THE ROLE OF L2 VOCABULARY EXPANSION IN THE PERCEPTION AND PRODUCTION OF AUSTRALIAN ENGLISH VOWELS BY ADULT NATIVE SPEAKERS OF JAPANESE

Rikke Louise Bundgaard-Nielsen
BA (English), MA (English and Psychology)
University of Aarhus, Denmark

Thesis submitted for the degree of Doctor of Philosophy
University of Western Sydney
March 2009
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 any other person, nor material which has been accepted for the award of any other degree or diploma at the University of Western Sydney, or any other educational institution, except where due acknowledgement 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.

___________________________________________
ACKNOWLEDGEMENTS

I would like to thank my panel of supervisors, Professor Catherine Best, Dr. Michael Tyler, Dr. Christian Kroos, and Dr. Satomi Kawaguchi for their expert guidance as well as generous support and kindness during my candidature. I have learnt so much from you, and I hope to continue to learn from you in the future.

I would also like to thank the MARCS Office Queens: Darlene Williams, Karen McConachie, Gail Charlton, Lurline Beard, and Mel Gallagher, as well as MARCS Lab Manager Colin Schoknecht, programmer Johnson Chen, and technical officer Steven Fazio for unfaltering support. I am also indebted to Dr. Takaaki Kuratate and Ms. Kikuko Nakamura for their assistance with countless things Japanese: どもありがとうございます。I also wish to thank my Japanese participants for their willingness to participate and their patience with a very taxing test battery; Mr. Koji Hotta of APC for assisting me in recruiting participants for my first experiment; my Australian English speakers; and my Australian English listeners.

My thanks also extends to my great friends at MARCS: My officemates Renee Glass, Iris-Corinna Schwarz, Dogu Erdener, Nathan Perry, Bronson Harry, and Karen Mulak; to my Noah’s Ark office mates Shaun Halovic and Kirk Olsen; to my RA-colleagues Anna Notley and Mary Broughton, and to my excellent proof-readers Tim Byron, Lizzie Beach, Damien Smith, Karen Mulak, Christa Lam-Cassettari, Mark Antoniou, Nathan Perry, and Carly Hayman, and the MARCS Speech & Language Writing Group for valuable input.

I will always be grateful to our Australian family for receiving us into their homes and hearts without reservation and for their continued support. Thank you, Elissa (Graunty), Dave (Poppa-Dave), Malda, Craig, and Kelcy. You made a faraway place feel like home! Likewise, I am grateful to my family and friends in Denmark whose support and encouragement always reminded me to do my best.

Last but not least, I wish to thank my ‘academic father’ Ocke-Schwen Bohn for his support and friendship over the past years.
For Thomas, *my* alchemist and fellow adventurer. Thank you!
ABSTRACT

Research indicates that adult Second Language (L2) learners typically do not become native-like in their perception or production of their L2, likely as a result of interference from their native language (L1). Research also indicates that L2 perception and production nonetheless improves with increased experience with the L2. Until recently, however, theories of L2 acquisition (e.g., SLM: Flege, 1995; PAM: Best, 1994) have been vague in their account of the processes underlying this improvement. The recent PAM-L2 (Best & Tyler, 2007), however, opens up new ways to understand experiential change in L2 perception and production. Centrally, PAM-L2 suggests that a large L2 vocabulary curtails change in L2 perception and production because it forces the learner to settle on an accented version of the L2 phonology.

The present thesis introduces the Vocabulary-Tuning Model of L2 Rephonologisation (Vocab Model). This model extends PAM-L2 by highlighting the facilitating effect of L2 vocabulary expansion, in early L2 immersion when the L2 vocabulary is still small, on the perception and production of an L2. It is further argued that the processes underpinning this improvement are analogous to those that underpin L1 acquisition in infants and toddlers. The thesis tests the Vocab Model in a series of studies (cross-sectional as well as longitudinal) of the perception and production of Australian English vowels by native speakers of Japanese, who have recently arrived in Australia for the purpose of acquiring English. The results show that L2 vocabulary size is indeed associated with L2 vowel perception and production and thus support the predictions of the Vocab Model. The thesis examines the usefulness of different criteria for L2-L1 vowel assimilation and discusses the findings in relation to results from L1-vowel perception research. The research design also pioneers a ‘whole system’ approach to cross-language vowel perception research that allows the learners to use all native vowels and all native vowel combinations (all three thesis studies), and to apply them to the full inventory of L2 vowels (Study 1). It is argued that results from such an approach more appropriately reflect the actual perceptual flexibility of the learners in a natural L2-immersion context than would a smaller subset of L1 and/or L2 vowels. This ‘whole system’ approach further suggests that L1 phonotactics is worthwhile to consider in future studies of L2 segmental perception and production.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ................................................................................................. III

ABSTRACT ..................................................................................................................... V

LIST OF FIGURES ......................................................................................................... XI

LIST OF TABLES ........................................................................................................... XIII

OVERVIEW ...................................................................................................................... XVIII

1 FACTORS IN SECOND LANGUAGE SPEECH PERCEPTION AND PRODUCTION ............................................................................................................. 1

1.1 Accent in the Mouth and on the Ears ..................................................................... 2

1.1.1 Factor 1: Native Language Background ............................................................. 3

1.1.2 Factor 2: Age of Acquisition ............................................................................. 5

1.1.3 Factor 3: Amount of L2 Experience ................................................................. 7

1.2 Theoretical Explanations ...................................................................................... 9

2 THEORIES ON SECOND LANGUAGE SPEECH PERCEPTION AND PRODUCTION ........................................................................................................... 10

2.1 The Speech Learning Model ................................................................................. 10

2.1.1 Interaction Between the L1 and L2 ................................................................. 10

2.1.2 SLM and Perception ......................................................................................... 12

2.2 The Perceptual Assimilation Model ..................................................................... 13

2.2.1 Direct Realism and Articulatory Phonology ....................................................... 13

2.2.2 PAM ................................................................................................................. 15

2.2.2.1 Perceptual Assimilation Patterns ................................................................. 16

2.2.2.2 Perceptual Discrimination of Non-Native Contrasts ................................... 17

2.2.3 PAM and Production ......................................................................................... 19

2.2.3.1 Motor Theory of Speech Perception and Production .................................. 19

2.3 SLM versus PAM: Differences Distilled ............................................................. 20

2.4 PAM-L2: New Inroads on Developmental Change in L2 Perception and Production ........................................................................................................... 20

3 THE ROLE OF L2 VOCABULARY EXPANSION IN L2 REPHONOLOGISATION ............................................................................................................... 22

3.1 The Vocabulary-Tuning Model of L2 Rephonologisation .................................... 22

3.2 Evidence for the Role of L1 Vocabulary Expansion in the Acquisition of L1 Phonology ........................................................................................................... 24

3.2.1 Early Acquisition of Phonetic Categories in the L1 .......................................... 24

3.2.2 Acquisition of Phonological Categories in the L1 .......................................... 25

3.2.3 Evidence for the Role of L1 Vocabulary Expansion in L1 Segmental Production ........................................................................................................... 27

3.2.4 L2 Vocabulary Development and L2 Segmental Production ................................ 29

3.3 Predictions for the Role of Vocabulary in Segmental L2 Perception and Production ........................................................................................................... 30

3.3.1 General Predictions for L2 Perception ............................................................. 30
3.3.2 Vocab Model Predictions for the Role of L2 Vocabulary Expansion in L2 Perception

4 OUTLINE OF RESEARCH PROJECT
4.1 The Nature of Vowels
   4.1.1 What are the Characteristics of Vowels?
      4.1.1.1 Tongue Height
      4.1.1.2 Tongue Position
      4.1.1.3 Lip-Rounding
      4.1.1.4 Vowel Visualisation
      4.1.1.5 Vowel Duration
      4.1.1.6 Vowel Dynamics
   4.1.2 The Vowel Inventories of Japanese and Australian English
      4.1.2.1 Australian English
      4.1.2.2 Japanese
   4.1.3 Experimental Series
      4.1.3.1 Chapter 5: Vocabulary Size Matters - The Assimilation of L2 Australian English Vowels to L1 Japanese Vowel Categories
      4.1.3.2 Chapter 6: Vocabulary Size is Associated with Improved L2 Vowel Perception Performance in L2 Learners
      4.1.3.3 Chapter 7: Second Language Learners' Vocabulary Expansion is Associated with Improved L2 Vowel Intelligibility

5 VOCABULARY SIZE MATTERS: THE ASSIMILATION OF L2 AUSTRALIAN ENGLISH VOWELS TO L1 JAPANESE VOWEL CATEGORIES
5.1 Abstract
5.2 Introduction
5.3 Method
   5.3.1 Stimuli
   5.3.2 Participants
   5.3.3 Procedure
5.4 Results
   5.4.1 General Assimilation Pattern
   5.4.2 Analysis by Vocabulary Size
   5.4.3 Uncategorised Vowels
5.5 Discussion
5.6 References

6 VOCABULARY SIZE IS ASSOCIATED WITH L2 VOWEL PERCEPTION PERFORMANCE IN ADULT L2 LEARNERS
6.1 Abstract
6.2 Introduction
6.3 Experiment 1: Vocabulary Size and L2 Segmental Perception in Newly-Immersed Adult L2 learners
   6.3.1 Method
   6.3.1.1 Participants
   6.3.1.2 Stimuli
   6.3.2 Procedure
   6.3.3 Predictions
   6.3.4 Results
8 THESIS DISCUSSION AND CONCLUSIONS .............................................161
8.1 The Foundations of the Thesis ....................................................... 161
  8.1.1 The Factors of L2 Acquisition .................................................. 161
  8.1.2 PAM-L2 .............................................................................. 163
  8.1.3 The Vocabulary-Tuning Model of L2 Rephonologisation ............ 163
8.2 The Present Findings on the Role of L2 Vocabulary Size in L2 Segmental Perception and Production ................................................. 164
  8.2.1 L2 Perception and L2 Vocabulary Size .................................... 164
  8.2.2 L2 Production and L2 Vocabulary Size .................................... 166
8.3 Conclusions ..................................................................................... 167
  8.3.1 L2 Vocabulary Size Matters .................................................... 167
  8.3.2 The Benefit of a ‘Whole Systems’ Approach .............................. 169
  8.3.3 Different Assimilation Criteria ................................................ 170
8.4 New Directions ............................................................................... 170
9 REFERENCES ...................................................................................... 174
10 APPENDIX 1: ARE AUSTRALIAN ENGLISH R-VOWELS MERGING WITH THEIR R-LESS MATES? A VOWEL SUBSYSTEM DISTINCTION IN COLLAPSE ................................................. 196
  10.1 Abstract .................................................................................... 196
  10.2 Introduction ............................................................................... 197
    10.2.1 Background ....................................................................... 197
    10.2.2 Application of the Shell Framework to a Specific Vowel System: AusE 199
  10.3 Method ....................................................................................... 202
    10.3.1 Stimuli .............................................................................. 202
    10.3.2 Participants ....................................................................... 209
    10.3.3 Procedure ....................................................................... 209
  10.4 Results ....................................................................................... 210
    10.4.1 AXB Discrimination ......................................................... 210
    10.4.2 Identification ..................................................................... 211
  10.5 Discussion ................................................................................... 213
  10.6 References ............................................................................... 216
11 APPENDIX 2: SECOND LANGUAGE SPEECH LEARNING WITH DIVERSE INPUTS ................................................. 219
  11.1 Introduction .............................................................................. 220
  11.2 Methods .................................................................................... 226
    11.2.1 Subjects .......................................................................... 226
    11.2.2 Procedure (Production Data) .............................................. 226
    11.2.3 Procedure (Intelligibility Data) .......................................... 227
  11.3 Results ....................................................................................... 227
LIST OF FIGURES

Figure 2.1. Gestural score for the English word *pan* (adapted from Browman & Goldstein, 1995). The first 150 ms indicates the articulation of /p/, while /æ/ is produced between 100 and 300 ms into the utterance, and finally, /n/ is produced in the last 150 ms of the utterance. As is clear from the grey boxes in the figure, the activation periods for the specified articulators overlap. These overlaps indicate gestural coordination as well as coarticulation between the coordinative structures............ 14

Figure 4.1. An overview of vowels (Handbook of the IPA). ............................................. 36

Figure 4.2. F1, F2, F3 shown on a spectrogram of the vowel /æ/. ................................. 38

Figure 4.3. The Australian English monophthongs............................................................ 39

Figure 5.1. Percent total of the three most selected Japanese categories for the five uncategorised AusEWS vowels for the High Vocabulary group (HV), and the Low Vocabulary group (LV). ........................................................................................................... 60

Figure 6.1. Discrimination accuracy for the contrasts SC (/i:-/æː/), TC (/i:-/ɪ/), UC (/æ/-/u:/), UU/UC (/æu/-/o:/), overall discrimination score (Overall) as well as overall discrimination score without the SC contrast (Overall-SC) for the two vocabulary size groups, HV and LV. Error bars reflect S.E.M............................................................. 92

Figure 6.2. Discrimination accuracy for the SC (/i:-/æː/), TC (/i:-/ɪ/), UC (/æ/-/u:/), and UU/UC (/æu/-/o:/) contrasts and overall discrimination accuracy including SC (Overall), as well as overall accuracy excluding SC (Overall-SC). Error bars reflect S.E.M......................................................................................................................... 103

Figure 6.3. Closer look at the change in discrimination for the four contrast types (SC, TC, UC, and UU) as well as the overall discrimination score (Overall) and the overall discrimination score excluding the SC contrast between the two testing times for the two LV-participants. Improvement in performance between the two testing times is indicated by ellipses. Error bars reflect S.E.M.......................................................... 105

Figure 6.4. Closer look at the change in discrimination for the four contrast types (SC, TC, UC, and UU) as well as the overall discrimination score (Overall) and the overall discrimination score excluding the SC contrast between the two testing times for the
seven HV-participants. Improvement in performance between the two testing times is indicated by ellipses. Error bars reflect S.E.M. .......................................................... 106

**Figure 10.1.** Stimulus segmentation in Praat. .......................................................... 203

**Figure 10.2.** Dynamic vowel trajectories for AusE<sub>WS</sub> monophthongs. .......................... 204

**Figure 10.3.** Dynamic vowel trajectories for AusE<sub>WS</sub> diphthongs.............................. 204

**Figure 10.4.** Dynamic vowel trajectories for AusE<sub>WS</sub> r-context vowels. .................. 205

**Figure 10.5.** Dynamic vowel trajectories for AusE<sub>WS</sub> word final /a/ (schwa). ............ 205

**Figure 11.1.** The monophthongs of Danish and 3 varieties of English in the F1/F2 space. ................................................................................................................................... 224
LIST OF TABLES

Table 5.1. Citation-context (C) vowel durations (ms) and F1, F2, and F3 values (in Hz) at 25%, 50%, and 75% of the target vowels and diphthongs of Australian English (AusE<sub>WS</sub>). ........................................................................................................................................................................ 54

Table 5.2. Sentence-context (S) vowel duration (ms) and F1, F2, and F3 (in Hz) values at 25%, 50%, and 75% of the vowels and diphthongs of Australian English (AusE<sub>WS</sub>). 55

Table 5.3. The assimilation of citation AusE (left item in ‘Identification’ columns) to L1 Japanese (right item in ‘Identification’ columns) vowel categories by all listeners. GR is the mean goodness rating on a 7 point scale (1 = poor fit, 7 = excellent fit)............... 58

Table 5.4. Assimilation patterns to the top three selected L1 Japanese vowel categories (vertical) for the five uncategorised AusE<sub>WS</sub> vowels (horizontal). Group refers to High Vocabulary learners (HV) and Low Vocabulary learners (LV). % indicates the total percent of each of the uncategorised AusEWS vowels which were identified as each of the top three L1 categories. ........................................................................................................ 61

Table 6.1. Citation vowel durations (ms), word duration (ms) and F1, F2, and F3 values (in Hz) at 25%, 50%, and 75% of the target monopthongs and diphthongs of Australian English (AusE). ..................................................................................................................................................... 79

Table 6.2. Perceptual assimilation pattern for the UU contrast /o//-/æu/, adapted from Chapter 5. The table presents the top three selected L1 Japanese vowel categories (columns) for the two uncategorised AusE<sub>WS</sub> vowels (rows). Group refers to High Vocabulary learners (HV) and Low Vocabulary learners (LV) of the particular study (Chapter 5). The % columns indicate the total percent of each of the uncategorised AusEWS vowels which were identified as each of the top three L1 categories. .......... 82

Table 6.3. Vocabulary Group comparisons in mean Length of Stay (LOS) in Australia, Age of Immersion (AOI) in English, Age of Learning (AOL) onset, number of years of L2 study (Study), and L2 vocabulary size (Vocab). ................................................................. 83

Table 6.4. Assimilation pattern (percentage of each Australian English vowel assimilated to specific Japanese vowel categories, averaged over participants) and Goodness Ratings (on a scale of 1-7) for the High Vocabulary (HV) group with a 50% assimilation criterion. Grey indicates an Uncategorised L2 vowel. .............................. 85
Table 6.5. Assimilation pattern (percentage of each Australian English vowel assimilated to specific Japanese vowel categories, averaged over participants) and Goodness Ratings (on a scale of 1-7) for the Low Vocabulary (LV) group with a 50% assimilation criterion. Grey indicates an Uncategorised L2 vowel. 

Table 6.6. Assimilation pattern (percentage of each Australian English vowel assimilated to specific Japanese vowel categories, averaged over participants) and Goodness Ratings (on a scale of 1-7) for the High Vocabulary (HV) group with a 70% assimilation criterion. Grey indicates an Uncategorised L2 vowel. 

Table 6.7. Assimilation pattern (percentage of each Australian English vowel assimilated to specific Japanese vowel categories, averaged over participants) and Goodness Ratings (on a scale of 1-7) for the Low Vocabulary (LV) group with a 70% assimilation criterion. Grey indicates an Uncategorised L2 vowel. 

Table 6.8. Paired t-tests of the significance (Bonferroni-corrected) of discrimination difference of the four Australian English vowel contrasts SC, TC, UC, and UC. 

Table 6.9. Most frequently selected Japanese vowel category, in percent, across participants, for assimilation of AusE /o:/ in both Citation and Sentence presentation in Bundgaard-Nielsen et al., submitted (Chapter 5). 

Table 6.10. Assimilation pattern and Goodness Rating (GR) at Time 1 for the nine participants of Experiment 2. Light grey indicates an uncategorised L2 vowel using a 50% assimilation criterion. Dark grey indicates an uncategorised vowel using a 70% assimilation criterion. 

Table 6.11. Assimilation pattern and Goodness Rating at Time 2, Experiment 2. Light grey indicates an uncategorised L2 vowel using a 50% assimilation criterion. Dark grey indicates an uncategorised vowel using a 70% assimilation criterion. 

Table 6.12. Paired t-tests of the difference in the discrimination average of the four AusE contrasts at the two test sessions. * indicates a statistically significant difference. 

