cole, & Rudnicky, 1985). The prime and the target may
share from one or more phonemes, usually at word onset (Goldinger, Luce, Pisoni, & Marcario, 1992; Radeau et al., 1989; Slowiaczek & Hamburger, 1992) or all phonemes (identity or repetition priming) (Hamburger & Slowiaczek, 1998). The facilitation (or inhibition) effect of phonologically related primes is compared to participants’ responses when the prime and the target are phonologically unrelated (baseline or control condition).

9.1.2 Tasks in phonological priming experiments

Facilitation due to phonological priming has been observed in perceptual identification tasks (Goldinger et al., 1992; Slowiaczek et al., 1987; Slowiaczek & Pisoni, 1986) when a target embedded in noise or briefly presented was preceded by clearly presented primes. The subjects were required to identify the targets. For example, Slowiaczek et al. (1987) found that with increasing number of phonemes shared by the prime and the target identification of words presented in noise improved. The largest priming effect was observed in the condition when the prime and the target were identical (i.e. repetition priming). Similarly, in lexical decision tasks (Burton, Jongman, & Sereno, 1993) (Goldinger, 1998; Norris, McQueen, & Cutler, 2002; Praamstra & Stegeman, 1993; Slowiaczek, McQueen, Soltano, & Lynch, 2000; Spinelli, Segui, & Radeau, 2001) with an increase of overlapping phonemes the participants made increasingly faster and more accurate decisions whether the target was a word or a nonword. The facilitation effect of phonologically related primes was also elicited in picture naming tasks (Meyer & Schriefers, 1991; Wheldon, 2003).

9.1.3 Amount and location of phonological overlap between the prime and the target

Distinct effects have been observed depending on the location of the overlapping segments and on the extent to which primes and targets overlap. For primes that overlap with the final phonemes of targets (rhyme primes, e.g. mean-bean) when both primes and targets are auditorily presented facilitation has been consistently reported (Burton et al., 1993; Dumay et al., 2001; Norris et al., 2002; Slowiaczek et al., 2000; Spinelli et al., 2001). The effects are more complex when primes overlap with the initial phonemes of the targets (onset primes). For example, inhibition was observed when the prime and the target shared all but the last phoneme (Dufour & Peereman, 2003a, 2003b; Hamburger & Slowiaczek, 1996, 1999; Slowiaczek & Hamburger, 1992). On the other
hand, facilitatory effects have been reported when the prime and the target overlapped in smaller number of phonemes (Radeau, 1995; Spinelli et al., 2001). Several studies suggest that the initial overlap facilitation results from strategic processes (expectancy) that listeners develop when they become aware of the phonological relationship between the primes and the targets (Goldinger et al., 1992; Hamburger & Slowiaczek, 1996, 1999). On the other hand, some studies suggest that the effect could result from automatic processes occurring in normal spoken word recognition (Spinelli et al., 2001).

The position of the prime and the target vowel in experiments reported in this chapter is medial, specifically medial (nuclear) vowel priming in CVC targets. No prior studies have looked at that particular position of overlap: all prior studies focused on onset primes (typically C_ or CV_) and rhyme primes (_V or _CV) or repetition priming.

9.1.4 Interstimulus interval (ISI) and Stimulus onset asynchrony (SOA)

Interstimulus interval (ISI) is the time between the offset of the prime and the onset of the target. Various ISIs have been used in the phonological priming experiments: 20 ms and 500 ms (Radeau, Morais, & Segui, 1995), 50 ms (Slowiaczek & Hamburger, 1992), 1000 ms (Goldinger, 1998), and 1500 ms (Goldinger et al., 1992). Another measure of the time distance between the prime and target is the stimulus onset asynchrony (SOA). SOA refers to the time from the onset of a prime to the beginning of a target (including the duration of the prime as well) (Lupker & Colombo, 1994; Pexman, Cristi, & Lupker, 1999). SOA is commonly used in semantic priming experiments when the availability of semantic information extracted from the prime is of great importance for processing of the target. ISI, on the other hand, is commonly used in experiments where phonemic and phonetic information are relevant for processing of the following stimuli.

Goldinger et al. (1992) compared phonemic and phonetic (feature) priming at short and long ISIs (50 ms, 500 ms and 1500 ms), while manipulating the proportions of different prime-target relationships (phonemically related, phonetically related and unrelated primes and targets). Phonemic priming is observed when perception of a spoken target (e.g. nurse) is modified by prior perception of a phonologically related prime (e.g. north). Nurse and north are phonologically related because they share the initial phoneme. Phonetic priming is observed when a prime and a target are
phonetically similar, but share no common phonemes. For example, robe and like share phonetically similar initial segments only, where all phonetic features are the same except that of place of articulation (/r/ vs. /l/), thus they are not the same phonemes. The results obtained suggested that phonemic and phonetic priming have different bases: the manipulation in proportion of relationships between primes and targets influenced phonemic priming, but not phonetic priming. With a higher number of phonemically related prime-target trials the participants reacted faster, relative to the unrelated prime-target trials. This effect was the same across 50 ms, 500 ms and 1500 ms ISIs. Phonetic priming, on the other hand, was observed only when the ISI was 50 ms. These results suggested that both ISI and the strategies that participants develop, based on the proportion of the related and unrelated trials, play a role in lexical decision (Hamburger & Slowiaczek, 1996).

In the experiments reported in this chapter, 1000 ms and 500 ms prime-target ISIs were used. The motivation for the use of these two ISIs was based on the perception results obtained in Studies 1 and 2 (chapters 6 and 7, respectively) and on the evidence for that listeners display greater sensitivity to non-native speech contrasts in a 500 ms ISI condition for non-primed discrimination (Werker & Logan, 1985).

9.1.5 The role of ISI in cross-language speech perception

The interstimulus interval (ISI) (Crowder, 1982; Pisoni, 1973) plays a determining role in discrimination of non-native speech categories. At the shorter ISIs, subjects show evidence of within-category auditory-level discriminations whereas at longer ISIs only phonetic categorization is evident. In addition to these effects due to the length of the ISI, research has shown that in experimental procedures that have high memory demands, such as the ABX task, subjects have access only to a phonemic code, whereas in procedures with low memory demands, access to the acoustic details is facilitated, allowing low-level auditory processing (Carney, Widin, & Viemeister, 1977; Pisoni & Lazarus, 1974). Stimulus characteristics also influence whether auditory or phonetic processing will be used. Both Pisoni (1973) and Studdert-Kennedy (1973) have speculated that it may be more difficult to demonstrate an auditory level of processing for consonant than for vowel stimuli, because the relevant acoustic cues differentiating consonants are so brief and transient, whereas those cues differentiating vowels are longer in duration and include steady-state parameters. The brief, transient
cues for consonants may not remain available in auditory short-term memory, especially when they are presented in the context of a longer, steady state vowel. Such speculation is supported by research showing better within-category discrimination of temporally extended consonant stimuli (Tartter, 1981) and more categorical perception of shortened vowels. (Fujisaki & Kawashima, 1970; Pisoni, 1973).

Werker and Logan (1985) argued that there might be three factors in speech perception, depending on the ISIs: phonemic, phonetic and psychoacoustic (or auditory) factors. When the ISI is 1500 ms subjects perceive non-native stimuli according to their native-language phonological categories demonstrating phonemic perception. With 500 ms ISIs subjects show sensitivity to phonetic distinctions that are used in some other (but not their native) languages, using phonetically relevant (or phonetic) perception. When the ISI is 250ms, a generalized psychoacoustic or auditory level of processing is demonstrated in which subjects show sensitivity to acoustic differences that do not correspond to phonetic boundaries that function phonologically (to contrast meaning) in any of the world’s languages. In such 250 ms ISI conditions listeners discriminate speech segments on the basis of any acoustic variability between individual exemplars.

The present experiments were conducted with two ISIs, 1000 ms and 500 ms. In Studies 1 and 2 late Serbian-English bilinguals had difficulties discriminating the AusE /e/-/æ/ contrast when the ISI was 1000 ms. Therefore, the same 1000 ms ISI was used in the following experiments. It was hypothesized, however, that with the 500 ms ISI late learners might benefit from phonetic processing and display greater sensitivity to the /e/-/æ/ contrast they had difficulties with in 1000 ms condition.

9.1.6 Predictions

In this study Serbian-English bilinguals’ (early and late learners) phonological and phonetic processing of AusE /æ/-/e/ and /i:/-/i/ vowel contrasts was investigated using task-appropriate adaptations of the phonological priming method. The main hypotheses derive from the results of the experiments reported in Study 1 (chapter 7) and Study 2 (chapter 8).

Two studies were conducted. A priming paradigm with 1000 ms ISI (Experiment 1) and 500 ms (Experiment 2) was used to further test Serbian-English bilinguals’ sensitivity to phonemic (phonological) and phonetic levels of processing. The relationship between the prime and the target was also manipulated: the prime and
target contained phonologically identical, phonetically similar or phonetically unrelated vowels.

According to the previous studies reported in this thesis, AOL has an effect on perceptual assimilation and discrimination of the AusE /æ/-/e/ contrast, but not on the /i:/-/ʊ/ contrast. Early bilinguals perceptually assimilated AusE /æ/-/e/ to two different Serbian vowel categories: /a/ and /e/, respectively. Late learners perceptually assimilated both of these AusE vowels to Serbian /e/. This effect of AOL was not observed for AusE /i:/-/ʊ/ which was relatively easy for both AOL groups to discriminate. The perceptual assimilation patterns were relatively good predictors of perceptual performance: early Serbian-English bilinguals had no difficulties discriminating between /æ/ and /e/, whereas late Serbian-English bilinguals performed poorly in discrimination of the same contrast. Somewhat surprising, initially, was the learners’ poor discrimination of the /æ/-/e/ vowels, considering that they were not perceived as an equally good match to Serbian /e/ and therefore were expected to be discriminated relatively well. It is possible that late learners ignored phonetic differences between /æ/ and /e/ in the discrimination task, but were able to perceive them in the identification task when one stimulus was presented at a time and when the task requirements allowed direct comparison to the chosen Serbian category. If that was the case, manipulating ISI to allow a shift from phonemic (1000 ms ISI) to phonetic (500 ms ISI) effects on perception, and using phonological primes (same vowel is repeated in the prime and the target) versus phonetic primes (opposing member of the vowel contrast – a single phonetic feature difference -- is used as the prime) should affect late learners’ and early learners discrimination of the /æ/-/e/ vowels in different ways.

The 500 ms ISI would facilitate late learners’ discrimination of the “difficult” /æ/-/e/ contrast in the phonologically primed target vowel. This was predicted because the repetition of the same vowel in the first item of the discrimination targets was expected to increase late bilinguals’ sensitivity to the phonetic difference between the two targets (acting in a similar way as a repetition prime in lexical decision tasks) and hence to improve their discrimination performance. Phonetic priming in the 500 ms
condition was expected to inhibit discrimination of that same contrast, however. This was expected because this short ISI allows listeners to perceive phonetic distinctions between non-native segments that are not used distinctively in their native language (Werker & Logan, 1985). Due to increased sensitivity to the non-native phonetic difference between prime and target, the prime in the shorter ISI condition was expected to interfere with late learners’ perception of the second target item, resulting in difficulties in discrimination. Alternatively, however, if the participants were not able to perceive this phonetic difference between the prime and the first target, their performance in both phonological and phonetic prime conditions would be the same.