Table 7.1. Overall identification pattern, in percent, summed across all participants. The intended L2 vowels are presented in the leftmost column, while the top row indicates what the native listeners selected/perceived from among all 18 AusE vowel choices. Bold-face with light gray cell indicates the percentage of each vowel identified as
intended. Darker gray cells indicate greater mis-identification as a non-intended vowel than as the correct target. ................................................................. 132

Table 7.2. Overall identification pattern, in percent, summed across LV participants. The intended L2 vowels are presented in the leftmost column, while the top row indicates what the native listeners selected/perceived from among all 18 AusE vowel choices. Bold-face with light gray cell indicates the percentage of each vowel identified as intended. Darker gray cells indicate greater mis-identification as a non-intended vowel than as the correct target. ........................................................................... 133

Table 7.3. Overall identification pattern, in percent, summed across HV participants. The intended L2 vowels are presented in the leftmost column, while the top row indicates what the native listeners selected/perceived from among all 18 AusE vowel choices. Bold-face with light gray cell indicates the percentage of each vowel identified as intended. Darker gray cells indicate greater mis-identification as a non-intended vowel than as the correct target. ........................................................................... 134

Table 7.4. Difference in the intelligibility score for the HV group versus the LV group. The leftmost column indicates the intended vowels, while the percent correctly identified vowels are presented horizontally. The reported score is calculated as the difference in the mean percent of a vowel identified as intended between the HV group and the LV group. The score thus reflects the increase in identifiability between the vowels produced by the LV and the HV groups, respectively. .................... 135

Table 7.5. Overall identification pattern at the Time 1, in percent, summed across the nine participants. The intended L2 vowels are presented in the leftmost column, while the horizontal indicates what the native listeners perceived. Bold-face with light gray cell indicates the percentage of each vowel identified as intended. Darker gray cells indicate greater mis-identification as a non-intended vowel than as the correct target. ........................................................................................................... 141

Table 7.6. Overall identification pattern at the Time 2, in percent, summed across the nine participants. The intended L2 vowels are presented on the leftmost column, while the horizontal indicates what the native listeners perceived. Bold-face with light gray cell indicates the percentage of each vowel identified as intended. Darker gray cells
indicate greater mis-identification as a non-intended vowel than as the correct target.

Table 7.7. Difference in the intelligibility score at the two testing times. The leftmost column indicates the intended vowels, while the difference between the identifiability of the vowels at the two testing times is presented horizontally. The reported score is calculated as the difference in the mean percent of a vowel identified as intended at the first test session and the second test session. The score thus reflects the increase in intelligibility between the vowels produced by the participants from the first to the second test session.

Table 7.8. Percent of each intended vowel (presented on the leftmost column) which was identified as intended by the native listeners (presented horizontally) for the two LV-participants in the first testing session of Experiment 2.

Table 7.9. Percent of each intended vowel (presented on the leftmost column) which was identified as intended by the native listeners (presented horizontally) for the two LV-participants in the second testing session of Experiment 2.

Table 7.10. Table of difference between the intelligibility (i.e., % of vowels identified as intended) of the Japanese accented vowels at the first and the second test session. The leftmost column indicates the intended vowels, while the difference between the identifiability of the vowels at the two testing times is presented horizontally. The numbers indicating the increase (or decrease) in the percent of vowels identified as intended between the first and the second testing session for the LV-participants.

Table 7.11. Acoustic measurements for the 7 male participants, Experiment 1.

Table 7.12. Acoustic measurements for the 19 female participants, Experiment 1.

Table 7.13. Acoustic measurements for the 2 male participants in Experiment 2, Test Time 1.

Table 7.14. Acoustic measurement for the 2 male participants in Experiment 2, at Test Time 2.

Table 7.15. Acoustic measurements for the 4 female participants in Experiment 2, at Test Time 1.

Table 7.16. Acoustic measurements for the 4 female participants in Experiment 2, at Test Time 2.
Table 10.1. Vowel duration and F1, F2, and F3 values (in Hz) at 25%, 50%, and 75% of the vowel. .......................................................... 207

Table 10.2. Confusion matrix in percentages for the 18 AusEWS vowels..................... 212

Table 11.1. Confusion matrix for 10 L1 Danish speakers’ production of English /bVt/ syllables as identified by 10 L1 Canadian English listeners. Each cell lists the percent responses (given horizontally) to the intended productions (listed vertically, three tokens/speaker). .......................................................... 228

Table 11.2. Confusion matrix for 5 L1 English speakers’ production of English /bVt/ syllables as identified by 10 L1 Canadian English listeners. Each cell lists the percent responses (given horizontally) to the intended productions (listed vertically, three tokens/speaker). .......................................................... 230

Table 11.3. Confusion matrix for 2 L1 speakers of American English ...................... 231

Table 11.4. Confusion matrix for 2 L1 speakers of Southern British English.............. 232

Table 11.5. Confusion matrix for 1 L1 speaker of Australian English ....................... 233

Table 12.1. The assimilation of AE (left) to JP (right) vowel categories ................... 239

Table 13.1. Formant values at 25, 50 and 75% of /ɪ/ and /ɜ/. * indicates statistically significant difference between these two vowels. ........................................ 244
And the Gileadites took the passages of Jordan before the Ephraimites: and it was so, that when those Ephraimites which were escaped said, Let me go over; that the men of Gilead said unto him, Art thou an Ephraimite? If he said, Nay; Then said they unto him, Say now Shibboleth: and he said Sibboleth: for he could not frame to pronounce it right. Then they took him, and slew him at the passages of Jordan: and there fell at that time of the Ephraimites forty and two thousand. 

Book of Judges, 12:5-6.

OVERVIEW

The above quote from the Book of Judges testifies to the great importance of hearing and speaking a Second Language (L2) well enough to be accepted by native speakers of that language. While the war-time tragedy reflected in the quote (for an in-depth discussion, see Faber [1992]) may not reflect the experience of most L2 learners today, many adult L2 learners continue to face serious socio-economic disadvantages, such as decreased chances of gaining employment as well as exclusion from various spheres of society. In this light, research into the processes of L2 acquisition leading to increased understanding of the phenomenon, as well as improvements L2 instruction, may have significant impact on the lives of countless of L2 adult learners and lead to a wide range of improvements to their opportunity to contribute to all aspects of society.

The present thesis concerns the perception and production of Second Language (L2) Australian English vowels by native (L1) speakers of Japanese who differ in their L2 vocabulary sizes. It tests the hypothesis that L2 vocabulary development drives the acquisition of L2 phonology and also examines the effect of immersion duration on L2 segmental perception and production, with the larger goal being to explore the timing and rate of perceptual change.

The presented series of experiments are based on a rich literature documenting the effects of L1-L2 interaction, Age of Acquisition (AOA) of the L2, and amount of L2 experience on L2 perception and production. The experiments are driven by questions raised by...
theories that explain the phenomenon of L2 perception and production (see Chapter 1). In particular, they draw upon the Perceptual Assimilation Model (PAM), and its extension PAM-L2, which provide useful theoretical frameworks for the examination of developmental change in L2 segmental perception and production. It also considers the Speech Learning Model (SLM), though this theory has yet to be expanded to describe the acquisition processes underlying change in L2 perception and production (see Chapter 2), and very briefly touches upon the Motor Theory of Speech Perception and its prediction for the timing of changes to L2 perception and production in relation to PAM predictions (also Chapter 2).

The core of the thesis consists of the PAM-L2 extension, the Vocabulary-Tuning Model of L2 Rephonologisation, or Vocab Model (see Chapter 3), that outlines the role of L2 vocabulary expansion in L2 perception and production, and results from the three experimental Chapters (Chapters 5-7) testing the outlined predictions for the role of L2 vocabulary expansion in L2 perception (Chapters 5 and 6) and production (Chapter 7). These chapters have been written as journal articles and submitted for publication in peer-reviewed journals. They are largely presented in their submitted form, though the subheadings have been re-numbered to reflect the chapters in which the papers are presented here.

The thesis proper concludes with Chapter 8 summarising the experimental findings of Chapters 5-7 and placing the presented experimental findings in the context of L2 perception and production research. Finally, it considers the contributions to L2 acquisition research and theory made by the completion of this work.

The Appendices presented on the enclosed CD consist of copies of the written materials used in the thesis experiments, such as recruitment posters, information to participants, consent forms, the relevant vocabulary size test, and the instructions given to the participants. Also included is a copy of an email received from a Japanese participant, which reflects very well some of the obstacles that had to be overcome in the recruitment process. Finally, the Appendices contain copies of two one-page papers presented at InterSpeech 2008 (Appendix 3 and 4), an additional journal article currently under revision
for resubmission (Appendix 1), as well as a copy of a peer-reviewed book chapter (Appendix 2), based on my MA thesis, co-authored and published during my PhD candidature. While the results from the papers in Appendices 1, 2, 3, and 4 are discussed briefly in relation to the findings presented in the thesis proper, these studies are not narrowly relevant to my thesis topic and are therefore not included as experimental thesis chapters.

A Note on Notation

As a matter of good house-keeping, I will use the word phonetic (and give examples in square brackets [ ]) to refer to the actual details of realisation of a given speech segment (phone), which may vary according to context, regional dialect or language (other L1), and I will use the word phonological (and give examples between slashes //) to refer to a ‘higher’ level of organisation/perception in which the abstract segmental category is crucial while actual phonetic details or the realization (context, dialect or even language) are less relevant or even irrelevant.
1 FACTORS IN SECOND LANGUAGE SPEECH PERCEPTION AND PRODUCTION

Human speech is a highly complex, continuous, and finely coordinated execution of detailed motor programs that regulate the frequencies, volume, and quality of vocalisations, and also facial movements, in a manner such that these speech acts can be perceived and recognised by other human listeners (Clark & Yallop, 1995). Speech production normally begins as air passes from the lungs in the thorax through the various speech organs and out through the oral and/or nasal opening, while a number of these speech organs vibrate, move, and occasionally obstruct the passing of the air partially or completely (Clark & Yallop, 1995) and in this way create oscillating changes of air pressure, that is, sound. Speech perception is the picking-up and categorisation of these speech acts by the combined auditory, visual, and (via proprioception) tactile systems, popularly demonstrated by the McGurk effect (McGurk & MacDonald, 1976; Green & Kuhl, 1989; 1991; Fowler & Dekle, 1991). The McGurk effect is the very robust phenomenon of the integration of visual and auditory speech perception such that, for instance, a visually presented /g/ and an auditorily presented /b/ will be perceived as a /d/. It is believed to indicate that speech perception is multi-modal, i.e., relies on the perception of input from more than one perceptual modality.

Speech perception displays both categorical (particularly for consonants: Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) and continuous properties (particularly for vowels: Fry, Abramson, Eimas, & Liberman, 1962; Beddor & Strange, 1982; Polka, 1995). Categorical perception is characterised by great ease in discriminating between segments (or phones) that are contrastive, such as English /t/ and /d/, and of great difficulty in discriminating between acoustically different tokens from the same phonological category. This is easily illustrated by an artificially generated equal-step acoustic continuum between, for instance /t/ and /d/, which to listeners will sound like a string of identical /t/s that suddenly changes to a string of identical /d/s, despite the fact that the tokens are equally spaced in terms of their acoustic properties. In perceptual experiment, this pattern will typically be reflected in consistent within-category phoneme identification as well as a sharp shift at the boundary between these categories, which in turn corresponds
to a peak in discrimination at the phoneme boundary. This pattern of discrimination parallels results from other domains of perception such as the visual domain (e.g., colour perception: Bornstein, 1987; Roberson, Davies, & Davidoff, 2000; face perception: Kotsoni, de Haan, & Johnson, 2001). Continuous perception refers to a behavioural pattern in which the listener remains sensitive to some within-category differences. In a perceptual experiment, this would translate into good discrimination throughout the continuum as well as the absence of a discrimination peak at the category boundary, while the identification pattern would be relatively consistent at the extremes of the continuum, with a gradual shift from one category label to the other across the category boundary.

Research has identified a number of key problems that must be solved by any account of speech perception/production. Central are the problems of how a listener successfully normalises the variance in the speech signal due to non-linguistic factors, such as individual differences in size and shape of the speech organs (Fant, 1968), and due to linguistic factors such as contextual variation in the speech output, i.e., co-articulation, allophonic distributions, and segment duration differences due to variation in speaking rate (Nearey, 1989).

In addition to the problems of ‘normal’ linguistic and non-linguistic variance in speech production and perception, problems arise when an individual has to acquire and use a non-native language, and these added problems are the focus of Second Language (L2) perception and production research.

1.1 Accent in the Mouth and on the Ears

Despite individual differences in language learning abilities, most adult L2 learners speak their new language with a foreign accent that is characteristic of their particular Native Language (L1) background (Flege, Munro, & MacKay, 1995; Flege, 2002; Munro, Flege, & MacKay, 1996). Less obvious, but equally important, is that learners of an L2 also have ‘accented’ perception of the new language. In particular, language learners have perceptual difficulties with L2 phonemes that do not occur in their L1 (Bohn, 1995; Munro et al., 1996, Flege & MacKay, 2004).
Accentedness in L2 perception and production is systematically related to three core factors. These factors, to which we will turn individually below, are:

1. the perceived similarities between the phonological inventory of the learner’s L1 and L2;

2. the age of the learner at the time of L2 acquisition; and

3. the amount of experience that the learner has had with the L2.

1.1.1 Factor 1: Native Language Background

The native language background of an L2 learner greatly affects their L2 acquisition, and speakers of two different languages (such as English and Japanese) will not necessarily have the same pattern of difficulties in perception and production of phones in the same foreign language. A widely used example of this L1-effect is the English vowel pair /ɪ/ and /ɨ/. These two vowels differ both with respect to spectral qualities, and with respect to duration, with /ɪ/ being longer than /ɨ/. However, they have been shown to be very difficult for native speakers of Spanish to discriminate, with both perceived as exemplars of a single Spanish vowel category /ɨ/. (Escudero & Broersma, 2004; Flege et al., 1997; Morrison, 2002). On the other hand, this contrast is discriminated much more easily by native listeners of German, a language with a parallel vowel contrast – this highlights the significance of the L1-L2 matching, i.e., the similarities and differences between the L1-L2 pair in question (Bohn & Flege, 1990; Flege et al. 1997). The consonantal context of the vowels, however, also matters, such that for instance American English rounded back, mid-high and high vowels are systematically fronted when they follow coronal consonants (Strange, Bohn, Trent, & Nishi, 2004). Furthermore, while experienced German speakers of English produce English /æə/ in a more native-like manner than less experienced learners, they do not show a similar effect of L2 experience on other English vowels such as /i, ɪ, ɛ/ (Bohn & Flege, 1992), presumably because the German learners successfully use their native German vowels for those three English vowels.

Another widely studied example of how the L1 of learners affects their L2 perception and production is the perception and production of the English /ɹ/ and /l/ (rock-lock) by
Japanese learners of English, whose native language contains only /r/, which is acoustically dissimilar to both /u/ and /l/ (Goto, 1971; Miyawaki et al., 1975; MacKain, Best, & Strange, 1981; Strange et al., 1998; Mochizuki, 1981).

In addition to ‘perceived similarity’, be it acoustic or articulatory, between the L1 and L2, the relative sizes of the L1 and L2 inventories are important for the learner; it is harder for speakers of L1s with relatively small vowel inventories (e.g. Spanish) to acquire a rich L2 vowel inventory than for speakers of L1s with larger vowel inventories (e.g., German, Norwegian), perhaps because native speakers of languages with small L1 inventories have fewer native categories to exploit in their attempt to make sense of L2 input (Iverson & Evans, 2007).

The influence of L1 background on L2 perception and production applies to both consonants and vowels, but studies of L2 vowel discrimination have shown that L1 has less influence on the perception of L2 vowels than it does on the perception of L2 consonants (Bohn, 1995). This suggests that, in addition to the influence from the L1, vowel perception may be subject to certain universal influences, such as the reported high salience of the so-called corner vowels /i, a, u/ (Polka & Bohn, 2003; Polka, Bohn, & Molnar, 2005; Bohn, 2007), as well as differences in the articulation of vowels and consonants. Consonants are produced with a clearly specified place of constriction between clearly specified articulators (e.g., tongue tip, blade, or back in combination with teeth, alveolar ridge or velum), while vowels are produced without a clear point of constriction, and with just one articulator, the tongue body, as the main contributor in all vowels (lips and tongue tip may or may not contribute secondarily).

Irrespective of native vowel inventory size and organisation, all L2 learners pay attention to the same spectral and durational information, and (vowel) duration often becomes an important cue to the non-native learner, particularly when the spectral/formant information is ambiguous or otherwise insufficient to allow the learner to perform reliable discrimination (Bohn, 1995; Cebrian, 2006; Gottfried & Beddor, 1988; Bohn & Flege, 1990). For speakers of languages in which durational differences are phonologically contrastive even for spectrally-identical vowels (such as Serbian or Japanese), durational
differences may be even more salient, particularly in L2 learning situations in which learners are forced to discriminate and categorise rather large L2 vowel inventories as, for example, in English.

However, the interaction between the phonological systems of the learner’s L1 and L2 and proposed universal influences are not the only factors affecting the L2 acquisition of a learner. Other factors, such as the learner’s age of acquisition (AOA; see Factor 2) and the amount of experience (AOE; see Factor 3) they have had with the L2, also play significant roles in L2 acquisition.

1.1.2 Factor 2: Age of Acquisition

A large number of studies have identified what is commonly observed; namely that child L2 learners typically outperform adult L2 learners, both in terms of perception and production (for consonants, see Flege et al., 1995, Yamada, 1995; MacKay, Meador, & Flege, 2001; for vowels, see Munro et al., 1996; Flege, Yeni-Komshian, & Liu, 1999; Piske, Flege, MacKay, & Meador, 2002; Flege, Schirru, & MacKay, 2003; for overall accentedness, see Seliger et al., 1975; Oyama, 1976; Yeni-Komshian, Flege, & Liu, 2000).

The single most influential hypothesis about the AOA effect on L2 learning has been the Critical Period Hypothesis (CPH: Lenneberg, 1967; see also Johnson & Newport, 1989; Oyama, 1976). Inspired by research on imprinting in non-human animals, the CPH posits that language learning must take place during a specific early period in human development, during which the brain is sensitive to linguistic input received from the surroundings. The CPH also claims that language acquisition taking place before the end of the period of maturational sensitivity is based on innate biological structures in the human brain. Thus, learning after the end of that critical period may not draw on the same innate biological structures. Late acquisition would instead depend on general learning mechanisms, with the result that the outcome of late language acquisition will be poorer than that of early acquisition.

The critical period is often claimed to end around puberty, when a number of profound maturational changes occur in the body and brain, though a number of researchers working
within the CPH framework have suggested an even earlier decline to brain sensitivity at 9-10 years or even 5-6 years (Flege, 1988).

A growing number of studies, however, have failed to find support for the CPH in some domains of language, such as the lexicon, while still identifying AOA effects in other domains, such as phonetics and phonology (Flege, 1999; Scovel, 1988). In contrast to the predictions of CPH, older learners may indeed have certain advantages over younger learners in the early stages of L2 acquisition, such as greater ability to learn grammar explicitly, and greater meta-linguistic consciousness of the acquisition of the prosodic pattern of the new language (Harley & Wang, 1997; Flege, 1999; Yeni-Komshian et al., 2000; Hakuta et al. 2003). Indeed, the CPH is not even necessarily supported by evidence of young learners outperforming older learners in certain language domains such as phonetics and phonology. A critical period should result in markedly poorer immersion-based L2 or classroom-based Foreign Language (FL) acquisition (including acquisition of new phones) in post-puberty learners, at 14+ years of age, as compared to learners acquiring the new language immediately prior to puberty, at 11 years of age, for instance. However, this CPH prediction has not been documented - rather, the performance of L2 learners becomes poorer in a linear, gradual fashion with no clear cut-off point at the onset of puberty (Flege et al., 1995; 1999; Hakuta et al., 2003). Moreover, a recent study of late (adult) and early (child) L2 learners of English with Korean backgrounds found that while the children always outperformed the adults, the children did not perform on par with L1 speaking control participants, as would have been predicted by the CPH. Furthermore, there was no difference in the performance of children who had started to acquire English prior to puberty and those who had not begun their L2 acquisition until reaching the age of 14 years (Tsukada et al., 2005).

In addition, evidence from FL acquisition indicates that an earlier onset of FL teaching in schools does not necessarily yield better FL speakers in the long run. Conversely, some older learners (i.e., those acquiring a new language after puberty) have been reported to become highly fluent and proficient L2 users with imperceptible L1-accents (Ioup, Boustagui, & Moselle, 1994).
Finally, pronunciation proficiency in both the native and non-native language of Korean-English bilinguals is more consistent with the interpretation that deviations from native pronunciation norms arise from various interactions between the two languages of a bilingual, rather than from a critical period for language learning (Yeni-Komshian et al., 2000). In fact, this study found that while age of L2 acquisition was inversely associated with L2 pronunciation proficiency, the degree of non-native accent in the L2 was also inversely associated with degree of non-native accent on the L1: A native-like L2 pronunciation and a native-like L1 pronunciation thus seem to be mutually exclusive, to some extent.

Therefore, language learning may be dominated by a sensitive rather than a critical period in language acquisition. Whereas a critical period for language learning means that learning must take place within a narrow time window (such as pre-puberty), a sensitive period means that learning is optimal in a certain time window but can also happen outside of that window, allowing for a less marked and more gradual demise of the language acquisition skills of L2 and FL learners (Flege, 1995; Hakuta et al., 2003; Harley et al., 1997). Other researchers have suggested that it is not a loss of language learning abilities as such that results in poorer performance from older learners, but perhaps interference from a better established native language in the older learners (Best & Tyler, 2007, see also Flege, 1995, though the implications from SLM may be narrower than those of PAM/PAM-L2). From the PAM/PAM-L2 (and SLM) perspective, AOA may be more clearly conceptualised as the learner’s developmental L1-age at the onset of L2 acquisition: a more firmly established L1 (and thus more advanced L1-age) may effect perception and production of an L2 more than a less firmly established L1 (and hence younger L1 age).

1.1.3 Factor 3: Amount of L2 Experience

A third important factor in L2 acquisition is the amount of L2 experience that a particular L2 learner has had; learners who have had extensive L2 experience typically perform better on both L2 perception and production tasks than learners who have only had minimal L2 experience (Bohn & Flege, 1992; Flege, Bohn & Jang, 1997; Best & Strange, 1992; MacKain et al., 1981). An exception to this pattern can however be imagined in situations when increased experience with the L2 leads to systematic reorganisation of the
L1/L2 phonetic and phonological inventories (through the addition of new speech categories from an L2, for instance), increasing the perceived similarity between certain L2 phones, in which case the learner’s perception does not become more native-like for that particular L2 contrast.

Amount of L2 experience is difficult to quantify, complicating the matter. Effects of experience have sometimes been confounded with effects of AOA (Factor 2) and duration of exposure or immersion - the length of time in which a learner has been in the L2 environment, often tallied as Length of Residency (LOR) in the L2 country. Adding to the confusion are idiosyncratic definitions of L2 learners as ‘novices’, ‘advanced’, or ‘experts’ in L2 acquisition. To exemplify, numerous studies with adult learners have considered these learners ‘experienced’ after anywhere between 1 to 10 years of L2 acquisition (Tsukada et al., 2005).

The interpretation of results from studies in which L2 experience, AOA, and LOR have been conflated is problematic, particularly when research participants are grouped together for an experiment based on AOA, but they differ widely in LOR and L1/L2 usage ratio (Piske, MacKay, & Flege, 2001). Usage ratio refers to the amount of an individual’s communication that takes place in the L1 and L2 respectively, and is related to degree of foreign-accentedness in L2 productions (Flege, Frieda, & Nozawa, 1997). A high continued L1 usage pattern is typically correlated with a greater degree of L1 accent and indicates that the non-native learner lacks exposure to or experience using the L2. By contrast, a limited continued L1 usage pattern typically suggests that the non-native learner is using the new language in many and varied settings; this is typically correlated with a lesser degree of L1 accent.

The importance of L1/L2 usage patterns has been highlighted by recent studies that suggest that the most significant changes in adult non-native perception are apparent after only 6-12 months of L2 immersion, increasing the importance of recruiting very tightly constrained participant groups (Flege & Liu, 2001; Aoyama et al., 2004; Tsukada et al., 2005). After this first year of immersion, there is only insignificant change in the direction of a more native-like perception of the non-native target contrasts. This differs markedly
from the patterns observed in child learners of an L2, who continue to show phonetic/phonological benefits from exposure to the non-native language, even after 12 months of L2 exposure (Aoyama et al., 2004; Tsukada et al., 2005).

1.2 Theoretical Explanations

A number of models have been formulated in the attempt to explain the effects of the interactions between the L1/L2 of a learner, AOA, and amount of L2 experience, on L2 speakers. Three models have been particularly influential: the Speech Learning Model (SLM: Flege, 1995), the Perceptual Assimilation Model (PAM: Best, 1994; 1995), and the Motor Theory of Speech Perception (MT: Liberman et al., 1967; Liberman & Mattingly, 1985). In Chapter 2, we present SLM, PAM, the extension of PAM to L2-perception (PAM-L2; Best & Tyler, 2007), and touch briefly on MT, in order to provide a novel outline of their respective strengths and limitations from a developmental perspective.
2 THEORIES ON SECOND LANGUAGE SPEECH PERCEPTION AND PRODUCTION

2.1 The Speech Learning Model

The Speech Learning Model (SLM: Flege, 1995, 2002) is based on the psychoacoustic approach to speech perception, i.e., the identification of acoustic properties of speech signals as the primitives of perception (Stevens, 2002). SLM mostly describes production of non-native phones in relatively experienced non-native speakers. The model assumes that the language-learning processes and mechanisms that children use when they acquire their L1 remain intact and accessible throughout life and are thus available to the adult L2 learner. The main application of SLM has been to explain the production behaviour of experienced L2 learners.

Even though SLM claims that the ability to learn the phones of a new language remains intact throughout life, it does not predict a complete lack of difference between the ultimate L2 proficiency of a child L2 learner and that of an adult L2 learner. According to SLM, other factors relevant to the acquisition of L2 vowels and consonants affect the formation of new long term memory categories for the non-native vowels and consonants that are similar to those in the native inventory. In effect, not all L2-only phones will be perceived as sufficiently dissimilar from those of the L1 to evade classification as equivalent. In such cases, an adult L2 learner will not easily be able to establish the new L2 categories. SLM argues that the relationship between the phones of the L1 and L2 may be described as falling along a continuum between new and identical.

SLM also claims that the phones of the L1 and the L2 coexist in a shared acoustic-phonetic space and will exert bidirectional influence: the establishment of new phonetic categories within the common phonetic space that frames the two languages will affect the already existing (L1) categories.

2.1.1 Interaction Between the L1 and L2

To account for the ways in which the phonetic categories of the L1 and the L2 interact, SLM proposes two possible types of L1 interaction with the non-native phones: category
assimilation and category dissimilation. Both are thought to influence the phonetic categories of the L1 as well as of the L2. Which of the two mechanisms is active in a given situation may depend on whether a new category has been formed for the new L2 phone or not, i.e., where it is located along the identical to new continuum.

Category assimilation is the perceptual inclusion of an L2 phone in an L1 category. It occurs when the learner is unable to establish a new and separate category for a new L2 phone because the mechanism of category formation has been hindered, for example, because the perceived similarity between the new phonetic categories and an already existing category is too great. In situations in which two non-native phones overlap, i.e. when two new L2 phones are perceived as belonging to the same L1 category, the result will most likely be poorer discrimination than when there is no overlap (Flege & Mackay, 2004; Guion et al., 2000).

However, category assimilation does not prevent all forms of phonetic learning from taking place in the learner, though SLM does not speculate about how this happens or what could drive it. Often, assimilation leads to subtle changes in the existing L1 category so that it will eventually present itself as a compromise between the L1 and L2 phones. This merged (and in some respects revised) category may cause the speaker to be perceived as foreign-accented even in his or her L1. The new category thus comes to be characterised by properties from both the L1 and the L2 categories.

Category dissimilation is the formation of a new phonetic category, unlike any other L1 category, to accommodate an L2 phone. Category dissimilation is based on the assumption that the native and non-native phonetic systems will interact (though how this interaction takes place is not explained) and allow for the formation of a new speech category that is found in the non-native language but not in the native language. Category dissimilation may have systemic effects on the phonetic inventory of the native and non-native languages as it works by increasing the phonetic distance among the different phonetic categories represented in the (L1-L2 combined) phonetic space. This may lead to a more extreme production of the L1 categories in proximity of the newly established L2 category, and for vowels, it may even lead to a general shift (frequently conceived of as raising or
lowering of the position of the vowel in the vowel space) in vowel quality in a group of vowels.

As SLM claims that the learning mechanisms in infant and child language acquisition remain intact into adulthood, the model relies on the development of the native phonetic inventory when explaining the correlation between increased AOA and stronger accent in a foreign language. A better developed L1 inventory provides a much less flexible frame in which the new L2 phones will be learned than does an L1 system that is still taking shape.

2.1.2 SLM and Perception

As SLM mainly concerns itself with production, it must be extended considerably in order to also account for perception. As introduced above, the central mechanisms of the SLM predictions for production are category dissimilation and category assimilation. We may propose that perception typically (i.e., in immersion acquisition with no specific pronunciation training) precedes production, as also suggested in Flege (1987), and Rochet (1995). Other research has, however, indicated that the relationship may be more complicated, such that improvements in one domain may translate into improvement in the other irrespective of the direction of improvement (Matthews, 1997; Sheldon & Strange, 1982), and even that perception and production may be independent of each other (Fujisaki, 1983; Paliwal, Lindsay, & Ainsworth, 1983). We also propose that perception, like production, is characterised by those same mechanisms of assimilation and dissimilation. That is, non-native phones may be perceived as similar or different to already existing native phonetic categories. Unlike the Perceptual Assimilation Model (PAM) presented below, SLM does not consider the relationship between the new non-native (L2) and the native (L1) phones, but only the relationship between:

1. The individual phone on a new-to-identical continuum (with respect to the L1 categories); and

2. The entire phone system of L1/L2,

as SLM predicts that the addition of new vowel may result in systemic shifts that maintain contrast between the different (new and old) categories.
According to SLM, we should therefore expect contrasting non-native vowels that are each perceived as different in relation to the existing native vowel categories to be easily discriminated, whereas those non-native vowels that are perceived as similar to the same native category are discriminated poorly.

2.2 The Perceptual Assimilation Model

2.2.1 Direct Realism and Articulatory Phonology

The Perceptual Assimilation Model (PAM; Best, 1994; 1995) is based on the ecological (or direct realist) viewpoint of Eleanor and James Gibson (Gibson, 1966; 1979) and Articulatory Phonology, typically associated with Carol Fowler, Louis Goldstein, and Catherine Browman (Goldstein & Fowler, 2003; Browman & Goldstein; 1989; 1990; 1992). In contrast to the psychoacoustic approach (Stevens, 2002) of SLM presented above, Articulatory Phonology (AP) posits that dynamically defined articulatory gestures are the most primitive units of both the phonetic realisations and the abstract phonological representations of speech (Browman & Goldstein, 1990).

In AP, and in PAM, articulatory gestures are organised in ‘gestural scores’ (Browman, Kelso, Rubin, & Saltzman, 1984; Saltzman, & Munhall, 1989; Turvey, Saltzman, & Schmidt, 1991) that involve coordinated movement of several articulators, such as the articulatory score for the English word *pan* presented in Figure 2.1 below.

The individual gestures in the gestural score overlap and are co-produced in a manner that explains, for instance, context-conditioned allophonic variance leading to the problem of one-to-many mappings between different phonetic realisations of the same phone and its underlying phonological category in traditional generative phonology.
Figure 2.1. Gestural score for the English word *pan* (adapted from Browman & Goldstein, 1995). The first 150 ms indicates the articulation of /p/, while /æ/ is produced between 100 and 300 ms into the utterance, and finally, /n/ is produced in the last 150 ms of the utterance. As is clear from the grey boxes in the figure, the activation periods for the specified articulators overlap. These overlaps indicate gestural coordination as well as coarticulation between the coordinative structures.

While AP has traditionally focused more on speech production than perception, PAM’s focus has traditionally been on perception and stresses the posit that *distal* events (i.e., objects and their events in the world) are perceived directly (i.e., extracted from multimodal incoming information) as well as entirely and satisfactorily by the general perceptual systems of an individual, regardless of which modality or modalities are carrying the information (visual, auditory, haptic) about the distal source and event. This kind of information thus differs from *proximal* information (i.e., sensory patterns on the modality-separated end organs, such as acoustic waveforms in the ear, optical patterns on the eyes, or skin pressure patterns on the fingers). AP thus differs fundamentally from the related Motor Theory of Speech Perception (see, for instance, Liberman & Mattingly, 1985), which assumes the gestural primitives of speech perception are perceived by a special speech encoder/decoder in the speaker. The view also differs from the psychoacoustic view which assumes that perceivers take in the proximal information from
the separate modalities, which they must then cognitively integrate and translate to match with abstract phonological structures. By emphasising the direct perception of the gestural primitives of speech, PAM and AP circumvent the problem of translating between the proximal psycho-acoustic properties of phones and their correspondence to abstract phonological categories, as well as the translation posited by a modular Motor Theory approach.

2.2.2 PAM

In contrast to SLM presented above, PAM mainly serves to explain the perception of non-native vowels and consonants by naive listeners, and to account for how experience with a native language (L1) changes the ways in which we perceive our native language as well as the vowels and consonants of unfamiliar languages.

PAM claims that individuals directly perceive the distal speech information in the world, as naturally produced by real speakers (including the perceiver’s own speech). In other words, we perceive the invariant properties of the world around us through integrated perceptual systems (auditory, visual and tactile) and active exploration and information-detection, and we do not rely on higher cognitive processes or abstract mental representations to interpret or extract meaning from the proximal input. As individuals develop greater familiarity with a certain type of information, they may further fine-tune their perceptual systems so that they, to a greater extent, focus only on information that is meaningful in their context or language, and they will be able to perceive higher-order information from their particular world niche, i.e., from their own ambient context. In the case of speech, the concept of higher-order invariants refers to the level of language-specific phonology, while differences in production of no phonological significance are considered to be phonetic.

For speech, the primitives of all phonological elements are the integrated articulatory gestures of the human vocal tract and talking face. Furthermore, as all speech sounds are produced within the human vocal tract, they are naturally defined by the possible articulatory gestures of the vocal tract and often characterised by some similarities across languages (Browman & Goldstein, 1989; 1990; 1992). Importantly, auditory and visual
speech information is not perceived separately, but unified by the meaningfully evolved integrated perceptual system of a human being. As a result, speech is perceived as amodal information about the object and events (in this case, the human vocal tract and its articulatory actions) that produced the dynamic changes in visual and auditory (and other types of) energy reaching the perceiver.

Furthermore, non-native speech segments (phones and articulatory gestures, respectively) are perceived on the basis of their similarities to and dissimilarities from the native segments (phones, gestures). Because both the human phonetic domain and native phonological space are defined by the shape of the vocal tract and its possible articulatory gestures, it is those specific distal properties that provide the frame for judgments of similarity between native and non-native speech sounds.

Like SLM, PAM assumes that perceptual learning is possible throughout the lifespan of an individual and that new phonetic categories may be established. However, it also assumes that this perceptual learning will be influenced by the learning history of the particular individual. That is, it assumes that the way in which someone perceives the L2 gestural information will reflect how that individual’s perceptual system has been fine-tuned to his or her L1. It follows from this that PAM must claim that there will be differences between the way in which infants, children and adults perceive both their native speech sounds and new, non-native sounds.

2.2.2.1 Perceptual Assimilation Patterns

PAM suggests several different ways in which non-native phones may be perceived relative to native phones. These are:

Assimilation to an existing native category, either as

1 a good example of that particular category,
2 an acceptable example of that category, or
3 a deviant example of that category;
Assimilation as an un-categorisable speech sound; or

No assimilation to any speech sound, for instance when a non-native phone is perceived as something other than speech, like a smacking or kissing sound of the lips.

Though not explicitly stated in PAM, good to deviant should be thought of as a continuum, similar to that of Flege’s identical to similar but short of actually including a new category (Best, personal communication).

We may of course ask whether there are indeed any non-assimilable vowel sounds, a question also raised by Strange et al. (1998). Due to the non-discrete nature of vowels, L2 learners may well be able to identify all vowels as vowel sounds, however unusual or strange in relation to the available L1 vowel categories. In fact, it is possible that the non-assimilable pattern may be the legacy of PAM’s first objects of study – click consonants – and not be relevant when studying vowels.

We may also ask whether there are any un-categorisable vowel sounds. While PAM, and its extension PAM-L2 (Best & Tyler, 2007) does not predict whether or not new phonetic (and phonological) categories will be established for initially un-categorisable phones – both scenarios are possible – PAM does claim that the ability to establish new phonetic and phonological categories remains intact even after the onset of puberty, and this learning may indeed happen as the perceptual systems attune to the initially un-categorisable phonetic categories of the new language.

2.2.2.2 Perceptual Discrimination of Non-Native Contrasts

On the basis of the assimilation patterns above, PAM suggests a number of ways in which minimally contrastive pairs of non-native phones may be perceived in relation to each other. These are:

1 Two-Category Assimilation (TC), where the two non-native phones are perceived as exemplars of two separate native categories. In this case, a non-native listener will find it very easy to discriminate between them.
2 *Category-Goodness Difference* (CG), where the two non-native phones are perceived as exemplars of one native category but are not equally good examples of this category. Here, discrimination is moderate to good, depending on the degree of perceived differences in goodness of fit.

3 *Single-Category Assimilation* (SC), where the two non-native phones are again perceived as exemplars of one native category but are perceived as equally good (or bad) examples of that category. Discrimination will be poor.

4 Both *Uncategorisable* (UU), where the two non-native phones both fall in between two or more phonetic categories of the native language. Discrimination will be between poor and very good, depending on the degree of similarity between the two sounds and their perceived similarities to the same or different subsets of native categories.

5 *Uncategorisable versus Categorisable* (UC), where one non-native phone is perceived as an exemplar of a native sound category and the other falls in between two or more native phonetic categories. Here, discrimination is expected to be very good because it involves a native category distinction (in-category vs. outside-of-that-category).