Early learners were expected to differ from late learners in discrimination of the /æ/-/e/ contrast: in both 1000 ms and 500 ms ISI conditions early learners should be able to access phonetically relevant information in discrimination of this contrast. This hypothesis was based on the results of Study 1 and Study 2 in which early learners’ performance did not differ from that of monolingual AusE speakers in the control group.

However, when making such predictions we must take into account the previous research on the phonological systems in “early” bilinguals (Pallier, Bosch, & Sebastian-Galles, 1997; Sebastian-Galles & Soto-Faraco, 1999) which indicated that even highly skilled early L2 speakers might fail to perform in the same way as monolingual speakers of their L2. In that case, if the tasks used in this study are sensitive enough to tap into the fine difference in segmental processing between early learners and AusE monolinguals, early learners’ perception of the targeted contrasts should depart from AusE monolinguals in the discrimination task.

In the 1000 ms ISI condition it was hypothesized that the prime would not have an effect on late learners’ discrimination of the /æ/-/e/ contrast, as phonemic processing in 1000 ms does not allow perception of those distinctions that are not contrastive in Serbian. Early learners were expected to display the same discrimination accuracy as in previous experiments.

No difference between early and late learners were expected for the /iː/-/ɪ/ contrast regarding the prime type.
9.1.7 Participants

Fifteen early and fifteen late Serbian-English bilinguals participated in two discrimination and two perceptual assimilation tasks. None of these bilinguals participated in the previous experiments. The requirements for participation (age of learning (AOL), experience with English prior to arrival in Australia for late learners, age of learning English as second language for early learners, the native Serbian dialect) were the same as in the previous experiment. The participants filled the same Language Background Questionnaire and participants information statement and consent form in Serbian. Furthermore, all communication between the participants and the experimenter took place in Serbian.

The mean age of early learners was 25.2 years (SD = 1), and that of late learners was 37.6 years (SD = 5.2. Mean AOL for the early learners was 3.7 years (SD = 1); that for the late learners was 30.26 years (SD = 4.7). The mean length of English use for the early learners was 21.25 years (SD = 5.6); that for the late learners was 7.6 years (SD = 2.4).

Fifteen monolingual AusE speakers participated as a control group in the Study 3 discrimination experiments (mean age = 22.37 years, SD = 2.7).

9.1.8 Stimuli

The same stimuli used in the previous experiment were used as the target words. In addition, the prime stimuli, the AusE vowels /æ/, /e/, /i:/, /u/ and /o/ , were produced in a b_p frame (/bæp/, /bep/, /bip/, /bi:p/, /bup/, /bop/) in the carrier sentence ‘________ is the next word to say. I say ______ again. ______’ by ten male monolingual speakers of AusE; thus the primes were produced by different speakers than the target stimuli. These speakers also produced tokens of the AusE vowel /o:/. In b_p context for the unrelated syllables used in the practice trials before the experimental tasks.

The AusE speakers were undergraduate students of Psychology and Languages & Linguistics at the University of Western Sydney. Prior to the recoding session, the speakers were screened for their language background via the Language Background Questionnaire. Only those who reported speaking no other language but AusE and having no exposure to other languages on regular daily bases were invited for recording.
The sentences with the stimuli were repeated 10 times for each vowel. The recorded speech material was digitised at 44 kHz using Cool Edit. Three tokens in citation form (final isolated token at end of the carrier sentences) for each vowel were chosen from three different speakers and used as the priming stimuli in the experiment. The vowels in the chosen tokens were matched according to their duration, F0, F1, F2 and F3 at the vowel midpoint. This way 56 prime stimuli were chosen (6 vowels x 3 speakers x 3 tokens). The amplitude of each word was normalized to 90% of the peak intensity of all words. The stimuli were recorded and presented to participants sing the equipment as in previous Study 1 and Study 2.

9.2 Experiments 7 and 8

All Serbian-English bilinguals completed two perceptual experiments: primed perceptual discrimination (AX) and primed perceptual assimilation and goodness-of-fit rating. The experiments were conducted as two separate test sessions. There were two primed AX discrimination tests (one with 500 ms ISI and one with 1000 ms ISI) and two primed perceptual assimilation tests (also with 500 and 1000 ms ISI). In Experiment 7 participants completed both of the primed AX discrimination tests. In Experiment 8 they completed both of the primed perceptual assimilation tests with goodness-of-fit ratings. In each experiment the stimuli were blocked by ISI condition, but the trials (determined by the relationship between the prime and target) within the condition were randomised. At each test session, the order of 500 and 1000 ms ISI conditions was alternated between the participants for counterbalancing, so that one participant started the testing session with the 500 ms ISI, and the next participant started the session with the 1000 ms ISI condition. AusE monolingual participants completed only the discrimination tests in a single session, with testing order for the 500 and 1000 ms ISI conditions counterbalanced between the participants.

9.2.1 Procedures

Experiment 7: Primed AX discrimination

In each trial of a standard AX (same/different) discrimination test participants are presented either with two tokens that contain two contrastive non-native categories,
or two repetitions of the same physical token. The listener’s task is to assess whether the two tokens presented are the same or different. The present AX discrimination task differed in two ways: a prime (one of three types) was presented before the A target, and the AX pair for “same” cases did not involve the same physical token but only the same phonological category, i.e., it was a primed AX task. The details of these modifications to the standard AX procedure are described next.

In the current primed-AX experiment, participants were presented with a vowel embedded in a b_p context to which no response was required (a prime) and two vowels in h_d context to which a response was required (AX targets). The participants were required to decide if these two vowels in the h_d context were from the same vowel category or not. The vowel in the prime and in the target stimulus A were the same category in the phonological prime condition, opposite members of the English vowel contrast in the phonetic prime condition and phonetically unrelated in the unrelated condition. The relationship between the primes and the targets in the AX task is presented in the table 9.1. The prime, the A and the X target items were always produced by different speakers (i.e., three different speakers). There were 324 trials in total for each ISI condition: 4 vowels x 3 speakers x 3 tokens x 3 repetitions x 3 priming conditions.
Table 9-1
The stimulus triads for trials in the phonological, phonetic, and unrelated prime-target conditions

<table>
<thead>
<tr>
<th>Trial type</th>
<th>Phonological prime</th>
<th>Phonetic prime</th>
<th>Unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different trials</td>
<td>ßæp/hæd /hed</td>
<td>ßæp/ hed/ hed</td>
<td>bop/ hed/ hæd</td>
</tr>
<tr>
<td></td>
<td>bep/hed/ hæd</td>
<td>bep/hæd/hed</td>
<td>bop/hæd/hed</td>
</tr>
<tr>
<td></td>
<td>bïp /hïd/hi:d</td>
<td>bïp /hi:d/ hi:d</td>
<td>bup/ hi:d /hïd</td>
</tr>
<tr>
<td>Same trials</td>
<td>ßæp/hæd/ hæd</td>
<td>ßæp/hed/hed</td>
<td>bop/hed /hed</td>
</tr>
<tr>
<td></td>
<td>bep/hed/ hed</td>
<td>bi:p/hïd/ hïd</td>
<td>bup/ hïd/ hïd</td>
</tr>
<tr>
<td></td>
<td>bi:p /hi:d /hïd</td>
<td>bep/ hæd/ hæd</td>
<td>bop/hæd/ hæd</td>
</tr>
<tr>
<td></td>
<td>bi:p/hi:d/hi:d</td>
<td>bïp /hi:d /hi:d</td>
<td>bup/hïd/ hïd</td>
</tr>
</tbody>
</table>

When participants had been presented all the stimuli in a given trial, two boxes appeared on the computer screen. One box was labelled “same” and the other box was labelled “different.” The listeners were instructed to click on the box “same” if the target A and X h_d tokens were instances of the same vowel category or click on the “different” box if the h_d targets were categorically different (containing two different vowel categories). The software used for stimulus presentations and response recording was DMDX.

The bilingual participants were given oral and written instructions in Serbian; the controls were given all instructions in English. In both oral and written instructions the participants were told to ignore the first word and to give their responses only with regard to the relationship between the second and third words. Before the experiment each participant completed the practice in the experimenter’s presence. The practice consisted of 4 trials (Table 9.2), in which all tokens each trial were always produced by different speakers.
Table 9-2
The tokens used in the practice for the primed AX task

<table>
<thead>
<tr>
<th>Trial type</th>
<th>Prime / target A / target X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same trials</td>
<td>bo:p / hud / hud</td>
</tr>
<tr>
<td></td>
<td>bo:p / ho:d / ho:d</td>
</tr>
<tr>
<td>Different trials</td>
<td>bo:p / ho:d / hud</td>
</tr>
<tr>
<td></td>
<td>bo:p / hud / ho:d</td>
</tr>
</tbody>
</table>

The experiment took place in the same room as the experiments in Study 1 and Study 2.

Experiment 8: Primed perceptual assimilation and goodness-of-fit ratings

In the second visit the participants completed two primed perceptual assimilation tests with goodness-of-fit ratings. In each test, consistent with the discrimination experiment, the stimuli were blocked by ISI condition, but the trials (determined by the relationship between the prime and target) within each condition were randomised. The prime and the target contained the same vowel category in the phonological prime condition, the opposite member of the vowel contrast in the phonetic prime condition and unrelated vowels in the unrelated condition (Table 9-3).

Table 9-3
The prime and the target relationships in the primed perceptual assimilation experiments.

<table>
<thead>
<tr>
<th>Phonological prime/target</th>
<th>Phonetic prime/target</th>
<th>Unrelated prime/target</th>
</tr>
</thead>
<tbody>
<tr>
<td>bæp/ hæd</td>
<td>bæp/ hed</td>
<td>bop/ hæd</td>
</tr>
<tr>
<td>bep/ hed</td>
<td>bep/ hæd</td>
<td>bop/ hed</td>
</tr>
<tr>
<td>bi:p/ hid</td>
<td>bi:p hid</td>
<td>bup/ hid</td>
</tr>
<tr>
<td>bip/ hid</td>
<td>bip/ hi:d</td>
<td>bup/ hi:d</td>
</tr>
</tbody>
</table>
The order of 500 and 1000 ms ISI tests was alternated between the participants for counterbalancing, so that one participant started the testing session with the 500 ms ISI, and the next participant started the session with the 1000 ms ISI condition. Each trial had two parts: in the first part the participants heard one of the four AusE vowels in the h_d frame following the prime by the specified ISI period, and were asked to choose a Serbian keyword given on the screen that contained the same vowel as the AusE target word. They were instructed to ignore the first word and to give their responses to the second word. The possible responses were the same keywords as in Study 2 (baba, beba, biba, bobo, buba, baaba, beeba, biiba, booba, buuba) containing all Serbian vowels, five short (/a/, /e/, /i/, /o/, /u/) and five long (/a:/, e:, /i:/, /o:/, /u:/), embedded in the b_ba frame (Table 9.1) The Serbian keywords were presented on the computer screen. After hearing the AusE prime and target, participants were to click on the box containing the Serbian keyword with the matching vowel. The responses were given using the computer mouse. Once the response was given, after 1000ms participants heard the same prime and target word again. This time the boxes with numbers from 1 to 7 appeared on the bottom of the screen and the participants were asked to rate the similarity between the English vowel in the target word and their selected Serbian vowel using a 7-point scale (“1”-very bad example of the Serbian vowel, “7”-exactly the same vowels). The next trial started 1000 ms after the rating response was given. There were 216 randomly presented trials in total for each ISI condition (4 vowels x 3 speakers x 3 tokens x 2 repetitions x 3 priming conditions).