6 *Non-assimilable* (NA), where both non-native speech sounds fall outside of the native speech domain altogether and are perceived as non-speech sounds. Discrimination will be good or very good, depending on the degree of acoustic similarity between the two phones.

With respect to the discrimination patterns above, some patterns may be more likely to occur than others, and the least likely may be that of finding an example of a Single Category Assimilation with non-native vowels, that are not in the state of merging or near-merging and therefore pose a problem even to native speakers (see for instance, Bundgaard-Nielsen et al., 2008; Bundgaard-Nielsen et al., under revision). One study did however find such a contrast (Best et al., 1996; Best et al., 2003) as American English listeners have a clear SC assimilation pattern for Norwegian /l/ and /y/.
2.2.3 PAM and Production

Extending PAM to L2 production, we may again look to its direct realist assumptions that the directly perceived gestural information available from other speakers is also available from the listener him- or herself as a result of his or her own speech production experience. This leads to the general prediction that production will follow perception (see for instance, Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997), unless the speaker receives specific articulatory training, as the speaker will perceive his or her own production as well as the productions of others, and subsequently be able to finetune the gestures involved in particular speech acts. This expectation of perceptual change occurring before change to L2 production sets PAM apart from the Motor Theory of Speech Perception, which predicts the opposite pattern of improvement in perception happening alongside – or even after – improvements to L2 production (MT; see Section 2.2.3.1), with which PAM is often compared.

2.2.3.1 Motor Theory of Speech Perception and Production

The Motor Theory of Speech perception and production (Liberman, 1967; Liberman & Mattingly, 1985) was borne out of the observation that, while speech perception and speech acoustics are lawfully related, this relationship is not a one-to-one mapping between an acoustic signal and a given phoneme. Rather, the fundamentals of speech perception are the articulatory movements resulting from a set of neuro-motor commands that correspond to the intended phonemes, not the varying acoustic signal. In other words, while the acoustic signature and articulatory realisation of a given utterance may vary (for instance by speaker and context), the relationship between the neuro-motor command and the phoneme does not, and speech perception is the perceiving of this intended gesture.

Furthermore, speech perception and production are interpreted as two sides of one coin: such that speech perception is the identification of intended neuro-motor commands, speech production is the execution of these same neuro-motor commands. Both processes are suggested to be executed by a network of specialised brain modules handling for speech perception, and, as a reverse processing of perception, of speech production. While speech perception and production are also considered sides of the same coin by PAM and Direct Realism, the MT view differs from that of PAM/PAM-L2, primarily in two ways:
Firstly, MT relies on specialised brain modules to form and interpret the neuro-motor commands that underlie speech perception and production (PAM/PAM-L2 claims that speech perception and production is handled by general mechanisms and that the goal for both perception and production is the distal events), and secondly, MT ties speech perception and production to the same modules, suggesting that the two aspects develop in parallel.

### 2.3 SLM versus PAM: Differences Distilled

While the general predictions for L2 acquisition by SLM and PAM frequently overlap, there are several important differences between these two theoretical models. One difference lies in their different foci: SLM has traditionally focussed on production, while PAM has focussed on perception, and each theory must be extended to include both perception and production. Another equally important difference lies in their different theoretical underpinnings in psychoacoustics (SLM) and direct realism (PAM) that, while difficult to directly test, are essential for their further development. A third central difference, which will be briefly addressed in the analyses of the experiment presented in Chapter 5, is SLM’s focus on a single phonetic/phonological level, and PAM’s focus on a separation of the phonetic and phonological levels. In the PAM extension, PAM-L2 (see Section 2.4 below), this separation of the phonetic and phonological levels will yield testable hypotheses for what drives changes in L2 perception and production.

### 2.4 PAM-L2: New Inroads on Developmental Change in L2 Perception and Production

PAM-L2 (Best & Tyler, 2007) is an extension of PAM to account for acquisition-related changes in L2 perception and production. PAM-L2 assumes, as does PAM, that perceptual learning is possible at all ages but will be influenced by the entire language learning history of the individual.

It is argued that the division of the phonetic and phonological levels is essential to explain how L2 phones are acquired. According to PAM-L2, L2 phones are first assimilated into already existing L1 categories or dissimilated from existing categories and established as new categories first on a phonetic level (i.e., on a level where individual tokens of the same
vowel category can be heard as *qualitatively*, rather than categorically, different across different contexts or dialects). PAM-L2 further posits that continued L2 vocabulary acquisition forces the L2 learner to attune to the higher-order organisation of the L2 phones so that this assimilation/dissimilation process also takes place on a phonological level where phones are only distinguished on the basis of meaningful categorical differences and where non-categorical information such as information on dialect differences, and even foreign-accentedness, is irrelevant.

PAM-L2 further points to one mechanism that may influence the L2 perception and production of an individual by proposing that L2 vocabulary acquisition may “exert forceful linguistic pressure” (Best & Tyler, 2007, p. 32) on the learner to attune to articulatory, phonetic, and phonological differences in the L2 that have previously been ignored in the L1. PAM-L2 further suggests that improvement in L2 segmental perception may reach an asymptote very early in L2 learning, while the L2 vocabulary is still fairly small, and that a larger L2 vocabulary may curtail further phonetic learning, though seemingly paradoxical at first thought, because a larger and more rapidly growing vocabulary forces the learner to settle on a particular L1-influenced version of the L2 phonology in the perception and production of the L2 phones *in words*.

In the following chapter, the PAM-L2 suggestion that L2 vocabulary development may curtail (or lead to an asymptote in) the acquisition of L2 phones is further revised and extended, supported by research from the fields of L1 acquisition and cross-dialect perception in children. Finally, general predictions for developmental change in the production of non-native vowels are presented from within the theoretical frameworks of SLM and PAM/PAM-L2 and the Vocab Model - a novel theoretical model, based in part on key ideas from PAM-L2, which is developed as a core concept in this thesis. This novel framework drives the experimental studies and their interpretations, presented in Chapters 5, 6, and 7 of this thesis.
3 THE ROLE OF L2 VOCABULARY EXPANSION IN L2 REPHONOLOGISATION

This chapter presents the Vocabulary-Tuning Model of L2 Rephonologisation, or Vocab Model for short, a revised and extended version of the PAM-L2 proposition that L2 vocabulary acquisition may “exert forceful linguistic pressure” (Best & Tyler, 2007, p. 32) on the learner to attune to articulatory, phonetic, and phonological differences in the L2.

The chapter also presents evidence for the role of vocabulary expansion in L1 perception and production. Finally, we present a number of predictions for the development of L2 vocabulary and L2 vowel perception and production in adult L2 learners.

3.1 The Vocabulary-Tuning Model of L2 Rephonologisation

The Vocab Model presented here is based on PAM/PAM-L2 (see Chapter 2), yet revises and extends the PAM-L2 proposition that L2 vocabulary expansion may curtail further L2 reattunement and rephonologisation by forcing the L2 learner to fossilise his or her perception of the L2 phones. Rather than focussing on the inhibiting effect of a large L2 vocabulary, the Vocab Model thus focuses on the facilitating effect of early L2 vocabulary expansion on L2-perceptual attunement and rephonologisation.

The concept of reattunement is best thought of as a re-setting of the perceptual systems of the learner so that she becomes sensitive to the phonetic variations in her L2, while rephonologisation is best thought of as the establishment of an L2 phonological system, by modification or addition to the L1 phonological system of the learner. This process is dependent on the learner correctly tuning into those L2 phonetic differences that are meaningful, i.e., used to differentiate meaning in the L2, but does not imply that reattunement and rephonologisation necessarily leads to the formation of a complete new set of L2 categories. Rather, it is suggested that the learner settles on a phonological system for the L2, perhaps by stretching the L1 inventory as far as possible, that is, consistent with a principle of phonetic economy, and perhaps by forming new L2 categories. This also entails that the learner is unlikely to be able to rephonologise to a degree where the L2 phonology becomes entirely nativelike.
The Vocab Model suggests that early improvements (i.e., in the first 6 months of acquisition) in L2 segmental perception in late learners may be positively associated with, or even driven by, an expanding L2 vocabulary. Like PAM/PAM-L2, the Vocab Model is based on the assumption that basic language-learning mechanisms remain available over the entire lifespan, though the language learning behaviour (L1+ acquisition) of an individual will reflect his or her language learning history, resulting in the amply documented L1 accent in L2 perception and production (see Chapter 1). The Vocab Model is based on the additional assumption that L2 segmental acquisition (perceptual attunement to an L2 and subsequent L2-rephonologisation) unfolds in a similar fashion to L1 attunement and phonologisation. In particular, it is proposed that L2 attunement and rephonologisation is vocabulary-associated or -driven in a way similar to the emergence of L1 phonological functions that appear to be associated with, or driven by, the so-called vocabulary explosion in children around 18+ months-old (e.g., Benedict, 1979; Metsala & Walley, 1998; Nazzi & Bertoncini, 2003).

In practice, this means that early L2 phone perception is based on phonetic similarities between the L2 phones and L1 categories (See PAM-L2 in Chapter 2), and that early L2 words (like early L1 words; see Section 3.2 below) are first perceived on the basis of their phonetic similarities to and differences from the phones of the L1 of the learner. That is, the learner perceives only the surface structure of the L2 (on the basis of L1-L2 similarities/differences) as they have yet to attune to the higher-order phonological organisation of the L2. Successful L2 phonological perception is suggested to arise from the acquisition of an increasingly complex vocabulary, facilitating the attunement to the higher-order phonological structures in the L2 through increased exposure to the invariant phonological properties of that particular language. It is furthermore likely that this lexically-based pressure to rephonologise may be particularly strong in densely populated lexical neighbourhoods with many minimal pairs (words that differ in just one phonetic feature of a single phoneme, such as *pat* vs. *cat*), or in situations when several contrastive L2 phones are assimilated into the same L1 category causing extraneous activation of competing lexical items (Weber & Cutler, 2004; Broersma, 2005; Cutler & Otake, 2004; Cutler, Weber, Otake, 2006). For example, in an eye-tracking experiment, Weber & Cutler
(2004) found that Dutch learners of English will behave in a manner that indicates that the presentation of English words with /æ/, such as *panda*, will result in activation of competing lexical items containing /ɛ/, such as *pencil* (but not the other way around), possibly because the learner’s native phonology does contain an /ɛ/ but does not contain an /æ/. In such situations, the effort spent on rephonologisation, i.e., the formation of a new /æ/ category, will have a significant pay-off - increased awareness of the higher-order invariant properties governing that language will be of great help to an L2 user, who will otherwise confuse phonetically similar (from an L1-perspective), but phonologically different L2 phones (i.e., the conflation of the English vowels in *hot* and *hut* by Danish speakers, see Bohn & Bundgaard-Nielsen [2008] or the perception of French rounded vs. unrounded front vowels /i:/ and /y/ as English /i:/ by American English listeners, see for instance, Levy & Strange, 2008), as well as phonetically different, but phonologically identical L2 phones (i.e., the contexts-conditioned allophonic variation between extremely fronted American English [u:] following an alveolar consonants, and a much less fronted [u:] following for instance bilabial consonants), and therefore struggle to follow normal conversation.

### 3.2 Evidence for the Role of L1 Vocabulary Expansion in the Acquisition of L1 Phonology

#### 3.2.1 Early Acquisition of Phonetic Categories in the L1

Infants attend to phonetic differences from birth and are able to discriminate almost all speech contrasts well (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Aslin, Pisoni, Hennessy, & Perey, 1981), until their perceptual systems attune to those contrasts meaningful in the surrounding language. This happens at approximately 5-6 months of age for vowels and at 8-10 months for consonants (Werker & Tees, 1984; Eimas, Miller, & Jusczyk, 1987; Best, McRoberts, Sithole, 1988; Werker & Lalonde, 1988; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Best 1994; Kuhl, 2000; 2004; Best & McRoberts, 2003; Polka & Bohn, 2003; Polka, Bohn, & Molnar, 2005; Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006; Bohn, 2007; Vouloumanos & Werker, 2007).
Infants’ speech discrimination ability at just 6 months-of-age predicts their vocabulary size in their second year of life (Tsao, Liu & Kuhl, 2004), suggesting that successful phonetic attunement facilitates phonologisation and word-learning\(^1\). However, this phonetic attunement does not extend into early word learning proper. Recent evidence from toddler dialect perception seems to indicate that although toddlers may be able to perceive fine-grained phonetic differences in non-lexical discrimination tasks, they are not able to apply this to word-learning tasks (Stager & Werker, 1997). Complimentary studies of toddler word recognition in native and nonnative dialects (Best, Tyler, Gooding, Orlando, & Quann, in press; Best, Tyler, Kitamura, Notley, & Bundgaard-Nielsen, 2008) found that toddlers fail to recognise familiar words spoken in an unfamiliar dialect, indicating that they are unable to ‘hear past’ the different surface phonetic realisation of the familiar word in the nonnative dialect to identify the shared underlying phonological form. This is taken to suggest that when toddlers learn their first words, these words are learned in their dialect-specific phonetic form before the toddlers have established a native phonology.

### 3.2.2 Acquisition of Phonological Categories in the L1

Toddlers first begin to pay attention to phonetic information in well-known words in a familiar dialect from 14 months. At that age, /dog/ is better recognised as a familiar word than the minimally mispronounced non-word /tog/ by the toddlers (Swingley & Aslin, 2000; 2002). Similarly, at that age, toddlers will look longer to a picture of a doll when presented with the (American English) word *doll*, than to the same picture when presented with the minimally different word *ball* paired with a picture of a doll (Fennell & Werker, 2003). This sensitivity to even very subtle minimal pair differences, however, does not extend to *novel* words until they are between 18- to 24-months-old (Werker, Fennell, Corcoran, & Stager, 2002). In fact, this study found that young infants of 14 months failed to learn to associate two phonetically similar words (*bih* and *dih*) with two different objects, while infants of 20 months-of-age successfully learned the new labels, and that 17

---

\(^1\) Reciprocally, vocabulary size has been found to predict the speech perception abilities of children between the ages of 2-6 years (Metsala, 1999), highlighting the continued interdependency of vocabulary development and speech perception.
month-old toddlers performed at intermediate level. Importantly, the performance by the toddlers in the two younger age-groups was predicted by their vocabulary size.

These findings have been interpreted to suggest that the phonetic similarity between well-known words and their non-word and real word minimal pairs is not adequately mapped onto different phonological structures in young toddlers. Similarly, cross-dialect word recognition studies with 15- and 19-month-old toddlers show that 19-month-old infants successfully recognise familiar words in an unfamiliar dialect, but that 15-month-olds do not (Best, Tyler, Gooding, Orlando, & Quann, in press). This may be because the latter group are unable to map the different phonetic realisation of the otherwise familiar word onto their native phonological categories. In a follow-up study with 14- and 19-month-old toddlers, the younger toddlers, who all had a limited expressive vocabulary under 25 words, did not even recognise familiar words in a familiar dialect (but from an unfamiliar speaker). On the other hand, the older children, who all had larger expressive vocabularies over 50 words, recognised the familiar words in both familiar and unfamiliar dialects (Best, Tyler, Kitamura, Notley, & Bundgaard-Nielsen, 2008). Again, this indicates that the early vocabulary of a child is phonetically-specific, and that knowledge of the underlying phonological system (and therefore of which phonetic differences are meaningful, and which are not) only develops with a larger vocabulary.

To summarise, these findings are interpreted as evidence that although infants are highly skilled in detecting phonetic details important in their L1 before their first birthday. They, and young toddlers, however, do not have phonological constancy, i.e., access to the underlying phonological structure of a word which allows the identification of differently produced instances of that word as ‘the same’, until the time of the vocabulary explosion, at approximately 18 months. In other words, prior to phonologisation of the L1 around 18-19 months, phones are stored and recognised in their detailed phonetic dialect-specific forms, rather than as abstract phonological entities. Thereby, words are simply recognised on the basis of their dialect-specific phonetic properties, rather than higher-order phonological properties, until the child has acquired enough words to realize they must ignore unimportant phonetic variation and attend only to more abstract phonological differences. Together, this early attunement to phonetic detail and later attunement to
phonological structure indicate that pressures from an increasingly complex vocabulary may indeed drive the attunement to the higher-order phonological structures in the language, as has also been indicated by for instance Metsala & Walley (1993), which will be reviewed in detail below.

By extension, under the assumption that the language learning mechanisms of an individual remain intact over the lifespan, parallel acquisition mechanisms may be observed for L2 learners. The Vocab Model thus proposes that early L2 perception is based on phonetic similarity between L2 words and phones and L1 categories, such that L2 phones will be perceived as instances of the most phonetically similar L1 phone (Best & Tyler, 2007). Only through the acquisition of a larger set of L2 vocabulary items will the L2 learner be familiar with a large enough L2 sample to be able to tune into the higher-order L2 phonological systems and ignore irrelevant phonetic differences. This move from phonetic to a phonological level of organisation may lead to shifts in L2-L1 assimilation patterns, or even to the formation of new L2 phonetic/phonological categories.

3.2.3 Evidence for the Role of L1 Vocabulary Expansion in L1 Segmental Production

Vocabulary size can be measured in two ways; as the receptive or productive vocabulary of the child (or adult). The receptive vocabulary contains all words a child understands (but may not produce), while the productive vocabulary contains only those words that the child uses spontaneously. Word comprehension (receptive vocabulary) is acquired quickly, from around eight months-of-age, while the productive vocabulary develops more slowly, from around 12 months (Benedict, 1979; Crain & Lillo-Martin, 1999).

Changes to the receptive vocabulary are more difficult to measure than the productive lexicon (Hamilton, Plunkett, & Schafer, 2000). This is particularly the case beyond the first birthday (Griffiths, 1986) when the vocabulary expands more rapidly. It is suggested, however, that the spurt in productive vocabulary in children at around 18 months follows a prior spurt in the receptive vocabulary (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998), which appears to occur approximately 5 months earlier, that is, around 12-14 months of age (Benedict, 1979). It is noteworthy, that this is the same age as the younger
toddlers tested in the cross-dialect studies reported above (Best, Tyler, Gooding, Orlando, & Quann, in press; Best, Tyler, Kitamura, Notley, & Bundgaard-Nielsen, 2008), and who fail to recognise familiar words in an unfamiliar dialect – and who may even struggle to recognise those familiar words in their own dialect if they have a low productive vocabulary. Together, this indicates that improvement in L1 production follows the perceptual attunement in the L1 on which the process of phonologisation rests, with the following caveats: Firstly, some of the delay in expressive vocabulary development is due to cognitive limitations as well as immature motor control of the speech organs (Reznick & Goldfield, 1992; Harris & Chasin, 1999). Secondly, the size of the expressive vocabulary is difficult to assess in very young children with highly idiosyncratic productions.