The participants were given oral and written instructions in Serbian (see Appendices 9-4 and 9-5), after which they completed 4 practice trials with the experimenter present. The practice trials consisted of the bo:p / hud and bo:p/ ho:d prime/target pairs repeated two times. The number of practice trials was limited to only two AusE vowels repeated 2 times, with the aim to familiarize participants with the task and the position of the Serbian keywords on the computer screen.

Before receiving the oral instructions, participants read the written instructions in Serbian. Next, the experimenter (a native speaker of Serbian) explained the differences between long and short vowels, giving an example of Serbian words in the appropriate context. These examples were the same as the ones used to illustrate this difference in the previous experiment (Na klupi sedi jedan deda/ Na klupi sedi mnogo seda; An old man is sitting on the bench/ The old man are sitting on the bench). As in
the previous perceptual assimilation and goodness-of-fit study (Study 2), when the participants were reminded of the difference between long and short vowels in Serbian, it was explained that the orthographic representation for each vowel would be doubled to indicate the long vowels. Finally, they were shown the list of the keywords that were going to be used in the task and asked to read them aloud. The familiarization with the keywords was followed with an explanation of how to use the 1-7 rating scale (“1”-very bad example of the Serbian vowel, “7”- exactly the same as the Serbian vowel). Participants were instructed to ignore the prime and to give their responses to the target only. After the practice trials, the participant began the experiment.

9.2.2 Design

The independent variables in discrimination experiments were age of learning (AOL) with three levels (EL, LL, and M) English vowel contrast with two levels (EVC) (/æ/-/e/ (EVC1) and /i:/-/ɪ/ (EVC2) and prime (P) with two levels (phonological prime (PhonolP) and phonetic prime (PhonetP) The unrelated prime (UP) was used a baseline. The dependent variable in the discrimination task was the percentage of accurate discrimination scores. In the perceptual assimilation test, the additional independent variables were Serbian vowel contrast (SVC) with two levels (SVC1: /a/ vs. /e/ and SVC2 /i/ vs. /i:/). As in Study 1 and Study 2, AOL was between-subject factor, EVC, P and SVC were within subject factors. The dependent variable was a percentage of identification as each Serbian vowel.

9.3 Results: Experiment 1 - Primed perceptual assimilation and goodness-of-fit rating

9.3.1 1000 ms ISI condition

Separate ANOVAs were conducted for each AusE vowel contrast (EVC): EVC1 /æ/-/e/ and EVC2 /i:/-/ɪ/ within each priming condition. The mean percentages of chosen Serbian keywords given in response to the English words for each group in the unrelated priming condition (baseline) was subtracted from the phonological and the phonetic priming conditions in order to assess the effect of the phonological and phonetic primes compared to the baseline, unprimed perceptual assimilation of two vowel contrasts. The mean percentages of chosen Serbian keywords for each condition
and the mean percentage of chosen Serbian keywords for phonological and phonetic prime condition when subtracted from unrelated prime condition are presented in tables 9-4 (/æ/-/e/ contrast) and 9-5 (/i/-/ɪ/ contrast).

The /æ/-/e/ contrast

Table 9-4. Perceptual Assimilation at 1000 ms ISI EVC1

<table>
<thead>
<tr>
<th>Mean percent of identified choices</th>
<th>/had/</th>
<th>/head/</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOL /babia/ /beba/ /babia/ /beba/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological prime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>60.74</td>
<td>12.59</td>
</tr>
<tr>
<td>LL</td>
<td>21.48</td>
<td>57.04</td>
</tr>
<tr>
<td>Phonetic prime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>31.85</td>
<td>38.52</td>
</tr>
<tr>
<td>LL</td>
<td>34.07</td>
<td>54.07</td>
</tr>
<tr>
<td>Unrelated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>40.74</td>
<td>38.52</td>
</tr>
<tr>
<td>LL</td>
<td>29.63</td>
<td>57.04</td>
</tr>
<tr>
<td>Phonological-Unrelated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>20.00</td>
<td>-25.93</td>
</tr>
<tr>
<td>LL</td>
<td>-8.15</td>
<td>-2.96</td>
</tr>
<tr>
<td>Phonetic-Unrelated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>-8.89</td>
<td>-0.00</td>
</tr>
<tr>
<td>LL</td>
<td>4.44</td>
<td>-2.96</td>
</tr>
</tbody>
</table>

Note: Positive values indicate an increase in performance (i.e. facilitation) in the given experimental condition relative to the unrelated baseline (because the values for the baseline condition have been subtracted from the experimental condition). Conversely, negative values indicate a decrease in performance relative to baseline i.e. response inhibition.

An ANOVA revealed a significant four-way interaction between AOL, P, EVC1 and SVC1 (Table 9-5).
Table 9-5.
Analysis of Variance for perceptual assimilation of the /æ/-/e/ contrast (1000 ms ISI).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>1</td>
<td>0.18</td>
<td>0.68</td>
</tr>
<tr>
<td>S between-group error</td>
<td>28</td>
<td>(61.71)</td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>0.01</td>
<td>0.94</td>
</tr>
<tr>
<td>EVC1</td>
<td>1</td>
<td>2.29</td>
<td>0.14</td>
</tr>
<tr>
<td>SVC1</td>
<td>1</td>
<td>0.14</td>
<td>0.72</td>
</tr>
<tr>
<td>AOL x P</td>
<td>1</td>
<td>1.09</td>
<td>0.31</td>
</tr>
<tr>
<td>AOL x EVC1</td>
<td>1</td>
<td>0.09</td>
<td>0.76</td>
</tr>
<tr>
<td>AOL x SVC1</td>
<td>1</td>
<td>1.46</td>
<td>0.24</td>
</tr>
<tr>
<td>PRIME x EVC1</td>
<td>1</td>
<td>0.90</td>
<td>0.35</td>
</tr>
<tr>
<td>PRIME x SVC1</td>
<td>1</td>
<td>1.89</td>
<td>0.18</td>
</tr>
<tr>
<td>EVC1 x SVC1</td>
<td>1</td>
<td>1.01</td>
<td>0.32</td>
</tr>
<tr>
<td>AOL x P x EVC1 x SVC1</td>
<td>1</td>
<td>8.79**</td>
<td>0.01</td>
</tr>
<tr>
<td>PRIME x EVC1 x SVC1 x S within group error</td>
<td>28</td>
<td>(484.42)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects.
* p < 0.05. ** p < 0.01.

Group comparisons (EL vs. LL) in each priming condition revealed a significant interaction between AOL (EL vs. LL) and SVC1 (/a/ vs. /e/) in phonological prime condition: F (1, 28) = 7.71, p < 0.05. The assimilation pattern for /æ/ and /e/ for both groups of bilinguals mirrored the pattern obtained in Study 2: early learners perceptually assimilated /æ/ to Serbian /a/, and /e/ to Serbian /e/. Late learners perceptually assimilated both /æ/ and /e/ to Serbian /e/. The significant interaction between AOL and SVC1 only means that absolute percentage of identification choice did change depending on AOL: for EL /baba/ responses increased relative to the
baseline condition (/a/ response facilitation) while /beba/ decreased relative to the baseline condition (/e/ inhibition), whereas for LL the opposite pattern was observed (/e/ facilitation and /a/ inhibition).

The /iː/-/ɪ/ contrast

The mean percentages of chosen Serbian keywords for each condition and the mean percentage of chosen Serbian keywords for phonological and phonetic prime condition when unrelated condition was subtracted from the phonological and phonetic subtracted from unrelated prime condition for the /iː/-/ɪ/ contrast are presented in Table 9-6.

Table 9-6
Perceptual Assimilation at 1000 ms ISI
EVC2

<table>
<thead>
<tr>
<th></th>
<th>Mean percent of identified choices</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/heed/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/biba/</td>
<td>/biiba/</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phonological prime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>6.25</td>
<td>81.25</td>
<td>95.14</td>
</tr>
<tr>
<td>LL</td>
<td>5.56</td>
<td>88.89</td>
<td>96.03</td>
</tr>
<tr>
<td></td>
<td>Phonetic prime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>6.94</td>
<td>86.11</td>
<td>91.67</td>
</tr>
<tr>
<td>LL</td>
<td>7.14</td>
<td>89.68</td>
<td>93.65</td>
</tr>
<tr>
<td></td>
<td>Unrelated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>5.56</td>
<td>88.19</td>
<td>93.06</td>
</tr>
<tr>
<td>LL</td>
<td>4.76</td>
<td>88.10</td>
<td>92.06</td>
</tr>
<tr>
<td></td>
<td>Phonological-Unrelated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>0.69</td>
<td>-6.94</td>
<td>2.08</td>
</tr>
<tr>
<td>LL</td>
<td>0.79</td>
<td>0.79</td>
<td>3.97</td>
</tr>
<tr>
<td></td>
<td>Phonetic-Unrelated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>1.39</td>
<td>-2.08</td>
<td>-1.39</td>
</tr>
<tr>
<td>LL</td>
<td>2.38</td>
<td>1.59</td>
<td>1.59</td>
</tr>
</tbody>
</table>

Note: Positive values indicate an increase in performance (i.e. facilitation) in the given experimental condition relative to the unrelated baseline (because the values for the baseline condition have been subtracted from the experimental condition). Conversely, negative values indicate a decrease in performance relative to baseline i.e. response inhibition.
An ANOVA revealed a significant interaction between P (phonological vs. phonetic) and EVC2 (/i:/-/u/) (Table 9-7).

Table 9-7.
Analysis of Variance for perceptual assimilation of the /i:/-/u/ contrast 1000 ms

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>1</td>
<td>0.40</td>
<td>0.53</td>
</tr>
<tr>
<td>S between-group error</td>
<td>28</td>
<td>(44.84)</td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>2.75</td>
<td>0.11</td>
</tr>
<tr>
<td>EVC2</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SVC2</td>
<td>1</td>
<td>2.15</td>
<td>0.15</td>
</tr>
<tr>
<td>AOL x P</td>
<td>1</td>
<td>0.71</td>
<td>0.41</td>
</tr>
<tr>
<td>AOL x EVC2</td>
<td>1</td>
<td>0.41</td>
<td>0.53</td>
</tr>
<tr>
<td>AOL x SVC2</td>
<td>1</td>
<td>0.00</td>
<td>0.98</td>
</tr>
<tr>
<td>Px EVC2</td>
<td>1</td>
<td>10.17**</td>
<td>0.00</td>
</tr>
<tr>
<td>P x SVC2</td>
<td>1</td>
<td>1.84</td>
<td>0.53</td>
</tr>
<tr>
<td>EVC2 x SVC2</td>
<td>1</td>
<td>0.00</td>
<td>0.97</td>
</tr>
<tr>
<td>AOL xP x EVC: SVC2</td>
<td>1</td>
<td>0.16</td>
<td>0.70</td>
</tr>
<tr>
<td>Px EVC2 x SVC2 x S</td>
<td>28</td>
<td>(42.92)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects.

*p < 0.05, **p < 0.01
Separate ANOVAs for each priming condition revealed only the main effect of SVC2 (/i/ vs. /i:/, \( F(1,28) = 4.93 \)) in the phonological priming condition. Overall, the /biba/ responses for both groups slightly increased relative to the baseline facilitation of this response while /biiba/ percentage of /biiba/ responses decreased (response inhibition). However, the assimilation pattern did not change as a consequence of the prime: both groups of bilinguals perceptually assimilated AusE /i:/ to Serbian /i:/ and AusE /i/ to Serbian /i/.

**Goodness-of-fit ratings**

The goodness-of-fit rating scores for the unrelated prime condition subtracted from the phonological and phonetic prime conditions in the 1000 msec ISI are presented in Table 9-8 (for the /æ/-/e/ contrast) and in Table 9-9 (for the /i/-/i/ contrast). Early and late learners’ rating scores for the /æ/-/e/ contrast were analysed separately owing to the fact that early learners perceptually assimilated two members of the AusE contrast to two separate Serbian categories, whereas late learners perceptually assimilated the two AusE vowels to a single Serbian category. Hence two sets of ANOVAs were conducted. As both groups of participants perceptually assimilated members of the /i/-/i/ contrast equally (to Serbian /i:/ and /i/) the rating scores for EVC2 were subjected to a single ANOVA in order to investigate a possible interaction between AOL and prime types for EVC2.

**The /æ/-/e/ contrast**

There were no significant effects for early learners. A significant main effect of prime for late learners (\( F(1,14) = 4.74, p < 0.05 \)) in the phonetic prime condition revealed an overall decrease in rating scores for both the /æ/ and /e/ vowels relative to the baseline. It appears that the prime condition influenced early and late learners differently. However, the change of less than a half-point in the goodness-of-fit rating scale may not be practically relevant. Therefore, it can be concluded that priming did not effect the goodness ratings to any substantial amount to distinguish perceived closeness of the AusE vowels and the chosen Serbian vowels in the EVC1 (/æ/-/e/) condition (Table 9-8).
Table 9-8.
Goodness-of-fit rating difference scores in 1000 ms ISI

<table>
<thead>
<tr>
<th></th>
<th>Late learners</th>
<th>Early learners</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>had (beba) 0.14</td>
<td>head (beba) 0.61</td>
</tr>
<tr>
<td></td>
<td>head (beba) -0.53</td>
<td>had (baba) -0.31</td>
</tr>
<tr>
<td>Phonological prime</td>
<td></td>
<td>head (beba) 0.31</td>
</tr>
<tr>
<td>Phonetic prime</td>
<td>-0.31</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>0.39</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Note: Positive values indicate an increase in ratings in the given experimental condition relative to the unrelated baseline condition (i.e., facilitation of assimilation); positive values indicate a decrease in performance in the experimental condition relative to the unrelated baseline (inhibition).

The /i:/-/ʔ/ contrast

The significant interaction between prime and EVC2 (F (1,28) = 8.34, p < 0.05) revealed that both early and late learners gave /t/ a lower rating in the phonological prime condition than in the phonetic prime condition. The ratings for /i:/ were equivalent in both prime conditions. There were no interactions between AOL and the other factors for this contrast: both early and late learners performed similarly.

Goodness-of-fit rating difference scores for the /i:/-/ʔ/ contrast are presented in Table 9-9.

Table 9-9.
Goodness-of-fit rating difference scores in 1000 ms ISI

Priming effect on rating scores

<table>
<thead>
<tr>
<th>AOL</th>
<th>heed (biiba)</th>
<th>hid (biba)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phonological prime</td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>-0.07</td>
<td>-0.11</td>
</tr>
<tr>
<td>LL</td>
<td>-0.90</td>
<td>-0.54</td>
</tr>
<tr>
<td></td>
<td>Phonetic prime</td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>0.35</td>
<td>0.82</td>
</tr>
<tr>
<td>LL</td>
<td>0.36</td>
<td>0.92</td>
</tr>
</tbody>
</table>
9.3.2 500 ms ISI condition

The mean percentages of chosen Serbian keywords for each condition are presented in tables 9-10 (the /æ/-/e/ contrast) and 9-12 (the /i:/-/ι/ contrast). Separate ANOVAs were conducted for each AusE vowel contrast (EVC): EVC1 /æ/-/e/ and EVC2 /i:/-/ι/ within each priming condition (see Table 9-11 and Table 9-13 for ANOVA results).

The /æ/-/e/ contrast

Table 9-10.
Perceptual Assimilation at 500 ms ISI EVC

<table>
<thead>
<tr>
<th></th>
<th>Mean percent of identified choices</th>
<th>/had/</th>
<th>/head/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/baba/</td>
<td>/beba/</td>
<td>/baba/</td>
</tr>
<tr>
<td>AOL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>86.67</td>
<td>1.48</td>
<td>1.48</td>
</tr>
<tr>
<td>LL</td>
<td>23.70</td>
<td>74.07</td>
<td>5.93</td>
</tr>
<tr>
<td>Phonological prime</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>54.07</td>
<td>4.44</td>
<td>2.96</td>
</tr>
<tr>
<td>LL</td>
<td>17.78</td>
<td>75.56</td>
<td>2.96</td>
</tr>
<tr>
<td>Phonetic prime</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>91.11</td>
<td>2.22</td>
<td>1.48</td>
</tr>
<tr>
<td>LL</td>
<td>17.04</td>
<td>80.74</td>
<td>2.22</td>
</tr>
<tr>
<td>Unrelated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>-4.44</td>
<td>-0.74</td>
<td>-0.00</td>
</tr>
<tr>
<td>LL</td>
<td>6.67</td>
<td>-6.67</td>
<td>3.70</td>
</tr>
<tr>
<td>Phonological-Unrelated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>-37.04</td>
<td>2.22</td>
<td>1.48</td>
</tr>
<tr>
<td>LL</td>
<td>0.74</td>
<td>-5.19</td>
<td>0.74</td>
</tr>
</tbody>
</table>

An ANOVA revealed a significant main effect of AOL, significant main effect of EVC1, a significant interaction between AOL and P, a significant interaction between AOL and EVC1, a significant interactions between AOL and SVC1, P and EVC1, and a significant interaction among AOL, P, EVC1 and SVC1 (see Table 9-10).
Table 9-11
Analysis of Variance for perceptual assimilation of the /æ/-/e/ contrast in 500 ms ISI condition

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>1</td>
<td>24.40**</td>
<td>0.00</td>
</tr>
<tr>
<td>S between-group error</td>
<td>28</td>
<td>(8.86)</td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>14.34</td>
<td>0.08</td>
</tr>
<tr>
<td>EVC1</td>
<td>1</td>
<td>19.35**</td>
<td>0.00</td>
</tr>
<tr>
<td>SVC1</td>
<td>1</td>
<td>0.07</td>
<td>0.80</td>
</tr>
<tr>
<td>AOL x P</td>
<td>1</td>
<td>8.06**</td>
<td>0.01</td>
</tr>
<tr>
<td>AOL x EVC1</td>
<td>1</td>
<td>9.58**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x SVC1</td>
<td>1</td>
<td>4.40*</td>
<td>0.05</td>
</tr>
<tr>
<td>P x EVC1</td>
<td>1</td>
<td>16.23**</td>
<td>0.00</td>
</tr>
<tr>
<td>P x SVC1</td>
<td>1</td>
<td>7.02**</td>
<td>0.01</td>
</tr>
<tr>
<td>EVC1 x SVC1</td>
<td>1</td>
<td>4.58*</td>
<td>0.04</td>
</tr>
<tr>
<td>AOL x P x EVC1 x SVC1</td>
<td>1</td>
<td>7.81**</td>
<td>0.01</td>
</tr>
<tr>
<td>P x EVC1 x SVC1 x S</td>
<td>28</td>
<td>(178.18)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects.
* p < 0.05. ** p < 0.01.

In the phonetic prime condition there was a significant main effect of AOL: \( F(1, 28) = 20.4, \ p < 0.05 \) and a significant main effect of EVC1 (/æ/ vs. /e/), \( F(1, 28) = 28.68, \ p < 0.05 \). A significant two-way interaction between AOL and EVC1: \( F(1, 28) = 14.63, \ p < 0.05 \) and a significant two-way interaction between AOL and SVC1: \( F(1, 28) = 6.28, \ p < 0.05 \) could be attributed to a 37% decrease in early learners’ /baba/ responses for the /æ/ vowel relative to the baseline.

A significant two-way interaction between SVC1 and EVC1 (\( F(1, 28) = 13.61, \ p < 0.05 \)) and a significant three-way interaction EVC1 x SVC1 x AOL (\( F(1, 28) = 21.42, \ p < 0.05 \)) again are most likely attributed to early learners’ change in perceptual assimilation of /æ/. This change is also confirmed in a high percentage of /baaba/ responses for had for early learners: 39.02% of time they perceptually assimilated /æ/
to /aː/. No significant effects or interactions were obtained for the phonological prime condition.

*The /iː/-/ɪ/ contrast*

Perceptual assimilation for the /iː/-/ɪ/ contrast in each priming condition and subtracted from unrelated condition are presented in Table 9-12.