As demonstrated above, a child’s L1 segmental perception skills predict L1 receptive vocabulary development, which in turn is related to the size of the child’s L1 productive vocabulary. Research into the role of L1 productive vocabulary size and L1 non-word repetition skills in older children also indicates a link between vocabulary size and production proficiency, bringing the connection from segmental perception over vocabulary development to segmental production, full circle. In one study, children with smaller productive vocabularies articulated low-frequency (rarely occurring in the children’s L1) two-phoneme combinations less accurately and with slower speaking-rate than high-frequency (frequently occurring in the L1) combinations (Edwards, Beckman, & Munson, 2004). A similar study also found that late talkers (late vocabulary onset and low vocabulary) between 18-33 months-of-age produce complex native consonant clusters in a manner less intelligible than typically developing toddlers (Thal, Oroz, & McGraw, 1995). In yet another study, children with larger receptive vocabularies performed better than children with smaller receptive vocabularies on non-word repetition tasks (Bowey, 2001). Together, these findings suggest that L1 production accuracy is predicted by vocabulary size, which in turn, as demonstrated above, is predicted by the perceptual abilities of the child.

Extending this integrative interpretation of L1 phonological development to late/adult L2 acquisition, we would expect that learners with poor ability to categorise and discriminate L2 contrasts will be those with smaller L2 vocabularies and poorer L2 production.
3.2.4 L2 Vocabulary Development and L2 Segmental Production

While the effect of L2 vocabulary size on L2 segmental speech perception and production has yet to be examined directly, a limited number of studies have observed an asymmetry in the sizes of the receptive and productive vocabularies for adult L2 learners (Fan, 2000; Laufer, 1998; Laufer & Paribakht, 1998; Webb, 2008). These findings could be viewed as consistent with the prediction of PAM/PAM-L2 and the Vocab Model that increases in L2 productive vocabulary and production accuracy in adult L2 acquisition follow increases in perceptive vocabulary size and improved perceptual realignment and rephonologisation, as in L1 acquisition. The similarities between child L1 acquisition and late/adult L2 acquisition must however be qualified as the relationship between perceptual abilities and production in L1-learning toddlers and late/adult L2 learners differ in a number of ways.

Firstly, any maturational processes in the brain that may influence the timing relationship between perception and production in young children are long completed in L2 acquisition in adolescence and adulthood. Secondly, late/adult L2 learners have already firmly established a pattern of native speech production behaviour. These two differences likely affect the timing between improvements in L2 perception and L2 production in two ways. On the one hand, the discrepancy in perception and production may be smaller for late/adult L2 learners than it is for young L1 learners due to differences in cognitive and motor skills. On the other hand, the necessity of learning new L2 motor patterns in L2 speech may counteract this logically-possible advantage of L2 learners over young L1 learners.

Despite possible cognitive advantages of adult L2 learners over child L1 learners, L2 learners still frequently misproduce (as well as misperceive) L2 phones. As outlined in Chapter 1, it is well-established that the nature of these misproductions (and misperceptions) is related both to the L1-L2 minimal segmental pair in question, and to the amount of experience that a learner has had with the L2. What is less clear is how and why experience with the L2 leads to improved L2 production. Based on evidence of the role of vocabulary expansion in the acquisition of L1 phonology, we offer the Vocab Model presented here as a possible account for learning-related changes not just to L2 perception, but also to L2 production.
According to the Vocab Model, perceptual reattunement and rephonologisation begin before evidence of this process finds its way to the L2 output of a learner. Despite cognitive advantages over L1-learning toddlers, the learned motor patterns of late/adult L2 learners interfere with L2 production, such that perceptual reattunement only forms a first step in the direction of improved L2 production; after tuning in to the L2 phonology, the learner must still fine-tune his/her motor patterns to better approximate those of native speakers of the L2. Improvement in L2 production therefore is expected to most often suffer a delay relative to improvement in L2 perception. Overall, this prediction is in line with the predictions of both PAM/PAM-L2, but not MT, as perception and production would be expected to improve simultaneously, or even with a production lead, within that theoretical framework.

3.3 Predictions for the Role of Vocabulary in Segmental L2 Perception and Production

3.3.1 General Predictions for L2 Perception

Based on the theoretical frameworks of SLM and PAM in combination (see Chapter 2) it is possible to generate a number of testable hypotheses about the role of L2 vocabulary expansion in L2 segmental perception and production. Recalling the three factors influencing L2 perception and production (see Chapter 1), it is predicted that:

1. The L1 phonetic and/or phonological inventory will affect L2 segmental perception and production, such that L2 phones will be perceived on a continuum from identical over similar to new (identical -> similar -> new) in relation to the inventory of the L1. In particular, SLM predicts this continuum to reflect only phonetic similarities, not phonological similarities, while PAM predicts that this continuum will firstly reflect phonetic similarities, but will also, at least for some learners, move on to reflect detection of higher-order phonological similarities;

2. A large L1 phonetic/phonological inventory aids the acquisition of a large L2 inventory, by providing the learner with more phonetic/phonological categories to relate the novel phones to;
3 Age of Acquisition of the L2 will affect the degree of difficulty experienced by the learner, such that younger learners will generally perform better than older learners. However, this is due not to a critical period effect, but rather because younger learners’ native phonetic/phonological system has not yet become firmly ‘tuned’ and therefore remains more malleable than that of older learners²; and

4 Increased Experience with and use of the L2 will improve L2 segmental perception and production, such that L2 users with more L2 experience and less continued L1 use will generally outperform L2 users with less experience and more continued L1 use.

3.3.2 Vocab Model Predictions for the Role of L2 Vocabulary Expansion in L2 Perception

Considering the role of vocabulary acquisition in the development of phonetic and phonological categories in L1 acquisition, and the PAM-L2-founded hypothesis that L2 segmental reattunement and rephonologisation is somewhat analogous to L1 attunement and phonologisation, it is possible to generate more specific predictions for the effect of L2 vocabulary expansion on L2 segmental perception and production in immersion acquisition.

We predict that L2 vocabulary size is positively associated with L2 segmental perception such that:

1 The perceptual assimilation of L2 phones to L1 categories will be more consistent in L2 learners with large L2 vocabularies than in learners with smaller L2 vocabularies. Consistency refers to a larger number of L2 phones successfully assimilated to L1 categories and an average higher percentage of

² This hypothesis will not be tested in the present thesis. Rather, we will make the assumption those adult speakers who do not differ in age, are very largely comparable in terms of their native-language perceptual tuning and the state of their L1 phonological system.
each L2 phone assimilated onto L1 categories, reflecting less perceptual confusion. This pattern of more consistent perceptual assimilation is hypothesised to arise from a greater ability to attend to only phonologically relevant differences of L2 learners with larger L2 vocabularies than learners with smaller L2 vocabularies.

2 L2 learners with larger L2 vocabularies will perform better than learners with smaller vocabularies on tests of their discrimination of L2 contrasts. Again this improved performance arises from a greater degree of re-attunement and rephonologisation, allowing the high L2 vocabulary learners to ignore phonologically irrelevant phonetic variations in the L2 input and attend only to phonological differences.

We are also able to make specific predictions for vocabulary-driven change in L2 perception for the PAM and PAM-L2 contrast types (SC, GC, TC, UC, and UU; see Chapter 2). We predict that:

Increased L2 vocabulary size affects the different L2 contrast types differently, such that:

1 Discrimination of Single Category (SC) contrasts will remain problematic to the L2 learner. This contrast type, however unusual for vowels, is based on the perception of (near-) absolute equivalence between the two phones.

2 Discrimination of Category Goodness (CG) contrasts will improve with increased vocabulary as the L2 learner continues to re-attune to the phonologically significant differences between the two phones (i.e., not only to their phonetic differences). Depending on the size of the L1 phonological inventory, the learner may eventually perceive the less typical of the two phones as belonging to a different L1 category than they originally perceived, or may even form a new L2 category for this phone;

---

3 See Appendix 1 for a further discussion of SC contrasts in the light of diachronic vowel change.
3 Discrimination of Two Category (TC) contrasts is not likely to improve significantly as the vocabulary develops, as the L2 learner already successfully discriminates this type of contrast and experiences no pressure to attune and rephonologise further. Thus, phonetic discrepancies between L1 and L2 phones for TC contrasts may persist and be evident in a poorer pronunciation of these L2 phones that would be expected, for instance, for the less nativelike of two phones in a CG assimilation;

4 Discrimination of Uncategorised-Categorised (UC) contrasts will most likely improve by increased vocabulary-driven attunement. As the L2 learner more successfully tunes out of the phonetic discrepancies, and into the phonological characteristics of the two phones, the uncategorised phone may be assimilated into an already existing L1 category or form a novel L2 category. The opposite pattern may, however, also be observed, if the Uncategorised phone is eventually assimilated into the same category as the Categorised phone in that particular pair.

5 Changes over time in discrimination of Uncategorised-Uncategorised (UU) L2 contrasts is more complicated to predict, as the phonetic similarity may be great or small between two such L2 phones, yielding initial discriminability ranging between excellent (for L2 phones that are similar to separate subsets of L1 categories) to very poor (for phones that are similar to the same subset of L1 categories). Depending on the degree of phonetic similarity of the two phones to existing L1 categories, increased vocabulary-driven attunement may lead to successful assimilation of these phones into specific L1 categories (rather than being distributed across two or more L1 categories), or to the formation of novel L2 categories. Indeed, the phones may eventually be perceived as a SC, CG, or UC contrast and discriminated accordingly.

Finally, based on the Vocab Model’s rationale, as outlined in this chapter, we make predictions for acquisition-related changes in L2 production relative to L2 perception, such that;
1 L2 learners with larger L2 vocabularies will produce L2 phones in a manner that is more intelligible to native listeners than learners with smaller L2 vocabularies; and

2 The improvement in the perception of non-native phones will be evident earlier during the L2 learning process than will improvement in the production of non-native phones.

In the following Chapter 4, we outline a series of Experiments to test the above PAM-L2 and Vocab Model-based research hypotheses.
4 OUTLINE OF RESEARCH PROJECT

The following series of experiments addresses the general predictions for L2 perception and production put forward by SLM and PAM/PAM-L2 (see Chapter 2), with particular focus on the Vocab Model predictions for the role of vocabulary development in L2 segmental perception and production (see Chapter 3).

To test the prediction that L2 vocabulary development plays a significant role in L2 perception and production, we examine the role of L2 Australian English (AusE) vocabulary size in the perception of L2 AusE vowels by adult Japanese learners of AusE who have recently arrived in Australia with the explicit purpose of acquiring conversational English.

Before describing the series of experiments undertaken, however, a brief introduction of the nature of vowels in general, and the vowel systems of Australian English and Japanese in particular, will be given. It highlights a number of important vowel features and differences in the vowel inventories of Japanese and English that make Japanese and AusE a highly suitable L1-L2 pair for testing our predictions.

4.1 The Nature of Vowels

4.1.1 What are the Characteristics of Vowels?

Vowels, such as the phonetic segments we associate, in English, with the letters a, e, i, o, and u are produced when air passes through the vibrating glottis (in most cases) and through the oral cavity (and the nasal passage for some vowels in some languages) without meeting any obstructions, i.e., without the lips, teeth, and/or tongue blocking the passing air. The particular acoustic pattern associated with each vowel is determined by the configuration of the oral cavity while the air passes through the nasal and/or oral cavity. A change in the shape of the oral cavity such as is caused by a movement of the tongue, raising or lowering of the jaw, or protrusion of the lips, results in a gradient (i.e., graded or gradual) change to the quality of the vowel.
The production of vowels differs from the production of consonants\textsuperscript{4}: Vowels are longer in duration, subject to much greater variability in production, and will change considerably as a result of co-articulation with preceding and following speech segments and as a result of changes in speaking rate. These characteristics make vowels difficult to study, in comparison to consonants, but at the same time a very fruitful area of research into human speech production and perception. Indeed, it could be argued that vowels are particularly interesting as they form a closely connected system among themselves (and share similarities and patterns of change over time) in ways that differ from consonants: While vowels form one large overarching system, consonants form smaller subsystems (nasals, fricatives, stops, etc.) which are much less connected and can change independently.

The quality of a vowel is generally understood to be determined by changes in vocal tract resonances, and formally, there are a number of parameters of interest and importance to us when we are investigating the production of vowels: tongue height, tongue position, and lip rounding, which in turn will be discussed below.

\textsuperscript{4} For an in-depth discussion of the articulatory and acoustic characteristics of consonants, see Ladefoged & Maddieson (1984), Ladefoged (1993), or Clark & Yallop (1995).
4.1.1.1 Tongue Height

The parameter of tongue height is used to describe how close to or distant from the roof of the mouth the tongue body is during the production of a specific vowel. Some vowels such as the /i/ sound in beat and the /u/ in boot are produced with the tongue body very close to the roof of the mouth (“high vowels”), whereas vowels such as the /æ/ in bat and the /o/ in British English hot are produced with the tongue very far down from the roof, or hard palate (“low vowels”). We control tongue height by raising our tongue body toward the palate as well as by opening and closing the jaw. In research, vowels are often visualised through the use of spectrograms. A spectrogram is a mathematically generated image of the acoustic characteristics of a given speech sound, that is, the spectral or frequency characteristics of the complex speech waveform as they unfold over time during speech production. In a spectrogram, higher spectral energy is shown as darker coloration. Consequently, when vowels are viewed on a spectrogram, the spectrogram will show a number of horizontal bands of high energy within restricted frequency ranges called formants, and for acoustic reasons, tongue height is generally associated with the first high-energy band from the bottom (other than the band of high frequency energy that indicates the fundamental frequency, or F0, of the speaker’s voice, which averages 125Hz for men and 200Hz for women [Yang, 1996]). Below, in Figure 4.2, is an example of a spectrogram of the vowel /æ/ as in the English word hat. The indicated band of energy is referred to as the first formant, or F1. A good example of two vowels mainly differentiated by F1 is the vowels in the English words beat and bet.
4.1.1.2 Tongue Position

Tongue position is used to describe the positioning of the tongue body relative to the roof of the mouth, i.e., which part of the tongue body is “highest” or closest to the hard palate. In the previous examples, the /l/ in beat is produced with the tip or front of the tongue being closest to the roof of the mouth. In boot, with the vowel /u/, the back of the tongue is closest to the roof of the mouth. When describing tongue position, or place of articulation, we use a frontness to backness continuum. Any sound produced with the tongue closest to the roof of the mouth close to the front teeth or alveolar ridge will be described as a front-vowel and any sound produce with the tongue highest at the back of the oral cavity will be described as a back-vowel. On a spectrogram, the frontness to backness dimension corresponds with the second formant, i.e., the second band of high frequency sound waves, also known as the second formant (F2), as indicated by arrow in Figure 4.2.

4.1.1.3 Lip-Rounding

Lip-rounding refers to the aperture (distance between upper and lower lip opening) and/or protrusion of the lips when producing a specific vowel sound. The lips may be unrounded, or relatively open, as in the articulation of the words beat or bat, or they may be rounded (closer together or protruded), as in the words boot or butcher. Lip rounding is often associated with the spectral acoustic properties of the third band of high frequency energy in a spectrogram, the third formant (F3). F3 may also be conceived of as arising from vocal tract lengthening and/or radiation characteristics at the lips (a narrowed aperture, thus
reducing acoustic radiation from the lips, regardless of how the reduction is achieved, e.g., by lip pursing and protrusion or by simple lip narrowing). The F3 is the next principal band of high-frequency energy and is influenced by the first two formants, just as these two formants are influenced by F3, as the vocal tract is best seen as a ‘serial resonator’, consisting of various time- and location-varying ‘chambers’ (Ladefoged, 1993; Stevens, 2002).

4.1.1.4 Vowel Visualisation

To visualise the vowel system of any given language in spectral characteristics, the F1 (on the y-axis) and F2 (on the x-axis) values of the individual vowel sounds are often used to generate a so-called vowel chart such as shown in Figure 4.3 below. This vowel chart corresponds with the articulatory descriptions of the vowels on the high-low, front-back continua, though it of course fails to capture the temporal dimension of vowels, as well as contributions to vowels from for instance F3 as well as the dynamic properties of formants.

![Figure 4.3. The Australian English monophthongs.](image)
4.1.1.5 Vowel Duration

Vowels can also be described in terms of relative duration, a dimension which the F1/F2 vowel chart fails to capture (just as the space also fails to indicate the contribution of F3 to vowel systems such as French, German, Danish etc., that have front rounding distinctions, such as the Danish words *bide* (to bite) and *byde* (to offer). For some languages, such as Japanese, vowel duration is traditionally understood to be the distinguishing feature between the vowels in a so-called long-short vowel pair.

For some languages, such as English, the long-short distinction is largely irrelevant, though there are exceptions in Australian English (AusE). Nevertheless, for English (and for instance German and Danish), it is more meaningful to distinguish between tense and lax vowels. The tense vowels of English are /i, ɪ, ɔ, o, u/, traditionally thought of as the long English vowels, and sometimes embellished with the /l/ that is used in phonetic transcription to indicate a long vowel sound. The lax vowels of English are /ɪ, ɛ, ə, ʌ, u, ø/. Tense and lax vowels however also differ on a number of parameters beyond duration; they also differ in spectral properties, reflecting a difference in articulatory positioning and/or tenseness of the tongue body (or root) during production, and in terms of phonotactics, as tense vowels may occur in both open and closed syllables, while lax vowels must be followed by a consonant.

4.1.1.6 Vowel Dynamics

Vowels are not always produced with the tongue and jaw staying in the same configuration throughout the articulation. Some movement of either tongue, jaw or both (and sometimes even the lips) may occur between the onset of the vowel and the offset\(^5\). When the vowel quality changes during production of the vowel, it is referred to as vowel dynamics. In some languages, such as English, even those vowel sounds classified as monophthongs,  

---

\(^5\) Indeed, listeners may be able to correctly identify vowels on the basis of only onglides and offglides, when the vowel nucleus has been edited out (“silent center” vowels: Strange et al., 1983; Hillenbrand & Gayvert, 1993a, 1993b; Hillenbrand & Nearey, 1999), indicating that articulatory transitions from co-articulation with preceding and following consonants may greatly influence the perception of vowels.
i.e., consisting of just one vowel sound, may be characterised by some change in the articulator position, and thus in acoustic quality during production. In other languages, such as Japanese, there is very little change in tongue and jaw position during production of any vowel, and therefore more purely monophthongal vowels.

In contrast to most monophthongs, certain other vowels are characterised by rather marked changes in tongue height and position as well as jaw (and sometimes lip) movement during production. These vowels are, in a sense, a combination of two separate monophthongs, and are thus referred to as diphthongs. An example of a diphthong is the vowel /æʊə/ sound in the English word *house*. Diphthongs are produced when the articulators first form one vowel, and then glide toward the position of another vowel which is then maintained, at least briefly, in the second portion of the diphthong articulation. The duration of a diphthong in Australian English is, for example, approximately 167% of the duration of a (short) monophthong (Cox, 1999). This may be seen as evidence that the two vowel sounds of the diphthong have indeed been combined into one unit and are no longer completely separate vowels, as is the case in Japanese. In Japanese, bi-moraic combinations remain two separate vowels: the durational relationship between Japanese single mora and bi-moraic vowels is typically between 1:2.4 and 1:3.2 in careful speech (Han, 1962; Tsukada, 1999). In citation form, Japanese /i/ and /iː/ have even been found to be respectively 60ms and 190ms long – a durational difference of more than 300% (Tsukada, 1999).

4.2 The Vowel Inventories of Japanese and Australian English

4.2.1 Australian English

Australian English (AusE) is a non-agglutinative language with a rich phonological system, stress-foot rhythmic structure, with a primarily trochaic stress-patterning. AusE has Central British English origins (Yallop, 2003) and is divided into three sociolects: Broad, General and Cultivated, with Cultivated AusE retaining more British English features than the General and Broad varieties (Mitchell & Delbridge, 1965). AusE is a non-rhotic variety of English due to the loss of overt pronunciations of the post-vocalic /tr/ in hear, hair, hard, horde, and heard.
The AusE monophthong inventory consists of /ɪ, ɪ, ɪ, ə, ə, ə, ə, ə, ʊ, ʊ, ɔ/ (Cox, 2006). The AusE vowels, even the monophthongs, are all characterised by some degree of spectral change during production. Exceptions to the observation that AusE vowels typically differ in spectral quality, are the vowel pairs /ɪ-/ɪ/ and /ɛ-/ɛ:/ which are mainly differentiated on the basis of duration (Cox, 2006).

The AusE diphthongs are /æʊ, æi, əe, oi, eu, əɛo/. They are typically divided into two subgroups: non-centring /æi, əe, oi, eu, əɛo/ and centring /ɪə/ (Cox, 2006).

4.2.2 Japanese

Japanese is an agglutinative language with a very limited phonological system and a moraic rhythmic structure. The only permissible syllable structures are V or CV. Japanese has a vowel inventory of just five unique vowels in long-short (phonologically distinct) pairs /i, i, e, e, a, a, ʊ, ʊ, o, o/ (IPA, 1999). The Japanese vowels are relatively monophthongal (“pure”). There are, however, systematic spectral differences between the long (bi-moraic) and short (mono-moraic) Japanese vowels with the long vowels occupying more peripheral places in the vowel space and the short vowels occupying more centralised places in the vowel space (Tsukada, 1999; Hirata & Tsukada, 2004).

4.3 Experimental Series

The experimental part of the present thesis consists of a series of five experiments, described in Chapters 5-7. A brief summary of each Chapter is given below.

4.3.1 Chapter 5: Vocabulary Size Matters - The Assimilation of L2 Australian English Vowels to L1 Japanese Vowel Categories

Chapter 5 presents Thesis Experiment 1 which tests the PAM-L2/Vocab Model prediction that L2 vocabulary size predicts L2 vowel assimilation to L1 categories. It does so by examining the perception of all 18 AusE vowels (presented in a bi-syllabic nonsense word in both citation form and in a carrier sentence) by newly-immersed Japanese learners who vary in the size of their L2 vocabulary. Adopting a ‘whole-system’ approach by allowing these learners to apply all native sensitivities (i.e., all Japanese mono- and bi-moraic vowels as well as all bi-moraic combinations) to the full L2 vowel system, the study also
tests the shared PAM and SLM prediction that L1 inventory size effects L2 perception by exploring the perceptual flexibility of the learners.

The results support the PAM-L2 vocabulary hypothesis and show that L2 learners with a larger L2 vocabulary are more consistent in their vowel assimilation patterns than learners with smaller L2 vocabularies. The results also indicate that learners with an L1 vowel inventory that is smaller than the L2 vowel inventory are able to successfully make use of their full range native categories in the categorisation of L2 phones.

4.3.2 Chapter 6: Larger L2 Vocabulary Size is Associated with Improved L2 Vowel Perception in Adult L2 Learners

The two experiments presented in Chapter 6 replicate and extend the experiment reported in Chapter 5.

The first one, Thesis Experiment 2, replicates the PAM-L2/Vocab Model hypotheses that a larger L2 vocabulary size is associated with more successful L2 vowel assimilation in a large group of Japanese learners of L2 AusE, who, again, differ in the size of their L2 vocabularies. The experiment also extends the work presented in Chapter 5 by the inclusion of a test of discrimination of a select set of L2 vowel contrasts. The L2 contrasts presented in the discrimination experiment was based on the perceptual assimilation patterns revealed in Thesis Experiment 1 (Chapter 5); they were selected specifically to examine differences in the handling of SC, TC and UC/UU assimilation types. The results again support the PAM-L2/Vocab Model predictions.

The next experiment, Thesis Experiment 3, investigates vocabulary-related change in L2 perception in a longitudinal study. We first present assimilation and discrimination data from learners with less than 12 weeks of immersion. We then compare this to data collected from the same group of learners after 6-8 months of continuous L2 immersion. Again, we present evidence in support of the PAM-L2/Vocab Model hypothesis that L2 vocabulary expansion drives, or at least heavily contributes to, change in L2 segmental perception. The effect of extended immersion on the perception of L2 vowels is also discussed, as Thesis Experiment 3 highlights a surprisingly rapid process of reattunement and rephonologisation, with respect to AusE vowels, in the Japanese learners.
4.3.3 Chapter 7: Second Language Learners’ Vocabulary Expansion is Associated with Improved L2 Vowel Intelligibility

The two experiments presented in Chapter 7 test the PAM-L2 and Vocab Model predictions for the role of L2 vocabulary development in L2 vowel production, specifically vowel intelligibility.

First, in Thesis Experiment 4, we test the Vocab Model hypothesis that newly immersed L2 learners with larger L2 vocabularies produced L2 segments (vowels) more intelligibly than otherwise comparable L2 learners with smaller L2 vocabularies. The results support the prediction of the Vocab Model by showing that early-stage L2 learners with larger vocabularies produce more L2 vowels in an intelligible manner than learners with smaller L2 vocabularies, and that learners with larger L2 vocabularies have a higher overall intelligibility than learners with smaller L2 vocabularies.

Next, in Thesis Experiment 5, we investigate vocabulary-related change in L2 production in a longitudinal study, and compare the common-sense prediction that increased L2 exposure and use leads to improved L2 production in all learners against the PAM-L2/Vocab Model prediction that increased L2 exposure and use does not necessarily lead to improved production in learners whose L2 vocabularies have already reached a certain size (>6000 words) and forced them to settle on an accented L2 phonology. We first present vowel intelligibility data from learners with large L2 vocabularies and with less than 12 weeks of immersion and then compare this to data collected from the same group of learners after 6-8 months of continuous L2 immersion. We present evidence in line with the PAM-L2/Vocab Model hypothesis of L2 vocabulary expansion driving change in L2 segmental production, and that ‘too large’ of an L2 vocabulary at onset of L2-environment immersion can curtail further reattunement and rephonologisation.
5 VOCABULARY SIZE MATTERS: THE ASSIMILATION OF L2 AUSTRALIAN ENGLISH VOWELS TO L1 JAPANESE VOWEL CATEGORIES

Revision submitted to Applied Psycholinguistics.

5.1 Abstract

Adult second language (L2) learners’ perception of L2 phonetic segments is influenced by L1 phonological and phonetic properties. It has recently been proposed that L2 vocabulary size in adult learners is related to changes in L2 perception (PAM-L2), analogous to the emergence of L1 phonological function associated with the ‘vocabulary explosion’ at 18 months. To investigate the relationship between L2 perception and vocabulary size, Japanese learners of Australian English (AusE) identified AusE vowels, provided goodness-of-fit ratings, and completed a vocabulary size questionnaire. We adopted a ‘whole-system’ approach, allowing learners to apply all native vowel system possibilities to the full L2 vowel system. Learners with a larger L2 vocabulary were more consistent in their vowel assimilation patterns, compatible with PAM-L2.
5.2 Introduction

It will not surprise many people that second language (L2) learners have foreign accents that are influenced by their particular native language (L1) (Flege, 2002; Flege, Munro, MacKay, 1995; Munro, Flege, MacKay, 1996). It is less obvious, but equally important, that learners of a L2 also have an ‘accent’ in their perception of the new language (Jenkins, Strange, & Polka, 1995) which is systematically related to the perceived similarities between the phonological segments of their L1 and L2 (Best & Strange, 1992; Flege, 1987). Although learners may find some non-native contrasts easy to discriminate, it is common for two or more L2 phones to be perceived as identical or similar to just one native phoneme (Werker & Tees, 1984; Flege, Bohn, & Jang, 1997; Bohn & Flege, 1992; Bohn, 1995; Flege & MacKay, 2004; Guion, 2000; Goto, 1971). However, difficulties in perceiving non-native phones do not always persist as L2 proficiency increases (Flege, Munro, MacKay, 1995; Ingram & Park, 1997; Tsukada, Birdsong, Bialystok, Mack, Sung, & Flege, 2005), though it is unclear what drives this change and results in increased nativelike perception or production. It is the purpose of this paper to investigate the role of vocabulary size in how L2 learners learn to perceive these non-native phones.

In particular, the paper investigates the role of vocabulary size on L2 vowel perception, taking into consideration the entire L1 and L2 vowel systems, rather than a subset of nonnative and native consonants or vowels. We focus on a whole vowel system for a number of reasons. As discussed in detail below, there is ample evidence that a learner’s L1 vowel inventory (size and organisation) influences how L2 vowels are perceived, and vowels are less discretely perceived (and articulated) than consonants (see for instance, Strange, 1998a; 1998b), likely resulting in vowels being inherently more interconnected as a system. If this is the case, then failure to include the entire vowel system in a perceptual experiment would result in an ecologically invalid estimation of the perceptual flexibility of L2 learners.

There is abundant evidence that the size and organisation of the L1 vowel inventory influences how L2 learners perceive the vowel contrasts in their new language. For example, native speakers of Spanish, a language with no temporal or tense-lax spectral contrasts, struggle to discriminate between British English /ɪ/ and /ɨ/, because both are
perceived as instances of Spanish /i/ and their native phonology is not attuned to durational differences as being significant for vowel identification (Escudero & Boersma, 2004). In contrast, native speakers of Serbian, a language that also lacks this vowel contrast but does have a long and short version of /i/, discriminate between these vowels very well on the basis of duration because they are attuned to durational differences as a marker of phonological identity (Krebs-Lazendic & Best, 2008). Moreover, native speakers of German, a language with an /ɪ/-/ɪ/ contrast, discriminate /i/ and /ɪ/ quite well on the basis of spectral differences, although the German realisation of these two vowels is slightly different to the English pronunciations (Bohn & Flege, 1992; Iverson & Evans, 2007).

In addition, the number of vowels in learners’ L1 influences their L2 vowel perception. The perceptual difficulty experienced by an L2 learner is partly determined by the size of the L1 vowel inventory relative to the L2 vowel inventory. Thus, it is harder for speakers of L1s with smaller vowel inventories (such as Spanish) to acquire a rich L2 vowel inventory, relative to speakers of L1s with larger vowel inventories (such as German and Norwegian), because several L2 vowels may be perceived as similar to just one L1 vowel category and consequently will be hard to discriminate (Iverson & Evans, 2007).

Perceptual sensitivity to L2 vowel duration has been shown to be a more highly salient cue to vowel identity than spectral information. For example, Japanese learners of English systematically distinguish American English tense (/ɪ/, /ɛɪ, ɛː, æː, ɔː, ʌː, ou, u:/) and lax (/ɪ, ε, æ, ʌ, u/) vowels (Strange, Akahane-Yamada, Kubo, Trent, Nishi, & Jenkins, 1998), AusE vowels /æe/ and /æːe:/ (Ingram & Park, 1997), and Canadian English /ɪ/ and /ɪ/ (Morrison, 2002) on the basis of duration, even though the durational differences per se are not phonemic in any of these three dialects of English, for which the relevant dimensions are tense/lax, leading English listeners to rely on spectral rather than purely-temporal differences. The high saliency of durational differences is thus particularly strong for speakers of languages where duration is indeed phonemic (including vowel germination contrasts), such as Estonian (McAllister, Flege, & Piske, 2002), though it has also, to a lesser extent, been documented for speakers of L1s where vowel duration is not phonemic, but who have a tense-lax distinction, such as English and German (Bohn, 1995; Bohn & Flege, 1992; Cebrian, 2006; Flege et al., 1997; Gottfried & Beddor, 1988; Goudbeek,
Cutler, & Smits, 2008), and even less so for speakers of languages without length (germination) and tense-lax distinctions, such as Spanish.

Nevertheless, differences and interactions between the L1 and L2 phonological systems of a learner do not mean that an L2 learner’s perception will never improve or come to more closely resemble that of a native speaker of the L2. A large body of literature documents that a learner’s perceptual difficulties are partly determined by their level of familiarity with and use of the L2, and that increased L2 exposure and use typically lead to improved perception and production in the L2 (Flege, Munro, & MacKay, 1995: Ingram & Park, 1997; Tsukada et al., 2005).

Unfortunately, it is not clear how this change in perception (and production) occurs, and the underlying mechanisms have not been adequately addressed by the two primary theoretical models of L2 production and perception: the Speech Learning Model (SLM: Flege, 1995), which focuses on experienced L2 speakers; and the Perceptual Assimilation Model (PAM: Best, 1995), which focuses on naïve listeners. Recently, an extension of PAM, PAM-L2 (Best & Tyler, 2007), which focuses on L2 development from within the general PAM framework, has attempted to fill this theoretical gap as well as to provide testable hypotheses for how perceptual changes come about.

PAM (Best 1994; 1995) assumes that L1 acquisition is essentially the fine-tuning of the perceptual systems to those articulatory gestures in the L1 that are meaningful, and that naïve perception of L2 phones will reflect this L1 tuning, both in terms of the phonetic realisations and the phonological organisation of the speaker’s native language. PAM-L2 (Best & Tyler, 2007) also assumes that perceptual learning is possible at all ages but will be influenced by the entire language learning history of the individual.

PAM-L2 further posits that an increased L2 vocabulary may “exert forceful linguistic pressure” (Best & Tyler, 2007, p. 32) on the learner to attune to articulatory, phonetic, and phonological differences in the L2 that have previously been ignored in the L1, and consequently causes the learner to rephonologise, i.e., establish an L2 phonology – by modification of or addition to – the learner’s existing L1 phonological system. This vocabulary-driven rephonologisation is proposed to work in a similar fashion to the
emergence of L1 phonological function that appears to be driven by the ‘vocabulary explosion’ in children around 18+ months (e.g., Metsala & Walley, 1998). This linguistic pressure may be particularly strong in densely populated lexical neighbourhoods with many minimal pairs, especially where several contrastive L2 phones are assimilated into the same L1 category. It is important to note that we do not imply that reattunement and rephonologisation in L2 learners necessarily leads to the formation of a complete new set of L2 phonological categories, but rather that the learner settles on a phonological system for the L2, perhaps by stretching the L1 inventory as far as possible, and perhaps by forming new L2 categories.

Given that most theories of L2 acquisition (such as PAM/PAM-L2 and SLM) assume that L2 acquisition is based on the same processes of acquisition as is L1 acquisition (though the language history of an individual most often results in L2 acquisition being affected by the L1 of that learner), the suggestion that L2 reattunement and rephonologisation is closely associated with L2 vocabulary development does have some support. Research into L1 acquisition in young infants and children suggests a circular or cyclic relationship between segmental speech perception, vocabulary size and speech production abilities (see for instance, Smith, McGregor, & Demille, 2006). Indeed, early vocabulary size has been found to predict the speech perception abilities of children between the ages of 2-6 years (Metsala, 1999), highlighting the interdependency of vocabulary development and speech perception. Segmental perception (of phonetic contrasts) in the first year of life has also been found to predict vocabulary size in the second year of life (Tsao, Liu & Kuhl, 2004), suggesting that successful phonetic attunement facilitates word-learning, which in turn supports the onset of phonologisation at the time of the ‘vocabulary expansion’ (see for instance, Stager & Werker, 1997; Swingley, 2003).

The present study tests the PAM-L2 hypothesis that a larger L2 vocabulary drives a process of rephonologisation for adult L2 learners. We do so by examining the similarities and differences in L2 speech perception in two groups of speakers (differing in the size of their L2 vocabulary, but not on any other central measure such as AOA, L2 immersion duration, and number of years of FLA) who speak a language with a limited vowel inventory and duration-based phonological vowel distinctions (Japanese), and who are
acquiring an L2 with a larger vowel inventory but no pure duration-based phonological distinctions (Australian English).

This L1-L2 combination has traditionally been seen to pose a significant challenge to Japanese learners, who will have to adjust their limited vowel inventory of just five unique, and relatively pure, vowels as these occur in five spectrally similar short-long pairs (/i, i:, e, e:, a, a:, u, u:, o, o:/) (IPA, 1999). However, when the possible Japanese bi-moraic combinations (/ie, ia, iu, io, ei, ea, eu, eo, ai, ae, au, ao, ui, ue, ua, uo, oi, oe, oa, ou/) are added to the L1 Japanese learners’ category inventory, the task of mapping the the 18 AusE vowels (monophthongs /i:, i, e, e:, ə, æ, e, ə:, o, u, o/ and diphthongs /æə, æɪ, əe, əɪ, əʊ, æə/) (Cox, 2006) may be less extremely taxing. Complicating the picture, however, is the fact that all AusE monophthongs are characterised by some degree of spectral change during production and are generally differentiated on the basis of F1 and F2, with the exception of /i:/ and /æɪ/, which are differentiated mainly on the basis of duration (Cox, 2006).

In a cross-language mapping experiment, we adopted a ‘whole-system’ approach and presented the learners with the entire AusE vowel inventory and all Japanese mono-moraic or bi-moraic categories possible. This approach allows a more complete assessment of the perceptual flexibility of the learners than would the presentation of a pre-selected sub-set of the L2 vowels because it provides the Japanese learners with the full range of L2 input as well as the opportunity to exploit all native sensitivities as well as L2 vowel differences, spectral as well as durational.

In line with the PAM-L2 hypothesis that L2 vocabulary size contributes centrally to L2 phonological reattunement (Best & Tyler, 2007), we expected L2 learners with a larger L2 vocabulary to more consistently identify L2 vowels in terms of their L1 vowel categories than learners with a smaller L2 vocabulary. This is based on the reasoning that the more advanced acquisition of L2 vocabulary by the former subgroup would have already driven them to perceptually attune to non-native phonetic and phonological differences and to begin to reorganise their native phoneme inventory to accommodate the L2 vowel system, at least to some degree (and/or to begin to establish new L2 vowel categories for certain
L1-L2 differences, as discussed below). This is compatible with both PAM and SLM, which propose (i) that learners are likely to perceive (and produce) non-native phones on the basis of their similarity to or dissimilarity from existing L1 phones (on a scale from new over similar to identical), and (ii) that more advanced learners may be more successful in integrating the L2 phones into their existing phonological system (or establishing new L2 phonemes), because their increased experience with the (vocabulary of their) L2 may have prompted them to better attune to the meaningful (i.e., phonological) differences in the L2 and more successfully use their L1 categories and -sensitivities to differentiate between (and possibly create novel categories for) these L2 phones.

In the case of L2 vowel acquisition, we further suggest that it is unlikely that an L2 vowel will be perceived as an entirely new category (i.e., as unrelated to any L1 vowel category) in early acquisition, but rather that ‘difficult’ L2 vowels will be initially be perceived as somewhat similar to a number of L1 vowels, in line with research indicating that vowels are produced in a manner less discrete than consonants and rather continuously perceived. We also suggest that such vowels will be more consistently identified as belonging to just one L1 category through further perceptual reattunement and rephonologisation, which may allow the learner to identify those aspects of the uncategorised L2 phone that most systematically relate to just one L1 category (facilitating L2 comprehension), or realise that the phone does not systematically relate to any L1 category, leading perhaps to the formation of a new L2 category.