*Table 9-12.*
Perceptual Assimilation at 500 ms ISI EVC2

<table>
<thead>
<tr>
<th>Mean percent of identified choices</th>
<th>/heed/</th>
<th>/hid/</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOL</td>
<td>/biba/</td>
<td>/biiba/</td>
</tr>
<tr>
<td>Phonological prime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>5.19</td>
<td>83.70</td>
</tr>
<tr>
<td>LL</td>
<td>1.48</td>
<td>96.30</td>
</tr>
<tr>
<td>Phonetic prime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>3.70</td>
<td>88.89</td>
</tr>
<tr>
<td>LL</td>
<td>2.96</td>
<td>90.37</td>
</tr>
<tr>
<td>Unrelated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>2.96</td>
<td>85.93</td>
</tr>
<tr>
<td>LL</td>
<td>0.74</td>
<td>96.30</td>
</tr>
<tr>
<td>Phonological-Unrelated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>2.22</td>
<td>-2.22</td>
</tr>
<tr>
<td>LL</td>
<td>0.74</td>
<td>0.00</td>
</tr>
<tr>
<td>Phonetic-Unrelated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>0.74</td>
<td>2.96</td>
</tr>
<tr>
<td>LL</td>
<td>2.22</td>
<td>-5.93</td>
</tr>
</tbody>
</table>

An ANOVA revealed no significant results for the perceptual assimilation of the /iː/-/ɪ/ contrast with the in 500 ms ISI for either the phonological or the phonetic prime conditions (see Table 9-13).
Table 9-13
Analysis of Variance for perceptual assimilation of the /i/-/i/ contrast in 500 ms ISI condition

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Between subjects</td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>1</td>
<td>0.05</td>
<td>0.68</td>
</tr>
<tr>
<td>S between-group error</td>
<td>28</td>
<td>(61.71)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within subjects</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>0.01</td>
<td>0.94</td>
</tr>
<tr>
<td>EVC2</td>
<td>1</td>
<td>0.00</td>
<td>0.97</td>
</tr>
<tr>
<td>SVC2</td>
<td>1</td>
<td>0.63</td>
<td>0.43</td>
</tr>
<tr>
<td>AOL x P</td>
<td>1</td>
<td>0.52</td>
<td>0.48</td>
</tr>
<tr>
<td>AOL x EVC2</td>
<td>1</td>
<td>0.29</td>
<td>0.59</td>
</tr>
<tr>
<td>AOL x SVC2</td>
<td>1</td>
<td>0.91</td>
<td>0.35</td>
</tr>
<tr>
<td>P x EVC2</td>
<td>1</td>
<td>0.01</td>
<td>0.93</td>
</tr>
<tr>
<td>P x SVC2</td>
<td>1</td>
<td>0.41</td>
<td>0.53</td>
</tr>
<tr>
<td>EVC2 x SVC2</td>
<td>1</td>
<td>2.97</td>
<td>0.10</td>
</tr>
<tr>
<td>AOL x P x EVC2 x SVC2</td>
<td>1</td>
<td>0.00</td>
<td>0.97</td>
</tr>
<tr>
<td>P x EVC2 x SVC2 x S within group error</td>
<td>28</td>
<td>(484.42)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects.

* *p < 0.05. ** *p < 0.01.

**Goodness-of-fit rating**

The /æ/-/e/ contrast

There was a significant main effect of prime on this contrast for early learners: F (1, 14) = 10.07, p < 0.05. Rating scores in the phonetic priming condition were reduced (inhibition of assimilation) compared to the unrelated baseline condition.

For late learners there was a significant main effect of EVC1: F(1,14) = 4.61m p < 0.05. The vowel /æ/ was given lower rating scores (inhibition of assimilation to Serbian /e/) with both the phonological and phonetic prime conditions, relative to the
unrelated condition, whereas there was no significant change in the rating scores for the vowel /e/. The goodness-of-fit rating scores in 500 ms ISI condition are presented in Table 9-14.

Table 9-14.
Goodness-of-fit rating difference scores in 500 ms ISI  EVC1

<table>
<thead>
<tr>
<th></th>
<th>LL</th>
<th>EL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>had (beba)</td>
<td>head (beba)</td>
</tr>
<tr>
<td>Phonological</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prime</td>
<td>-0.16</td>
<td>0.11</td>
</tr>
<tr>
<td>Phonetic prime</td>
<td>0.01</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

The /i:/ - /ɪ/ contrast

The only significant result obtained here was a main effect of prime: F (1, 28) = 12.48, p < 0.05. In phonological prime condition both /i:/ and /ɪ/ vowels were given higher rating scores (facilitation of assimilation) whereas in the phonetic prime condition both AusE vowels were given lower rating scores (inhibition of assimilation). The goodness-of-fit rating difference scored for phonological and phonetic prime condition are presented in Table 9-15.
Table 9-15.
Goodness-of-fit rating difference scores in 500 ms ISI EVC2

<table>
<thead>
<tr>
<th></th>
<th>AOL</th>
<th>hid (biba)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonological prime</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>-0.07</td>
<td>-0.11</td>
</tr>
<tr>
<td>LL</td>
<td>-0.90</td>
<td>-0.54</td>
</tr>
<tr>
<td><strong>Phonetic prime</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EL</td>
<td>0.89</td>
<td>-0.18</td>
</tr>
<tr>
<td>LL</td>
<td>-0.24</td>
<td>0.13</td>
</tr>
</tbody>
</table>

9.4 Results: Experiment 2-Primed discrimination of /æ/-/e/ and /iː/-/ɪ/

First the groups’ d prime scores (d’ scores) were calculated for discrimination of each vowel contrast in each priming condition. These scores were derived from the proportion of “hits” (correct response when two stimuli represented different vowel categories) and “false alarms” (incorrect response to a different item when both stimuli represented the same vowel category) obtained for each contrast. D’ scores were calculated using the same formula as in Study 1 (d’ = z(H) – z(F)) to calculate the difference between standardised scores of “hits” and “false alarms” (see Experiment 2, Study 1 for detailed explanation of d’ scores calculation). Then the d’ scores for the phonological and phonetic prime conditions were subtracted from the unrelated (baseline) condition. The d’ scores for each contrast subtracted from the baseline condition were analysed using ANOVAs (see Table 8-17 for ANOVA results for the 1000 ms ISI condition and 8-19 for ANOVA results for the 500 ms ISI condition).

9.4.1 1000 ms ISI condition

The d’ scores for the baseline condition were subtracted from each prime condition and the d’ scores obtained by subtraction from each priming condition are presented in Table 9-16.
Table 9-16
Discrimination for the 1000 ms ISI

<table>
<thead>
<tr>
<th></th>
<th>(had/head)</th>
<th>(hid/heed)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonological prime</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>2.48</td>
<td>2.57</td>
</tr>
<tr>
<td>EL</td>
<td>2.10</td>
<td>2.31</td>
</tr>
<tr>
<td>LL</td>
<td>1.75</td>
<td>2.22</td>
</tr>
<tr>
<td><strong>Phonetic prime</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>2.53</td>
<td>2.51</td>
</tr>
<tr>
<td>EL</td>
<td>2.19</td>
<td>2.22</td>
</tr>
<tr>
<td>LL</td>
<td>1.08</td>
<td>2.41</td>
</tr>
<tr>
<td><strong>Unrelated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>2.53</td>
<td>2.47</td>
</tr>
<tr>
<td>EL</td>
<td>2.13</td>
<td>2.35</td>
</tr>
<tr>
<td>LL</td>
<td>1.81</td>
<td>2.27</td>
</tr>
<tr>
<td><strong>Phonological-Unrelated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>-0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>EL</td>
<td>-0.03</td>
<td>-0.05</td>
</tr>
<tr>
<td>LL</td>
<td>-0.07</td>
<td>-0.05</td>
</tr>
<tr>
<td><strong>Phonetic-Unrelated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>-0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>EL</td>
<td>0.06</td>
<td>-0.14</td>
</tr>
<tr>
<td>LL</td>
<td>0.73</td>
<td>0.14</td>
</tr>
</tbody>
</table>

The ANOVA revealed a significant interaction between AOL and EVC. Overall, early learners discriminated both /i:/-/u/ and /æ/-/e/ contrasts accurately, whereas late learners had difficulties discriminating the /æ/-/e/ contrast. There was also a significant three way interaction among AOL, P and EVC (see Table 9-17).
Table 9-17

Analysis of Variance for perceptual discrimination in the 1000 ms ISI condition

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>2</td>
<td>1.21</td>
<td>0.31</td>
</tr>
<tr>
<td>S between-group error</td>
<td>42</td>
<td>(0.13)</td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>2.32</td>
<td>0.14</td>
</tr>
<tr>
<td>EVC</td>
<td>1</td>
<td>3.09</td>
<td>0.09</td>
</tr>
<tr>
<td>AOL x P</td>
<td>2</td>
<td>2.19</td>
<td>0.12</td>
</tr>
<tr>
<td>AOL x EVC</td>
<td>2</td>
<td>3.76*</td>
<td>0.03</td>
</tr>
<tr>
<td>PIME x EVC</td>
<td>1</td>
<td>3.75</td>
<td>0.06</td>
</tr>
<tr>
<td>AOL x P x EVC</td>
<td>2</td>
<td>11.55**</td>
<td>0.00</td>
</tr>
<tr>
<td>P x EVC x S within group error</td>
<td>42</td>
<td>(0.11)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects. * p < 0.05. ** p < 0.01.

Group comparisons (EL vs. LL, EL vs. M and LL vs. M) in simple effects test of the AOL x P x EVC interaction revealed a significant interaction among AOL, P and EVC when early and late learners were compared (AOL [EL vs. LL] x P x EVC: $F(1,42) = 18.69, p < 0.05$) and when late learners and monolinguals were compared (AOL [LL vs. M] x P x EVC: $F(1,42) = 15.34, p < 0.05$). In the phonetic prime condition late learners’ discrimination accuracy of the /æ/-/e/ contrast decreased significantly whereas early learners and monolingual speakers in the control group showed no evidence of priming.
9.4.2 500 ms ISI condition

The d’ discrimination scores for the /æ/-/e/ and /i:/-/u/ contrasts for each prime condition and the d’ scores for the phonological and phonetic prime obtained by subtraction of the unrelated condition from the prime condition are presented in Table 9-18.

Table 9-18
Discrimination 500 ms ISI

<table>
<thead>
<tr>
<th></th>
<th>AOL (had/head)</th>
<th>AOL (hid/heed)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonological prime</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>2.68</td>
<td>2.72</td>
</tr>
<tr>
<td>EL</td>
<td>2.10</td>
<td>2.00</td>
</tr>
<tr>
<td>LL</td>
<td>1.75</td>
<td>1.62</td>
</tr>
<tr>
<td><strong>Phonetic prime</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>2.59</td>
<td>2.52</td>
</tr>
<tr>
<td>EL</td>
<td>1.44</td>
<td>1.90</td>
</tr>
<tr>
<td>LL</td>
<td>1.19</td>
<td>1.84</td>
</tr>
<tr>
<td><strong>Unrelated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>2.68</td>
<td>2.61</td>
</tr>
<tr>
<td>EL</td>
<td>2.02</td>
<td>2.09</td>
</tr>
<tr>
<td>LL</td>
<td>2.50</td>
<td>2.38</td>
</tr>
<tr>
<td><strong>Phonological-Unrelated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>EL</td>
<td>0.08</td>
<td>-0.09</td>
</tr>
<tr>
<td>LL</td>
<td>-0.75</td>
<td>-0.76</td>
</tr>
<tr>
<td><strong>Phonetic-Unrelated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>-0.09</td>
<td>-0.09</td>
</tr>
<tr>
<td>EL</td>
<td>-0.58</td>
<td>-0.19</td>
</tr>
<tr>
<td>LL</td>
<td>-1.31</td>
<td>-0.54</td>
</tr>
</tbody>
</table>
The ANOVA revealed a significant main effect of AOL, a significant main effect of P, a significant interaction between P and EVC and a significant three-way interaction among AOL, P and EVC (see Table 8-19). A significant main effect of the prime condition indicated an overall decline in discrimination accuracy (inhibition) in the phonetic priming condition for all three groups, relative to the unrelated baseline condition.