Further, in line with PAM/PAM-L2, we expected that the Japanese learners would be highly sensitive to durational differences as well as to some spectral differences in the AusE vowel system. Specifically, we expected that short AusE monophthongs would be identified as Japanese mono-moraic vowels, long AusE monophthongs would be identified as bi-moraic identical (‘long’) Japanese vowels, and diphthongs would be identified as instances of bi-moraic Japanese vowel-combinations.

Also in line with PAM, we expected that the perceptual assimilations of Australian English vowels by the learners would reflect not only phonetic similarities of their native vowel
inventory with the non-native phones, but also the perceived L1/L2 similarities on a phonological level.

5.3 Method

5.3.1 Stimuli

Eight male L1 Western Sydney Australian English (AusEWS) speakers, \(M_{age} = 22.7\) years, produced five randomised repetitions of the 11 AusE stressed monophthongs and seven diphthongs in a first syllable stressed \(/hVba/\) context in citation form (C) and in a carrier sentence context (S) (‘I say \(/hVba/\) for fun’). The \(/hVba/\) context was chosen to minimise any effect of consonant co-articulation on the vowels (Strange, Weber, Levy, Shafiro, Hisagi, & Nishi, 2007). While we tentatively suggest that lexical neighbourhood density may generally play a role in rephonologisation by increasing pressure on the learner to attune to the phonological organisation of the L2, the present study did not directly assess the affect of lexical neighbourhood effects. Rather, the focus was on the role of vocabulary size in perceptual reattunement and rephonologisation. The fact that the selected \(/hVba/\) context yields but one AusE real word (harbour) may be a prudent approach as any effects that the non-word nature of the disyllables may have on the performance of the learners is likely to be in the direction of reducing the effect of the learners’ vocabulary on their L2 perception, that is, we took a conservative approach that mitigates against our hypothesis.

Speakers were instructed to talk as if to a friend, at a normal conversational speaking rate. Three speakers were selected on the basis of auditory judgments of similar voice quality, general reading style and speed, by two phonetically trained experimenters. Three tokens of each vowel were selected from each of those three speakers on the basis of similar intonation pattern and speaking rate.

The recording took place in a sound-attenuated room at the University of Western Sydney, using a Shure SM10A headset microphone, an LG laptop computer and an external soundcard (Edirol UA-25). The recordings were high-pass filtered in Cool Edit with a cut-off frequency of 75Hz to eliminate a low-frequency rumble from the recording. This did not impair the intonation of the utterances, as that none of the F0 contours for the selected
tokens dipped below 75 Hz. The intensity of all words was adjusted in Praat so that the RMS intensity of the target vowel was equal to 70dB (Boersma & Weenink, 2008).

Vowel onset was defined as the beginning of the first regular and recognisable pitch pulse, and vowel offset was defined as the cessation of regular pitch pulsing. Following previous research, such as Strange et al. (1998), vowel duration as well as F1, F2, and F3 values at 25%, 50%, and 75% of the target vowel of the first syllable were estimated. The formant values were estimated using the Praat command ‘To Formant (burg)’. Time-step was 2.5 ms, maximum number of formants 5, maximum formant frequency 5 kHz, with a 25 ms window length. Pre-emphasis was from 50 Hz. The obtained measurements are shown in Tables 5.1 and 5.2.

Six dependent variables (F1 and F2 at 25%, 50%, and 75% of the vowel) were analysed in an 18 x 2 Multivariate Analysis of Variance (MANOVA) with the independent variables of vowel type and presentation context (Sentence [S] vs. Citation [C] presentation) revealed that there was no main effect of presentation context, $F(6,102) = 1.207, p > .01$. We did not include vowel duration, as any difference in duration would most likely have reflected speaking-rate rather than intrinsic differences in the vowels between the S and C conditions.

The formant values reported in Tables 5.1 and 5.2 are consistent with the vowel data reported by Cox (2006), with the exception of the diphthong /æə/. In our data, this diphthong is produced with a much higher (and also slightly more fronted) first target than /æə/ and should probably be transcribed as [œr].
Table 5.1. Citation-context (C) vowel durations (ms) and F1, F2, and F3 values (in Hz) at 25%, 50%, and 75% of the target vowels and diphthongs of Australian English (AusE).  

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Dur. (ms)</th>
<th>F1 (Hz) 25%</th>
<th>F1 (Hz) 50%</th>
<th>F1 (Hz) 75%</th>
<th>F2 (Hz) 25%</th>
<th>F2 (Hz) 50%</th>
<th>F2 (Hz) 75%</th>
<th>F3 (Hz) 25%</th>
<th>F3 (Hz) 50%</th>
<th>F3 (Hz) 75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a:/</td>
<td>128</td>
<td>335</td>
<td>317</td>
<td>310</td>
<td>2169</td>
<td>2216</td>
<td>2219</td>
<td>3102</td>
<td>3148</td>
<td>3032</td>
</tr>
<tr>
<td>/al/</td>
<td>59</td>
<td>333</td>
<td>342</td>
<td>344</td>
<td>2119</td>
<td>2085</td>
<td>1910</td>
<td>2983</td>
<td>2706</td>
<td>2424</td>
</tr>
<tr>
<td>/æl/</td>
<td>129</td>
<td>327</td>
<td>325</td>
<td>341</td>
<td>2172</td>
<td>2167</td>
<td>2142</td>
<td>3132</td>
<td>3118</td>
<td>2405</td>
</tr>
<tr>
<td>/e:/</td>
<td>74</td>
<td>547</td>
<td>558</td>
<td>528</td>
<td>1873</td>
<td>1829</td>
<td>1697</td>
<td>2565</td>
<td>2474</td>
<td>2388</td>
</tr>
<tr>
<td>/æ:/</td>
<td>143</td>
<td>512</td>
<td>523</td>
<td>526</td>
<td>1833</td>
<td>1869</td>
<td>1789</td>
<td>2622</td>
<td>2552</td>
<td>2543</td>
</tr>
<tr>
<td>/æl/</td>
<td>85</td>
<td>711</td>
<td>695</td>
<td>646</td>
<td>1535</td>
<td>1487</td>
<td>1412</td>
<td>2618</td>
<td>2570</td>
<td>2538</td>
</tr>
<tr>
<td>/u:/</td>
<td>130</td>
<td>540</td>
<td>489</td>
<td>409</td>
<td>1910</td>
<td>2006</td>
<td>2020</td>
<td>2673</td>
<td>2634</td>
<td>2479</td>
</tr>
<tr>
<td>/æːl/</td>
<td>142</td>
<td>708</td>
<td>731</td>
<td>681</td>
<td>1524</td>
<td>1367</td>
<td>1226</td>
<td>2477</td>
<td>2417</td>
<td>2540</td>
</tr>
<tr>
<td>/æːl/</td>
<td>147</td>
<td>708</td>
<td>702</td>
<td>603</td>
<td>1185</td>
<td>1374</td>
<td>1575</td>
<td>2589</td>
<td>2533</td>
<td>2562</td>
</tr>
<tr>
<td>/æl/</td>
<td>73</td>
<td>698</td>
<td>678</td>
<td>628</td>
<td>1198</td>
<td>1171</td>
<td>1102</td>
<td>2719</td>
<td>2734</td>
<td>2703</td>
</tr>
<tr>
<td>/æːl/</td>
<td>141</td>
<td>672</td>
<td>682</td>
<td>657</td>
<td>1262</td>
<td>1265</td>
<td>1227</td>
<td>2636</td>
<td>2640</td>
<td>2543</td>
</tr>
<tr>
<td>/ɪ:/</td>
<td>70</td>
<td>618</td>
<td>608</td>
<td>552</td>
<td>933</td>
<td>930</td>
<td>904</td>
<td>2760</td>
<td>2771</td>
<td>2745</td>
</tr>
<tr>
<td>/o:/</td>
<td>119</td>
<td>454</td>
<td>497</td>
<td>491</td>
<td>698</td>
<td>694</td>
<td>944</td>
<td>2825</td>
<td>2838</td>
<td>2859</td>
</tr>
<tr>
<td>/oːl/</td>
<td>142</td>
<td>499</td>
<td>474</td>
<td>387</td>
<td>1079</td>
<td>1549</td>
<td>1935</td>
<td>2530</td>
<td>2486</td>
<td>2537</td>
</tr>
<tr>
<td>/uːl/</td>
<td>76</td>
<td>378</td>
<td>385</td>
<td>361</td>
<td>929</td>
<td>951</td>
<td>986</td>
<td>2528</td>
<td>2532</td>
<td>2531</td>
</tr>
<tr>
<td>/ʊːl/</td>
<td>110</td>
<td>376</td>
<td>378</td>
<td>371</td>
<td>1579</td>
<td>1544</td>
<td>1449</td>
<td>2197</td>
<td>2201</td>
<td>2184</td>
</tr>
<tr>
<td>/uːːl/</td>
<td>116</td>
<td>570</td>
<td>522</td>
<td>467</td>
<td>1240</td>
<td>1310</td>
<td>1321</td>
<td>2411</td>
<td>2365</td>
<td>2302</td>
</tr>
<tr>
<td>/aːl/</td>
<td>144</td>
<td>487</td>
<td>504</td>
<td>490</td>
<td>1462</td>
<td>1468</td>
<td>1425</td>
<td>2409</td>
<td>2421</td>
<td>2405</td>
</tr>
</tbody>
</table>
Table 5.2. Sentence-context (S) vowel duration (ms) and F1, F2, and F3 (in Hz) values at 25%, 50%, and 75% of the vowels and diphthongs of Australian English (AusE₃₅).

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Dur. (ms)</th>
<th>F1 (Hz)</th>
<th>F2 (Hz)</th>
<th>F3 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>25%</td>
<td>50%</td>
<td>75%</td>
</tr>
<tr>
<td>/i:/</td>
<td>140</td>
<td>330</td>
<td>304</td>
<td>308</td>
</tr>
<tr>
<td>/æ/</td>
<td>66</td>
<td>330</td>
<td>347</td>
<td>364</td>
</tr>
<tr>
<td>/æ/</td>
<td>145</td>
<td>305</td>
<td>316</td>
<td>339</td>
</tr>
<tr>
<td>/e:/</td>
<td>92</td>
<td>541</td>
<td>559</td>
<td>534</td>
</tr>
<tr>
<td>/e:/</td>
<td>162</td>
<td>488</td>
<td>506</td>
<td>512</td>
</tr>
<tr>
<td>/æ:/</td>
<td>91</td>
<td>806</td>
<td>763</td>
<td>678</td>
</tr>
<tr>
<td>/æ:/</td>
<td>149</td>
<td>607</td>
<td>489</td>
<td>399</td>
</tr>
<tr>
<td>/æ:/</td>
<td>160</td>
<td>752</td>
<td>752</td>
<td>684</td>
</tr>
<tr>
<td>/æ:/</td>
<td>159</td>
<td>710</td>
<td>738</td>
<td>630</td>
</tr>
<tr>
<td>/e:/</td>
<td>69</td>
<td>736</td>
<td>697</td>
<td>628</td>
</tr>
<tr>
<td>/e:/</td>
<td>149</td>
<td>726</td>
<td>725</td>
<td>699</td>
</tr>
<tr>
<td>/e:/</td>
<td>69</td>
<td>611</td>
<td>586</td>
<td>527</td>
</tr>
<tr>
<td>/o:/</td>
<td>127</td>
<td>441</td>
<td>440</td>
<td>441</td>
</tr>
<tr>
<td>/o:/</td>
<td>151</td>
<td>469</td>
<td>477</td>
<td>403</td>
</tr>
<tr>
<td>/o:/</td>
<td>67</td>
<td>342</td>
<td>359</td>
<td>348</td>
</tr>
<tr>
<td>/o:/</td>
<td>141</td>
<td>513</td>
<td>494</td>
<td>429</td>
</tr>
<tr>
<td>/o:/</td>
<td>141</td>
<td>513</td>
<td>494</td>
<td>429</td>
</tr>
<tr>
<td>/o:/</td>
<td>165</td>
<td>487</td>
<td>519</td>
<td>511</td>
</tr>
</tbody>
</table>
5.3.2 Participants

Eleven native speakers of Japanese (8 female, $M_{\text{age}} = 26.4\text{ years}$) participated in the study. All had studied English in middle and high school mainly with native Japanese teachers, as well as some native English guest teachers ($M_{\text{study}} = 7\text{ years}, M_{\text{onset age}} = 11.6\text{ years}$). While two participants had had only Japanese teachers, the others had also been exposed to either Australian English or British English, though to a lesser extent than their exposure to Japanese accented English. Five participants had been exposed more than one native English dialect (with two reporting exposure to four different dialects of English). A tally of the countries of origins of the teachers showed that nine participants had had teachers that spoke Australian English; seven had had British English-speaking teachers, three had had American English speaking teachers, two had New Zealand English, and two Irish English speaking teachers. All had of course also had contact with various dialects of English through music and films, and other mass media. The highly dialectally varied L2 input that these students had received most likely contributed to some degree of target-confusion (such as has been reported for native Danish learners of English in a study by Bohn & Bundgaard-Nielsen [2008]), rather than a specific dialectal bias. All participants reported that they did not yet feel ‘confident’ in speaking English. They were all students at English colleges in Sydney, their goal being to learn spoken English, and all had spent less than 12 weeks in Australia ($M_{\text{stay}} = 7\text{ weeks}$). None had previously lived outside of Japan.

5.3.3 Procedure

All participants were tested in a quiet, sound-attenuated experiment room at MARCS Auditory Laboratories at University of the Western Sydney. Stimuli were presented over studio headphones from a Mac Book using PsyScope. The participants first heard a randomised presentation of all AusE_{WS} vowels over headphones ($N = 324$) in C and S contexts (presentation order of the two contexts was counter-balanced across participants) and identified them in terms of their L1 vowels on a computer screen, using a grid of Japanese katakana symbols representing all short, long and combination-vowels possible: $/i, i:, e, e:, a, a:, u, u:, o, o:, i_e, i_a, i_u, i_o, e_i, e_a, e_u, e_o, a_i, a_e, a_u, a_o, u_i, u_e, u_a, u_o, o_i, o_e, o_a, o_u$/ preceded by $/h/$. Additionally, the participants rated goodness-of-fit of each vowel.
on a scale from 1 (poor) to 7 (excellent). Finally, they completed a multiple-choice L2-English vocabulary size test (Nation & Beglar, 2007). The Nation & Beglar vocabulary size test was selected for several reasons: Firstly, and most importantly because, rather than assessing vocabulary acquisition on different ‘levels’ of vocabulary (see for instance Laufer & Nation [1999] and Nation [1993]), it offers an estimate of a learner’s recognition vocabulary irrespective of which level of vocabulary the learner has focussed on in his or her L2 acquisition process. Secondly, while tests of general proficiency will likely also be associated with segmental perception (perhaps as a result of improvements to L2 vocabulary size and the proposed related improvements to segmental perception), they include other measures of L2 proficiency which have not previously been directly posited to contribute centrally to the process of (re)phonologisation, as has vocabulary development in L1 learning infants and children. The possibility of an analogous vocabulary-growth effect on L2 speech perception was originally proposed in PAM-L2 (Best & Tyler, 2007), and was examined for the first time in the present report.

5.4 Results

5.4.1 General Assimilation Pattern

An L2 vowel was defined as ‘categorised’ if it was identified as an instance of one L1 category in more than 50% of presentations (following Best, Faber, & Levitt, 1996). Thirteen of the 18 AusE vowels were consistently assimilated to Japanese vowel categories in both the C and S contexts. Five were uncategorised, that is, not identified as any one Japanese vowel category for 50% or more of tokens (see Table 5.3).
Table 5.3. The assimilation of citation AusE (left item in ‘Identification’ columns) to L1 Japanese (right item in ‘Identification’ columns) vowel categories by all listeners. GR is the mean goodness rating on a 7 point scale (1 = poor fit, 7 = excellent fit).

<table>
<thead>
<tr>
<th>Categorised to L1</th>
<th>Uncategorised</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single mora</strong></td>
<td><strong>Double mora</strong></td>
</tr>
<tr>
<td>Identification</td>
<td>GR</td>
</tr>
<tr>
<td>/i/ → /i/</td>
<td>4.90</td>
</tr>
<tr>
<td>/e/ → /e/</td>
<td>5.12</td>
</tr>
<tr>
<td>/æ/ → /æ/</td>
<td>5.13</td>
</tr>
<tr>
<td>/o/ → /o/</td>
<td>4.86</td>
</tr>
</tbody>
</table>
| /o/ → /o/         | 5.05 | /oː/ → /oː/ | 4.88 | /æː/  

The pattern of assimilation did not differ for C and S presentations; both reflected the Japanese learners’ sensitivity to spectral as well as durational information. Most L2 vowels were assimilated into unique L1 categories, though both AusE /iː/ and /tə/ were assimilated to Japanese /iː/. Long (i.e., tense) AusEWS vowels were typically identified as bi-moraic Japanese vowels, and short (i.e., lax) AusEWS vowels were typically identified as mono-moraic Japanese vowels. The spectrally similar (but durationally different) vowel-pairs /eː/-/eː/ and /æː/-/æː/ were successfully assimilated to four separate Japanese vowel categories, most likely on the basis of systematic durational differences, though the slight spectral differences may also have contributed.

The average goodness ratings differed for vowels in C (M = 4.96) and S contexts (M = 5.25), F(1,24) = 8.691, p < .01. All 30 possible Japanese categories were used at least once by at least one participant in both C and S contexts. The identification of AusEWS /æː/ as Japanese /æː/ seems to reflect the different realisation of this diphthong between our data and that of previous work on AusE (Cox, 2006). The identification of AusEWS /æː/ and
AusEWS /e:/ as two different Japanese categories /e/-i and /e/-u thus reflects both phonetic similarities between the L2 vowels and the chosen L1 categories, and phonological as both of these two Japanese categories are actually realised as [e:].

5.4.2 Analysis by Vocabulary Size

The mean estimated vocabulary size (from the Nation & Begar test) for the 11 listeners was 6009 words (range 4100-7800 words). To test the hypothesis that L2 vocabulary development is a core contributing factor in L2 vowel phonological reorganisation, they were divided into high and low vocabulary groups using a median split (HV: \( n = 5, \ M_{\text{vocabulary}} = 7200 \); LV: \( n = 6, \ M_{\text{vocabulary}} = 5017 \)). A series of two-tailed paired \( t \)-tests revealed that the HV and LV group differed in terms of their L2 vocabulary size (\( t(9) = 5.40, p = .001 \)), but that they did not differ in terms of years of English study in Japan (\( M_{\text{HV}} = 7.8 \) years, \( M_{\text{LV}} = 6.3 \) years, \( t(9) = .140, p = .89 \)) nor in their length of residence in Australia on the day of testing (\( M_{\text{HV}} = 7.2 \) weeks, \( M_{\text{LV}} = 6.8 \) weeks, \( t(9) = 1.976, p = .10 \)).