A significant interaction of the prime and EVC suggested that both early and late learners’ discrimination of the /æ/-/e/ contrast worsened (was inhibited) in phonetic prime condition.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>2</td>
<td>24.98**</td>
<td>0.00</td>
</tr>
<tr>
<td>S between-group error</td>
<td>42</td>
<td>(0.11)</td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>16.24**</td>
<td>0.00</td>
</tr>
<tr>
<td>EVC</td>
<td>1</td>
<td>3.21</td>
<td>0.08</td>
</tr>
<tr>
<td>AOL x P</td>
<td>2</td>
<td>1.60</td>
<td>0.21</td>
</tr>
<tr>
<td>AOL x EVC</td>
<td>2</td>
<td>1.01</td>
<td>0.37</td>
</tr>
<tr>
<td>P x EVC</td>
<td>1</td>
<td>13.30**</td>
<td>0.00</td>
</tr>
<tr>
<td>AOL x P x EVC</td>
<td>2</td>
<td>5.68**</td>
<td>0.01</td>
</tr>
<tr>
<td>P x EVC x S within group error</td>
<td>42</td>
<td>(0.56)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Value enclosed in parenthesis represent mean square errors. S = subjects. 
*p < 0.05. **p < 0.01.

Group comparisons (EL vs. LL, EL vs. M and LL vs. M) in simple effects tests of the AOL x P x EVC interaction (see Table 9-19) revealed that the interaction was significant for the comparison of early learners and monolinguals (AOL [EL vs. M] x P x EVC, $F(1,42) = 5.88$, $p < 0.05$ and for late learners and monolinguals learners (AOL
[LL vs. M] x P x EVC, $F(1, 42) = 10.05$, $p < 0.05$. These results suggested a decrease (inhibition) in discrimination performance relative to baseline and relative to the control group on the /æ/-/e/ contrast for both early and late learners in the phonetic priming condition.

9.5 Summary of the results and discussion of Study 3

In this study Serbian-English bilinguals’ (early and late learners) phonological and phonetic processing of AusE /æ/-/e/ and /ɪː/-/iː/ vowel contrasts was investigated using task-appropriate adaptations of the phonological priming method. The phonologically identical, phonetically similar and unrelated prime and the target relationships were manipulated using 1000 ms and 500 ms interstimulus intervals.

Priming under the 1000 ms ISI condition did not have any effect on assimilation patterns for the tested contrasts. Both groups of bilinguals perceptually assimilated /æ/-/e/ and /ɪː/-/iː/ contrasts in the same way as in Study 1 and Study 2. Early learners perceptually assimilated /æ/ to /a/ and /e/ to /e/ whereas late learners perceptually assimilated both /æ/-/e/ to Serbian /e/. Both groups perceptually assimilated members of the /ɪː/-/iː/ contrast to two Serbian categories, /iː/ and /i/, respectively. These assimilation patterns were good predictors of discrimination accuracy: early learners discriminated both /ɪː/-/iː/ and /æ/-/e/ contrasts accurately, whereas late learners had difficulties discriminating the /æ/-/e/ contrast. Furthermore, in the phonetic prime condition late learners’ discrimination accuracy of the /æ/-/e/ contrast decreased significantly below the unrelated condition, whereas early learners and monolingual speakers in the control group did not change their performance relative to the baseline.

It is not clear what causes this decrease in late learners’ discrimination accuracy on the /æ/-/e/ contrast in this condition. It seems that 1000 ms allows late learners to perceive a phonetic distinction between /æ/ and /e/ that inhibits their discrimination of the contrast. This response is not surprising: although the difference compared to the unrelated condition is significant, the inaccuracy still follows the same pattern for late learners. They still have difficulties discriminating the otherwise problematic contrast.
Goodness-of-fit ratings suggested that for both groups, /e/ and /i/ were better instances of the chosen Serbian categories (/e/ and /i/, respectively) than /æ/ and /i:/ were to the chosen Serbian categories (/a/ or /e/ and /i:/, respectively). However, overall rating scores were lower in the phonetic prime condition for both tested contrasts. Furthermore, /æ/ was given a lower rating scores in both the phonological and phonetic prime conditions.

In the 500 ms ISI condition the perceptual assimilation pattern of the /æ/-/e/ contrast for early learners was changed: early learners had a tendency to perceptually assimilate both /æ/ and /e/ to Serbian /e/, just as late learners did across ISI conditions. This assimilation pattern was reflected in discrimination accuracy: early learners’ discrimination accuracy in the phonetic priming condition decreased significantly, indicating the same difficulties in discrimination of the /æ/-/e/ contrast as late learners had in both ISI conditions. Both groups perceptually assimilated members of the /i:/-/i/ contrast to two different Serbian categories, /i:/ and /i/ and had no difficulties discriminating between them.

Contrary to expectations, phonetic processing at 500 ms ISI did not facilitate late learners’ discrimination of the /æ/-/e/ contrast when a phonological prime was used nor did it inhibit discrimination when a phonetic prime was used. Surprisingly, early learners had difficulties in discrimination of the /æ/-/e/ contrast when phonetic processing at the shorter ms ISI allowed them to access phonetic distinctions between the prime and the target. This discrimination pattern as well as their perceptual assimilation pattern seem to match the late learners’ performance on the same contrast in other conditions. However, the fact that this decrease in early learners’ discrimination of the /æ/-/e/ contrast happens in both ISI and priming conditions that caused no changes in late learners’ performance indicates different processing strategies in early learners and late learners. It is possible, given that the prime always preceded the AX target sequence, that early learners were confused due to the high memory demands: they remembered the phonetic differences between the prime and the first target (A), but were unable to focus on the differences between the two targets. If that is the case, one could also argue that late learners faced the same memory demand problem in the 1000 ms ISI condition, although it is still unclear why they were able to perceive the
phonetic distinctions between the prime and the first target at such a long ISI. Another possibility may be the methodology itself: the different prime and target consonant contexts (b_p vs. h_d) may add to the perceptual confusion of the less well established vowel contrasts: in order to attend to the task, the participants might have been paying too much attention to the additional segmental information in the prime and the targets which, again, was memory demanding and prevented them from focusing on the discrimination of the /æ/-/e/ contrasts.

Perceptual assimilation of the /æ/-/e/ contrasts under the 500 ms condition ISI revealed that in phonetic prime condition early learners less frequently identified /æ/ as an instance of Serbian /a/ and very often (39.02% of the time) identified /æ/ as an instance of Serbian /a:/ . Again, a possible explanation for this response could be attributed to the methodology. It is possible that once early learners decided that /æ/ could not be identified as /a/ the next closest match in Serbian vowel inventory was /a:/ . Since the participants did not hear auditory representations of Serbian keywords, they had to rely on their own memory representations of the Serbian keywords (and vowels they contained). Early learners’ years of exposure to English might have caused the decrease in sensitivity to the vowel length in Serbian resulting in somewhat deviant memories of the distinctions between short and long vowel in Serbian. That is, early learners might have underestimated the difference in distinctions between the short and long Serbian vowels. If that was the case, the temporal distinction between Serbian /a/ and /a:/ would be underestimated and /æ/ would be related to /a:/ . This pattern of response also indicates that /æ/ is the less stable member of the /æ/-/e/ contrast. The reason for this asymmetry may be found at the lexical, rather than segmental level. For example, the beginning portion of English panda and pencil should be indistinguishable for Dutch speakers, as /æ/ and /e/ are phonetically indistinguishable in Dutch. In an eyetracking study (Weber & Cutler, 2004) Dutch listeners were presented with spoken English words beginning with pan or pen, and had to click on items in a display containing, for example, a panda and a pencil. The results revealed that when Dutch participants were instructed to click on panda, they were likely to look at the pencil during the presentation of the first syllable of the target word. However, when instructed to click on pencil, the Dutch participants did not look at the panda. Weber and Fox
(2004) interpreted these findings as indicating that one category, in this case /e/ was
ominant, causing Dutch listeners to look at the items containing /e/ first. For early
Serbian-English bilinguals, it appears that /e/ may have been dominant and therefore
more stable in the perceptual assimilation task across conditions. Overall early learners’
perceptual performance for the 500 ms ISI was consistent with previous results
(Sebastian-Galles & Soto-Faraco, 1999) suggesting that in certain tasks even highly
skilled early bilinguals fail to perceive L2 segments like native speakers of their L2.

The rating scores across the conditions and participants do not show a particular
strong pattern. Overall, the magnitude of change across conditions is too small to
indicate the difference from the patterns obtained in previous studies (that is, the
priming did not affect the perceived similarity between AusE vowels and the chosen
Serbian categories). However, in the phonological and phonetic priming conditions
under 500 ms ISI the late learners’ rating scores for /æ/ were lower than in the baseline
condition. It seems that phonetic processing allows late learners to make better
comparisons of the less stable category from the /æ/-/e/ contrast (/æ/) and their chosen
Serbian counterpart (/a/). There was no change in the rating scores for /e/ relative to the
unrelated baseline performance level.
CHAPTER 10
GENERAL DISCUSSION

This chapter presents the findings of this thesis in light of interpretations based on existing models and findings on age effects in perception and production of L2 segments. Specifically, the perception and production of the Australian English /e/ - /æ/ and /i/ - /ɪ/ vowel contrasts by early and late Serbian-English bilinguals was investigated. The aim of these studies was to investigate age effects on L2 vowel production and perception by testing predictions generated from two models, the Speech Learning Model (SLM, Flege, 1995) and the Perceptual Assimilation Model (PAM, Best, 1995) and to further test early and late learners’ perceptual performance using modified phonological and phonetic priming tasks. The results are discussed with respect to both theoretical and practical issues: L2 learning in early childhood vs. L2 learning in adulthood, stability of L2 speech segments in early learners, and implications for future research and L2 teaching and learning. Before addressing the interpretation of the results, however, we will turn to a brief review of the issues and an overall summary of the current findings.

Language experience constrains perception of non-native speech contrasts that differ phonologically or phonetically from those in the bilinguals’ native language. These effects are most obvious in adult second language learners, but recent findings at other laboratories, and in this thesis, are beginning to provide evidence that residual L1 effects persist even in learners who acquired their second language in early childhood. These effects are present in both perception and production of non-native segments. For example, studies examining the production of English by individuals who have learned English as an L2 have shown a wide range of phonetic divergences from the speech of monolinguals in both early (childhood) and late (adult learners) learners. Age of learning effects have been observed in the production of individual vowels (Munro, Flege, & MacKay, 1995), indicating that late learners’ production of vowels is more likely to differ from monolinguals’ production than early learners. Furthermore, second language speakers often differ from native speakers of the second language in segmental perception as well (MacKay, Meador, & Flege, 2001).
These age effects in second language segmental production and perception appear to be the result of an interaction between bilinguals’ first and second languages. Completely native-like segmental production and perception in a second language would imply that bilinguals can circumvent their L1 and L2 sub-systems’ influence on one another. The evidence, both from this thesis and from other recent reports, suggests that they cannot.

10.1 Summary of the findings

10.1.1 Study 1: Implications for the Speech Learning Model

L2 vowel production accuracy and vowel perception were tested in three experiments in Study 1 (Production experiment, Perceptual Assimilation Experiment and Perceptual Discrimination Experiment). The questions addressed in the study were: a) whether early and late Serbian-English bilinguals would differ in production and perception of the AusE /e/-/æ/ and /i/-/i/ contrasts, b) whether one of these two contrasts would be more “difficult” than the other, and c) whether production inaccuracy reflected a tendency on the part of the Serbian-English bilinguals to perceptually relate the targeted AusE vowels to their native vowels. The predictions for this study were based largely on the Speech Learning Model (SLM), according to which the likelihood of category formation for L2 phones is affected by both age of L2 learning (AOL) and the perceived phonetic distance of an L2 phone from the closest L1 phone. It was hypothesized that early and late Serbian-English bilinguals would differ in production and perception of AusE /e/-/æ/ and /i/-/i/ vowel contrasts. Late learners were expected to perceive the vowels from the /e/-/æ/ contrast as dissimilar to any of their native vowel categories and therefore to produce them more accurately. The vowels from the /i/-/i/ contrast were expected to be perceived as close to Serbian /i/ and therefore difficult for late learners to produce and discriminate.

Quite opposite to those predictions, it was found that the late learners failed to produce the /e/-/æ/ contrast accurately but had no difficulties in production of the /i/-/i/ contrast. Instead of establishing new categories for both /e/ and /æ/ vowels, late learners produced both /e/ and /æ/ as /e/, yet established separate categories for both
AusE /i/ and /ɪ/ vowels. Early learners did not differ from native Australian English monolinguals in production of the tested contrasts.

Furthermore, early and late learners’ perception of the /ɛ/ - /æ/ contrasts also differed: early learners perceptually assimilated two members of the contrast to two different Serbian categories (/e/ and /a/, respectively). Late learners perceptually assimilated both AusE /ɛ/ and /æ/ to a single Serbian category (/e/).

Perceptual assimilation and rating data obtained for the AusE /i/ - /ɪ/ contrast did not differ between the groups: both groups of Serbian-English bilinguals identified the vowels from the /i/- /ɪ/ contrast as instances of the Serbian /i/, with /ɪ/ being rated as a closer match to the Serbian category than /i/. In keeping with the production findings, both groups of bilinguals discriminated the /i/- /ɪ/ contrast accurately but the /ɛ/ - /æ/ contrast was difficult for late learners to discriminate, whereas early learners’ discrimination accuracy did not differ from monolinguals’.

The most significant findings of the first experiment were that the production accuracy of L2 segments was a predictor of perception accuracy, and that the perceived distance between L2 and L1 segments changed with increasing AOL (Flege, 1995). However, it is important to note that L2 perception and production are not necessarily always correlated (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Flege, Mackay, & Meador, 1999), and that production of some L2 segments may be easier to learn than perception of the same segments (Flege & Eefting, 1987). Furthermore, the two groups of bilinguals tested in this study differed not only in AOL, but in the length of exposure to AusE as their second language. Therefore, as in previous research within the SLM framework (Flege, Frieda & Nozava, 1997; Hakuta, Bialystok & Wiley, 2003; Yeni-Komshian, Robbins & Flege, 2001) age effects are a summary reflection of the effects related to, above all, the length of use of English as a second language. While the younger learners had at least 15 years of exposure and use of AusE, most of the older learners had only used AusE for between 5 and 10 years. It is possible that given the same time of use of AusE, late learners would eventually master the “difficult” contrast as well. Beyond the selection criteria regarding the length of AusE use, the amount of L1 and L2 use was not controlled in these studies. For example, it is known from previous research (Guion, Flege & Loftin, 2000) that amount of L1 use affects L2 production. Specifically, speakers who use their L1 more frequently might have more
difficulties in L2 segmental production and perception than speakers who use their L1 less frequently. Very often, older learners may make behavioural choices that bring them into frequent contact with other speakers of their L1 and limit their contact with L2 speakers (and the amount of L2 use). Early learners, on the other hand, may have weaker motivation to maintain their L1 and by not restricting their language interactions to the speakers of their L1, may gain more exposure to their L2. Taken together, it is important to note that AOL effects found in these studies are likely to be a reflection of other factors, such as the length of L2 use, the amount of L1 use, motivation, etc.

10.1.2 Study 2: Implications for the Perceptual Assimilation Model

In this study the relationship between the extent to which L2 phonetic segments can be assimilated to native vowel contrasts, and the discriminability of these L2 phonetic segments, was investigated based on predictions generated from the Perceptual Assimilation Model (Best, 1995).

Production and perception of the AusE /æ/ - /ə/ and /i/ - /u/ contrasts were tested in one production and two perception experiments, discrimination and perceptual assimilation with goodness-of-fit ratings. This time, in the perceptual assimilation experiment Serbian bilinguals were permitted to select from Serbian short and long vowels (in contrast to having their L1 choices restricted to only short vowels in Study 1).

Production results replicated those obtained in Study 1: the /æ/ - /ə/ contrast was difficult for late learners to produce (again both members of the contrast were produced as /ə/), whereas early learners did not differ from native speakers in producing both vowels accurately.

Early bilinguals perceptually assimilated the AusE /æ/ - /ə/ contrast to two different Serbian vowel categories (TC assimilation): /æ/ was perceptually assimilated to Serbian /a/, and /ə/ to Serbian /e/. Late bilinguals perceptually assimilated both /æ/ and /ə/ to Serbian /e/, but rated them as differing significantly in degree of perceived similarity to /e/ (CG assimilation type). AOL did not have an effect on perceptual assimilation of the /i/- /u/ contrast: both groups of bilinguals perceptually assimilated AusE /i/ to Serbian /i:/ and AusE /u/ to Serbian /i/ (TC assimilation).
The TC assimilation types obtained for the tested AusE contrasts in these experiments were good predictors of accuracy in discrimination. Early learners discriminated both /æ/ - /e/ and /i/ - /u/ contrasts well, and late learners discriminated /i/ - /u/ well. According to the PAM model, CG-categorized L2 vowel types are also expected to be discriminated well. Contrary to this hypothesis, however, the late learners had great difficulty discriminating the /æ/ - /e/ contrast. The poor discrimination performance on the AusE contrast that these same listeners assimilated to their L1 vowel as a CG difference might be the result of a confounding of CG and SC assimilation: neither of the AusE vowels in the /æ/ - /e/ contrast was perceived as an ideal match to the chosen Serbian vowel /e/, therefore suggesting that the late learners perceived them both as poor instances of a single Serbian vowel category, which caused them to ignore this intra-category difference in the discrimination task. When presented with one stimulus at a time in the goodness-of-fit rating task, the participants actually acknowledged the perceived intra-categorial differences by rating /e/ as better example of /e/ than /æ/.

There were three major findings from Study 2. First, the perceived quantitative distinction between AusE tense-lax vowels /i/ and /i/ was reflected in the vowels’ perceived similarities to Serbian long (\(\text{i:}\)) and short (\(\text{i}/\)) vowels. This indicates that experience with vowel duration in the bilinguals’ first language plays a role in perception of non-native vowel distinctions (Bohn, 1995). Second, the way bilinguals perceptually assimilated second language vowels to the closest vowel(s) in their first language was a relatively good predictor of discrimination accuracy (Best, 1995), though not a perfect predictor for the contrast on which late learners showed CG assimilation. Finally, AOL significantly affected the perceptual assimilation of non-native vowel contrasts.

10.1.3 Study 3: Phonological and phonetic priming effects on L2 vowel perception

The differences between early and late bilinguals in production and perception of AusE /æ/ - /e/ and /i/ - /u/ contrasts obtained in Study 1 and Study 2 suggest that AOL has an effect on perceptual assimilation of non-native L2 vowel contrasts to bilinguals’ native vowel categories. Late learners consistently failed to produce and
discriminate the /æ/ - /e/ contrast accurately, whereas early learners did not differ from native monolinguals in their performance on this contrast. The question addressed in this study was how stable the /æ/ - /e/ contrast in early learners actually is, given the rich body of evidence that even highly skilled second language speakers may not perform in all details exactly as native speakers of their L2 (Tees & Werker, 1984), and may rely on different acoustic parameters to differentiate non-native segments than those used by native listeners (Best & Strange, 1992).

Early and late Serbian-English bilinguals’ sensitivity to phonological and phonetic properties of the AusE /æ/ - /e/ and /i/ - /i/ contrasts was investigated using task-appropriate adaptations of the phonological priming method originally developed for probing the structure of the mental lexicon (i.e., word recognition). The relationships between the prime and the target (phonologically identical, phonetically similar and unrelated) and the interstimulus interval (1000 ms and 500 ms) were manipulated.

Priming under the 1000 ms ISI condition did not have a major effect on either AOL group’s perception of the tested contrasts. The perceptual assimilation patterns and discrimination accuracy obtained mirrored those obtained in the previous studies: early learners perceptually assimilated /æ/ to /a/ and /e/ to /e/ whereas late learners perceptually assimilated both /æ/ and /e/ to Serbian /e/. Both groups perceptually assimilated members of the /i/ - /i/ contrast to two Serbian categories, /i:/ and /i/, respectively. These assimilation patterns were good predictors of discrimination accuracy: early learners discriminated both the /i/ - /i/ and /æ/ - /e/ contrasts accurately, whereas late learners had difficulty discriminating only the /æ/ - /e/ contrast. In the phonetic prime condition, late learners’ discrimination accuracy for /æ/ - /e/ decreased relative to the baseline unrelated-prime condition; that is, they showed inhibitory phonetic priming, whereas early learners and monolinguals failed to show significant priming in either direction. It seems that in this condition only, the prime inhibited late learners’ discrimination of only this target vowel contrast. This inhibition may seem puzzling, given that late learners did not discriminate well between /æ/ and /e/ and therefore were not expected to perceive the phonetic differences between the prime and the target in 1000 ms condition. The most significant finding of the study was early learners’ change in perceptual assimilation pattern and discrimination accuracy of the
/æ/ - /ɛ/ contrast in the phonetic priming condition with the 500 ms ISI. In contrast to their robust and high performance on this contrast in the previous studies, in the phonetic priming condition with the shorter 500 ms ISI, now even the early learners had a tendency to assimilate the /æ/ - /ɛ/ contrast to a single Serbian category, /ɛ/. This assimilation pattern was reflected in early learners’ decrease in accuracy of distinguishing between /æ/ and /ɛ in the 500 ms condition of the discrimination test. An additional dimension of interest is that, in this short ISI condition, early learners often perceptually assimilated AusE /æ/ but not /ɛ/ to Serbian /aː/, when they did not assimilate it to Serbian /ɛ/, indicating that they nonetheless retained some degree of sensitivity to the differences in duration between the tense-lax /æ/ - /ɛ/ English vowel pair.

Neither facilitatory nor inhibitory priming was found for perceptual assimilation or discrimination of the /i/ - /ɛ/ contrast, regardless of the type of prime presented, the ISI differences or the AOL of the bilinguals.

10.1.4 Summary of the findings in the light of SLM and PAM

The choice of AusE and Serbian vowels tested in the studies reported were based on SLM’s predictions that those L2 phonetic segments that are “new” to L2 learners will be easier to perceive (and produce) than those L2 phonetic segments that are “similar” to the closest L1 segment. Successful acquisition of an L2 phonetic segment depends on successful differentiation of that segment from the closest L1 segment. Along with the perceived distance between L1 and L2 phonetic segments, the experiments are also concerned with the relationship between AOL and the perception of L2 vowels. Overall results have confirmed a strong relationship between AOL and the accuracy of L2 segmental perception and production. However, the distinction between “new” and “similar” L2 segments was not sufficient to explain why certain L2 vowels are easier to perceive and produce than others. A more reliable method for determining how L2 learners treat articulatory and acoustic differences between L1 and L2 vowels might come from the PAM’s view that perceptual accuracy of L2 vowels (perceptual discrimination) depends on how each member of the contrast is assimilated to a native vowel according to the different assimilation types.
The fact that younger learners demonstrated greater ability to perceptually differentiate between L1 and L2 vowels and greater accuracy in the production of L2 vowels is consistent with the SLM.

Given the clear evidence that not only L1 background but also the age when L2 is learned determines the way L2 vowels are perceived and produced, PAM’s predictions, when applied to bilinguals, should consider the AOL as well. Although PAM’s original framework does not explicitly deal with AOL (nor L2 segmental perception), the influence of AOL is compatible with PAM’s direct realist perspective. According to this perspective, perception of articulatory gestures changes over the lifespan and adults do not learn new speech information in exactly the same way as children exposed to the same information. Individual perceptual learning history cannot be separated from experience with one’s native language. Therefore, an adult L2 learner differs from a child L2 learner in that he has more exposure to the L1 but shorter perceptual learning history in the L2. Although not explicitly incorporated into PAM’s framework, a possible modification of the perceptual mechanisms in bilinguals differing in AOL comes from the evidence reported in this thesis that bilinguals exploit their (increasing) experience with L1 segments when learning L2 segments is consistent with the PAM (and PAM-L2) perspective. PAM’s assimilation types change over the course of L2 development.

10.2 Theoretical implications and future research: AOL and L2 vowel perception

Both the production and perception experiments in Study 1 and Study 2 and the perception experiments in Study 3 provide support for the hypotheses that AOL systematically affects acquisition of second language vowels. The first two studies differ from the third one in the strength of evidence they provide. In Study 1 and Study 2 early learners consistently matched the production and perception of AusE monolinguals on both vowel contrasts, whereas late learners accurately discriminated and produced the /ɪ/ - /ɨ/ contrast but failed to discriminate and produce the /æ/ - /e/ contrast accurately. Study 3 provided novel and intriguing evidence that the /æ/ - /e/ contrast is not perfectly established even in early bilinguals, and conversely that under
long-ISI conditions even late learners display some systematic sensitivity to phonetic differences between /æ/ and /ɛ/.

The results obtained in these studies are in agreement with previous findings (Flege, 1992; Munro, Flege, & Mackay, 1996; Yamada, 1995) that age of L2 learning, L1 phonology and phonetic dissimilarities between L2 and L1 vowels (Flege, Bohn, & Jang, 1997), as well as language independent auditorily based strategies (Bohn, 1995), can all influence L2 segmental production and perception. The initial hypothesis of Study 1, based on the Speech Learning Model (Flege, 1992, 1995, 2002), that late learners would have perceptual (and therefore production) difficulties with the /i/-/ɪ/ contrast and no difficulties with the /æ/-/ɛ/ contrast, were not supported by the data. Firstly, it was expected that the /æ/-/ɛ/ contrast would be “new” for late learners and therefore easy to acquire, whereas the /i/-/ɪ/ contrast was expected to be perceptually close to the Serbian /i/ and difficult to acquire. Serbian-English bilinguals’ experience with vowel length in their native vowel inventory had not been taken into consideration in the SLM-based approach. Hence, the explanation for excellent late learners’ performance on the presumably “similar” /i/-/ɪ/ contrast was provided by Study 2. The perceptual assimilation pattern of the /i/-/ɪ/ contrast as a two category (TC) assimilation revealed that late learners perceived the tense vowel /i/ as most similar to Serbian /iː/, and the lax /ɪ/ as (similar to) Serbian /ɪ/. This result is in line with the previous research on L2 learners’ use or weighting of a feature used in L1 for their learning of L2 phonology (Bohn, 1995; Gottfried & Suiter, 1997; McAllister, Flege, & Piske, 2002), which might be a result of language independent auditorily based strategies in L2 segmental perception (Bohn, 1995). For example, Bohn (1995) studied the perception of American English vowels varying in first formant frequency (F1) and duration by native speakers of Spanish and German. When identifying the stimuli, the listeners relied more heavily on duration than on spectral characteristics of the vowels. Although tense and lax vowels are contrasted in German by means of both durational and spectral cues, the native German speakers displayed a perceptual strategy which included the use of duration in the perception of non-native vowel distinctions. This perceptual strategy was evident even in Spanish participants, although Spanish does not use duration to signal phonological contrasts. On the basis of these findings, the near
native-like performance of late Serbian – English bilinguals on the tense-lax /i/ - /h/ vowel pair is not surprising, considering their experience with vowel duration in Serbian. In the case of the /æ/ - /ɛ/ contrast, where this temporal difference is reduced and is overridden by larger spectral differences between the vowels, late learners experienced difficulties, whereas early learners were able to rely on the spectral differences to discriminate the vowels.

We can discuss the late learners’ excellent performance in discrimination of the /i/ - /h/ contrast with greater confidence than their performance in the production experiments. In the production experiments, the native AusE monolinguals judged the Serbian-English bilinguals’ production of the tested contrasts simply in terms of the AusE vowels produced, using the keywords given on the screen. They did not rate the degree of foreign accent in those productions. One can speculate that the late learners are likely to have used a Serbian-like strategy in production of the /i/ - /h/ contrast, relying on a temporal distinction only (rather than both temporal and spectral) in production. If so, then their productions should be rated as foreign accented even though the AusE native listeners were able to accurately identify the intended vowels from the late learners’ pronunciations.

In Experiment 3, early learners’ specific perceptual assimilation pattern and discrimination performance for the /æ/ - /ɛ/ contrast, although generally in agreement with the Perceptual Assimilation Model (Best, 1995), raises one important implication. The early learners’ sensitivity to the “difficult” /æ/ - /ɛ/ phonemic distinction demonstrated in Study 1 and Study 2 must be attributed to the particular properties of the tasks employed in those experiments. Study 3 suggests that the /æ/ - /ɛ/ contrast might not be fully established even in early learners, yet the perceptual assimilation and discrimination experiments in Study 1 and Study 2 were not sensitive enough to tap into the difficulties that early learners experienced in the phonological priming condition with the 500 ms ISI. When exposed to a target preceded by a prime that contained the opposite member of the /æ/ - /ɛ/ contrast in the shorter ISI condition, early learners were able to perceive the difference between the prime and the target in both the perceptual assimilation and discrimination tasks (the same ability they demonstrated in Study 1 and Study 2), but this perceived distinction caused confusion similar to that
which the late bilinguals experienced with the same contrast across all conditions. Thus, the phonetic prime in the shorter, 500 ms ISI inhibited early learners’ perception of the /æ/ - /e/ contrast, indicating that they had failed to establish this contrast in a fully native-like way.

The present data contribute to ongoing debates concerning the way L1 shapes the perceptual system in early stages of development in such a way that it will determine the perception of non-native phonemic contrasts, even if there is extensive and early exposure to the L2, and even if the listener is a fluent daily user of the L2 (Sebastian-Galles & Soto-Faraco, 1999). That means that some L2 phonemic categories might not been processed with the same degree of accuracy as first language ones even in natural circumstances. For example, Sebastian-Galles and Soto-Faraco (1999) have found that highly proficient Spanish-Catalan bilinguals, who acquired Catalan in the first year of life and were selected for their ability to correctly identify Catalan phonemic distinctions, performed worse than Catalan-Spanish bilinguals on the same distinctions when a more sophisticated, sensitive gating technique was used. This result suggested that there are very strong constraints on the organization and initial acquisition of phonemic categories. The early exposure to the L1 might impose a strong influence on perception of non-native segments, preventing even early learners from forming new L2 categories perfectly. If this influence of L1 cannot be overcome, as the Spanish-Catalan findings and the data obtained in Study 3 in this thesis suggest, the unanswered questions are still how to establish the nature of the L1 influence in L2 segmental learning and how the acquisition of L2 segments influences perception of L1 categories. Despite the fact that both early and late bilinguals may have difficulties with the same non-native contrasts, the overall results of the studies reported here indicate that early learners still have a significant advantage over late learners in production and perception of a “difficult” vowel contrast.

A possible explanation for this advantage might be the state of development of the L1 at the time when L2 commences (Flege, 1997). That is, during early L2 acquisition for early learners, the L1 phonetic segments may not be fully established yet, allowing better establishment of newly acquired L2 segments. With increasing age, the L1 phonetic segments become stronger, imposing an increasing influence on L2 phonetic segments. If that is the case, then L2 phonetic segments should also impose some influence on bilinguals’ L1 phonetic segments, at least in early childhood. The
nature of bi-directional influences of bilinguals’ two languages is beyond the scope of this thesis. However, given that the studies reported in this thesis provided clear evidence not only for L1-L2 interaction, but also for changes in the way L1 and L2 interact with increasing AOL, future research should also investigate how L2 may influence L1 perception and production in both early and late Serbian-English bilinguals.

Furthermore, the results obtained in this thesis demonstrate the need for research concerning a fuller range of native and non-native segments as well as a wider range of participants differing in AOL and daily experience in both their languages. In the current studies the amount of L1 use was not controlled. According to previous research (Flege, Frieda, & Nozawa, 1997; Guion, Flege, & Loftin, 2000; Meador, Flege, & MacKay, 2000), amount of L1 use rather than AOL itself may be the better predictor of bilinguals’ performance in an L2. Even early bilinguals who use their L1 frequently (Flege, Schirru, & MacKay, 2003) may fail to establish a new vowel category for an L2 segment.

The results of the studies reported in this thesis attempt to account for the continuing interaction between bilinguals’ L1 and L2. Instead of providing definite answers, the interpretation of these findings is evidence that we need a fuller range of research concerning the phonological system in bilinguals. This research will only become valuable if the results are applied to bilinguals’ language learning in the classroom and out of the classroom environment. However, one ought not forget that second language learners differ in how successfully they adapt to and profit from instructions. While making an attempt to explain how a bilingual’s two languages interact, we have to keep in mind that the outcome of L2 learning is also closely related to the interaction between the learners’ characteristics and the learning context. Therefore, future research must take into account not only the nature of L1 and L2 segments, but individual differences between the speakers as well. Theories of intelligence, L2 aptitude, motivation, anxiety and emotion, and the relationship of native language abilities to L2 learning will have to be taken into account in L2 speech perception research.
REFERENCES


Fowler, C. A. (1991). Auditory perception is not special: We see the world, we feel the world, we hear the world. *Journal of the Acoustical Society of America, 89*, 2910-2915.