As predicted, differences in the identification patterns of the learners were related to their vocabulary size. The general identification pattern for the HV and LV groups was almost identical (see Table 3 for that common pattern), given that only one L2 vowel, AusEWS /æ/, reached the ‘categorised’ criterion for the HV group (94% perceived as /ai/) but not for the LV group (49% perceived as /ai/) in the Citation context. However, the two groups differed systematically in two other aspects of their performance on the assimilation task. Firstly, the HV group selected a significantly smaller number of L1 categories for the AusEWS vowels (\( M = 25 \)) than the LV group (\( M = 27.67 \)). Paired two-tailed \( t \)-tests revealed that this difference was significant: HV-LV for C context: \( t(17) = 2.77, p < .001 \), HV-LV for S context: \( t(17) = 2.40, p < .001 \) Secondly, paired two-tailed \( t \)-tests revealed that the HV group was more consistent in its selection of L2 phone to L1 category matching than the LV group: the mean identification score of the categorised vowels for the HV group was 85% (S) and 80% (C), while the LV group’s mean was lower at 73% (S) and 74% (C), C context: \( t(12) = 1.26, p < .23 \), S context: \( t(12) = 2.38, p < .04 \).
5.4.3 Uncategorised Vowels

There was considerable variation in the overall identification pattern for the five un categorised AusEWS vowels. For example, in the case of AusEWS /ɔː/, the three most frequently selected Japanese categories included 91% of the tokens. In contrast, for AusEWS /ʃ/, the three most frequently selected Japanese categories included only an average of 53% of the tokens (See Figure 5.1).

The HV and LV groups differed in the consistency of their pattern of identification for the un categorised vowels, with the top three selections for the HV group averaging 74% of responses, and the LV group averaging only 65% of responses, $t(9) = -14.11, p = .001$ (See Table 5.4). However, the difference in identification consistency between the two presentation contexts was not significant, $t(9) = 2.17$.

![Figure 5.1](image.png)

**Figure 5.1.** Percent total of the three most selected Japanese categories for the five un categorised AusEWS vowels for the High Vocabulary group (HV), and the Low Vocabulary group (LV).
Table 5.4. Assimilation patterns to the top three selected L1 Japanese vowel categories (vertical) for the five uncategorised AusEWS vowels (horizontal). Group refers to High Vocabulary learners (HV) and Low Vocabulary learners (LV). % indicates the total percent of each of the uncategorised AusEWS vowels which were identified as each of the top three L1 categories.

<table>
<thead>
<tr>
<th>AusE Vowel</th>
<th>Context</th>
<th>Group</th>
<th>1st</th>
<th>%</th>
<th>2nd</th>
<th>%</th>
<th>3rd</th>
<th>%</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>/o:/</td>
<td>Citation</td>
<td>HV</td>
<td>/ou/</td>
<td>42</td>
<td>/o:/</td>
<td>32</td>
<td>/o:/</td>
<td>21</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>/o:/</td>
<td>47</td>
<td>/ou/</td>
<td>28</td>
<td>/o:/</td>
<td>15</td>
<td>90</td>
</tr>
<tr>
<td>/æ/</td>
<td>Sentence</td>
<td>HV</td>
<td>/a:/</td>
<td>42</td>
<td>/æ:/</td>
<td>37</td>
<td>/æ:/</td>
<td>16</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>/o:/</td>
<td>44</td>
<td>/æ:/</td>
<td>33</td>
<td>/æ:/</td>
<td>9</td>
<td>86</td>
</tr>
<tr>
<td>/æ:/</td>
<td>Citation</td>
<td>HV</td>
<td>/æ:/</td>
<td>39</td>
<td>/æ:/</td>
<td>18</td>
<td>/æ:/</td>
<td>13</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>/æ:/</td>
<td>36</td>
<td>/æ:/</td>
<td>26</td>
<td>/æ:/</td>
<td>16</td>
<td>77</td>
</tr>
<tr>
<td>/æ:/</td>
<td>Sentence</td>
<td>HV</td>
<td>/æ:/</td>
<td>33</td>
<td>/æ:/</td>
<td>28</td>
<td>/æ:/</td>
<td>20</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>/æ:/</td>
<td>42</td>
<td>/æ:/</td>
<td>20</td>
<td>/æ:/</td>
<td>9</td>
<td>70</td>
</tr>
<tr>
<td>/æ:/</td>
<td>Citation</td>
<td>HV</td>
<td>/æ:/</td>
<td>42</td>
<td>/æ:/</td>
<td>20</td>
<td>/æ:/</td>
<td>18</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>/æ:/</td>
<td>23</td>
<td>/æ:/</td>
<td>24</td>
<td>/æ:/</td>
<td>19</td>
<td>65</td>
</tr>
<tr>
<td>/æ:/</td>
<td>Sentence</td>
<td>HV</td>
<td>/æ:/</td>
<td>45</td>
<td>/æ:/</td>
<td>18</td>
<td>/æ:/</td>
<td>11</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>/æ:/</td>
<td>19</td>
<td>/æ:/</td>
<td>19</td>
<td>/æ:/</td>
<td>17</td>
<td>55</td>
</tr>
<tr>
<td>/æ:/</td>
<td>Citation</td>
<td>HV</td>
<td>/ou/</td>
<td>48</td>
<td>/u:/</td>
<td>13</td>
<td>/u:/</td>
<td>4</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>/ou/</td>
<td>26</td>
<td>/u:/</td>
<td>19</td>
<td>/o:/</td>
<td>9</td>
<td>54</td>
</tr>
<tr>
<td>/æ:/</td>
<td>Sentence</td>
<td>HV</td>
<td>/u:/</td>
<td>41</td>
<td>/ou/</td>
<td>19</td>
<td>/au/</td>
<td>9</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>/u:/</td>
<td>31</td>
<td>/a:/</td>
<td>10</td>
<td>/au/</td>
<td>9</td>
<td>51</td>
</tr>
<tr>
<td>/æ:/</td>
<td>Citation</td>
<td>HV</td>
<td>/u:/</td>
<td>31</td>
<td>/ua/</td>
<td>16</td>
<td>/au/</td>
<td>12</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>/u:/</td>
<td>39</td>
<td>/e:/</td>
<td>9</td>
<td>/ua/</td>
<td>9</td>
<td>56</td>
</tr>
<tr>
<td>/æ:/</td>
<td>Sentence</td>
<td>HV</td>
<td>/u:/</td>
<td>18</td>
<td>/eu/</td>
<td>18</td>
<td>/ua/</td>
<td>14</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>/u:/</td>
<td>17</td>
<td>/e:/</td>
<td>16</td>
<td>/ue/</td>
<td>15</td>
<td>47</td>
</tr>
</tbody>
</table>

61
5.5 Discussion

This study examined the relationship between L2 vocabulary size and L2 vowel perception in two groups of native Japanese speakers who differed in the size of their L2-English vocabulary. The purpose was to test the PAM-L2 proposal that L2 vocabulary acquisition is central to a phonological reorganisation in the learner (Best & Tyler, 2007). We hypothesised that a larger L2 vocabulary would be positively associated with more L2 vowel identification. We expected that all Japanese learners, irrespective of vocabulary, would make use of spectral and temporal phonetic information in the L2 target items to identify the AusE<sub>WS</sub> vowels with reference to their native Japanese vowel categories. We also expected that the assimilation pattern would reflect not just phonetic similarity, but also phonological similarities.

Our results support the hypothesis that HV and LV L2 learner groups differ in a manner consistent with PAM-L2 predictions. Though the general patterns of identification were similar for the HV and LV groups, the HV learners were not only more consistent in their identification of the L2 phones in terms of L1-assimilation scores, but also in terms of how many alternative L1 categories they selected for each L2 vowel. Further support for a link between vocabulary size and L2 phonological reattunement was found in the consistency of the HV group in their identification of the un-categorised vowels. We interpret this greater consistency as a sign of the HV group being more advanced in integrating the L2 phones into their existing phonological system. Likewise, we interpret the fact that the HV group (but not the LV group) reached identification criterion when identifying AusE<sub>WS</sub> /æə/ as consistent with the PAM-L2 hypothesis about L2 vocabulary effects on L2 vowel perception.

While we acknowledge that these results do not preclude the possibility that influences other than vocabulary size, such as general proficiency, may underlie the observed differences, our explanation is consistent with the PAM-L2 theoretical framework. A general-proficiency explanation would be less parsimonious because it does not provide a clear theoretical link to L2 segmental perception. Indeed, a prior report on Japanese learners of English differing in L2 fluency alone (vocabulary size was not reported to have been assessed independently in that study) found that high and low proficiency learners did
not differ in their identification of AusE vowels /i/, /ɪ, e, æ, ɛ/ using the appropriate AusE labels (Ingram & Park, 1997), indicating that L2 perception is not straightforwardly related to L2 proficiency.

The results are also consistent with PAM-L2 claims that L2 learners will use both temporal and spectral information in their identification of the L2 vowels, irrespective of their L2 vocabulary, as the Japanese learners consistently used both mono- and bi-moraic L1 categories in a pattern that reflected both native phonetic and native phonological properties. The systematic use of single and bi-moraic Japanese categories to identify the short (i.e., lax) and long (i.e., tense) AusE vowels, respectively, is also consistent with the claims that durational information plays an important role in the discrimination of non-native phones, in particular when the spectral information is ambiguous to the listener (Bohn, 1995; Bohn & Flege, 1992; Cebrian, 2006; Flege et al., 1997; Gottfried & Beddor, 1988). The higher mean goodness rating in the sentence ($M = 5.25$) versus citation context ($M = 4.96$) is likely caused by durational information provided in the carrier sentence that helps learners disambiguate phonological duration information to which these native Japanese listeners are highly sensitive in their L1. This additional information in the sentential context may also be the reason for the Japanese listeners’ use of a smaller number of L1 categories in the sentence context, relative to the citation context.

The benefits of the whole-system approach to L2-L1 vowel perception, adopted here, are evident in the fact that Japanese L2 learners, irrespective of their L2 vocabulary size, identified the majority of the AusE vowels as members of Japanese vowel categories (mono- as well as bi-moraic) in a systematic manner. This pattern of successful assimilation would not have been clear from study of a subset of AusE vowels, whether based on a contrastive analysis of the relatively limited Japanese singleton-vowel inventory and the much larger AusE vowel inventory, or the previous literature on English vowel acquisition by Japanese speakers. Those earlier studies did not make the full L1 inventory of tools to differentiate vowels available to the learners. However, the relative success of the Japanese learners in exploiting both spectral and durational vowel settings of the L1 is consistent with PAM claims that L2 phones are classified in terms of L1 phonemes when
they are not perceived as too discrepant from the L1 phonemes, and learners will make use
of whatever information they can to make sense of the non-native system.

We propose that future research continues to explore the role of L2 vowel acquisition in L2
segmental perception and that the ‘whole-system’ approach, pioneered here, could
fruitfully be adopted to gain a better understanding of the flexibility of a learner’s
perceptual system, in particular with respect to studies of L2 vowel perception.
Furthermore, in light of research suggesting that a learner’s L2 phonological development
plateaus within the first year of acquisition in immersion settings, and that the most
significant changes in non-native vowel perception are apparent after only 6-12 months of
L2 immersion for adult learners (Flege & Liu, 2001; Aoyama, Flege, Guion, Akahane-
Yamada, & Yamada, 2004; Tsukada et al., 2005), we encourage a meticulous examination
of the relationship between L2 vocabulary growth and L2 phonological acquisition during
the first 12 months of immersion. Such studies would help determine exactly when during
the first 12 months of immersion the learner reaches this plateau, when and how the
phonological re-organisation begins or peaks, and the extent to which it asymptotes. In
conclusion, our findings are consistent with the PAM-L2 prediction that L2 vocabulary
acquisition drives a learner to rephonologise early in immersion-based conversational
acquisition.
5.6 References


6 VOCABULARY SIZE IS ASSOCIATED WITH L2 VOWEL PERCEPTION PERFORMANCE IN ADULT L2 LEARNERS

Submitted to Studies in Second Language Acquisition

6.1 Abstract

Improvement in second-language (L2) perception has been posited to asymptote early in L2 learning when the L2 vocabulary is still small, whereas a large L2 vocabulary curtails perceptual learning (PAM-L2: Best & Tyler, 2007). We extend this proposition by suggesting that early L2 lexical development facilitates establishment of phonological categories in a manner analogous to toddlers’ L1 acquisition before versus after the vocabulary explosion. According to this revised view, L2 speech should be assimilated more consistently to phonological categories, and cross-boundary contrasts discriminated more accurately, by learners with larger versus smaller L2 vocabularies. In order to test this proposition, we applied a novel whole-system approach to evaluate perception of L2 vowels with respect to the L1 vowel system in two experiments. In experiment 1, Japanese learners of Australian English with less than 12 weeks of L2 immersion completed labelling and goodness ratings on all Australian vowels, using all mono- and bi-moraic Japanese vowels and vowel combinations. They also completed discrimination tasks on four L2 vowel contrasts, representing a range of PAM-L2 contrast types, and a L2 vocabulary size assessment. Learners with larger vocabularies had more consistent vowel assimilation and more accurate cross-boundary discrimination than those with smaller vocabularies, supporting the proposition that lexical development assists L2 phonological acquisition. In experiment two, we compared the perception of L2-English vowels by a group of Japanese learners after only weeks of L2 immersion with their perception after 6-8 months of L2 immersion in Australia. The results also supported the predicted positive association between L2 vocabulary size and L2 vowel perception rather than a general prediction of increased immersion duration leading to improved perception.
6.2 Introduction

It is well known that late (adolescent or especially adult) second-language (L2) learners almost invariably have a foreign accent when they speak in their L2 (Flege, 2002; Flege, Munro, MacKay, 1995; Munro, Flege, MacKay, 1996, but see Ioup, Boustagui, El Tigi, & Moselle, 1994). It is less well known that L2 learners have a perceptual foreign accent when they listen to non-native speech (Jenkins, Strange, & Polka, 1995; Best & Strange, 1992; Flege, 1987). The perceptual accent reflects the native language (L1) background of the learner and, in particular, the perceived similarity of the learner’s L1 and L2 phones. For example, the British and American English vowels /i/ and /ɪ/ are distinguished both spectrally and durationally, but native speakers of Spanish struggle to discriminate between them as their native vowel inventory contains only /i/ and lacks phonemic vowel length (duration) or tense/lax contrasts (Escudero & Boersma, 2004; Flege, Bohn, & Jang, 1997).

In contrast, native speakers of Serbian, a language also lacking an /i/-/ɪ/ contrast, but with a long and short version of /i/, discriminate Australian English /i/-/ɪ/ well on the basis of the durational difference, which they perceptually assimilate to Serbian long versus short /i/ (Krebs-Lazendic & Best, 2008).

A second important factor of L1-L2 interaction is the relative phoneme inventory size of their L1 and L2, as a larger L2 than L1 inventory provides the learner with a limited number of native phonological categories to exploit in his or her attempt to tease apart the novel L2 phones. A classic example of how one’s native phonological inventory interferes with the perception of nonnative contrasts is the failure of Japanese learners of English to distinguish and produce the two English liquids [l] and [ɾ], as both are assimilated to the single Japanese liquid [ɾ] (Guion, Flege, Akaha Akahane-Yamada, & Pruitt, 2000; Goto, 1971; Miyawaki et al., 1975; MacKain, Best, & Strange, 1981, Mochizuki, 1981), while native speakers of Germanic languages, such as Danish, which has both /l/ and /ɹ/, have very minor problems with this contrast (Best, & Bohn, 2002; Best et al., 2003), despite their realisation of /ɹ/ (pharyngeal approximant [ɾ]) being very different from that of English (Groennum, 2007). For vowels, there is evidence that native speakers of Norwegian and German, with relatively large (oral) vowel inventories, outperform native
speakers of Spanish and French, with their smaller (oral) vowel systems, on a task of L2 English vowel recognition (Iverson & Evans, 2007).

A third important factor in L2 acquisition is the Age of Acquisition (AOA) of the L2 learner. AOA is related to L2 acquisition success, with respect to both L2 perception and production, such that younger learners generally outperform older learners (for consonants: Aoyama et al., 2004; Tsukada et al., 2005; Flege et al., 1995; Yamada, 1995, Mackay et al., 2001; for vowels: Flege, Yeni-Komshian, & Lui, 1999; Birdsong, 2006; Tsukada, Birdsong, Bialystok, Mack, Sung, & Flege, 2005; Munro et al., 1996; Piske et al., 2002; Flege et al., 2003; for overall accentedness: Seliger et al., 1975; Oyama, 1976; Yeni-Komshian et al., 2000). Despite this advantage for young L2 learners, two influential current theories of L2 acquisition (PAM: Best, 1994; 1995; PAM-L2: Best & Tyler, 2007; and SLM: Flege, 1995) propose that the mechanisms underlying L1 acquisition are still available to the L2 learner in their adult years, well past the completion of their L1 acquisition, and are used in the subsequent acquisition of other languages. This claim is supported by evidence that L2 perception and production improve even in adult learners, irrespective of individual differences in aptitude, as a function of increased experience with and use of the L2 (Flege, Munro, MacKay, 1995; Ingram & Park, 1997; Tsukada, Birdsong, Bialystok, Mack, Sung, & Flege, 2005).

It is not clear, however, what drives this change in adult-learner L2 perception, though both social, such as motivational and usage (for a discussion, see Piske, MacKay, & Flege, 2001), and maturational factors (i.e., the Critical Period Hypothesis; see Lenneberg, 1967, but see also Johnson & Newport, 1989; Oyama, 1976) have been suggested. In the following, we will focus on the influence of L1 phonetic and phonological tuning on L2 acquisition (PAM-L2: Best & Tyler, 2007). According to PAM/PAM-L2, L1 acquisition involves the attunement of the perceptual systems to the meaningful phonetic features and phonological structure/organisation of the language spoken in an infant’s immediate environment (Best & Tyler, 2007). Studies of infant speech perception have demonstrated that infants attend to phonetic differences from birth and are able to discriminate most speech contrasts well until their perceptual systems attune to those contrasts that are meaningful in the surrounding language at approximately 5-6 months of age for vowels,

However, data from the relatively new research topic of early cross-dialect word recognition suggest that, though infants and young toddlers are highly skilled in detecting the phonetic details that are important in their L1, they do not display phonological constancy, i.e., are not able to ‘hear past’ irrelevant phonetic information (like dialect or speaker specific variance) in order to perceive speech on the basis of the underlying phonology. In a study of 15 and 19-month-old learners of Connecticut English, the older toddlers successfully recognised familiar words spoken in the markedly different and unfamiliar Jamaican Mesolec English accent whereas the 15-month-olds failed to do so, although both groups recognized the same words in their native Connecticut accent (Best, Tyler, Gooding, Orlando & Quann, in press). This study was extended to 14- and 19-month-old Australian English-learning toddlers and expanded to address the potential role of vocabulary development. The younger children, who had been selected for having limited vocabularies under 25 expressive words, did not recognise familiar words in an unfamiliar (or even in the familiar) dialect, while older ones, who were selected for notably larger vocabularies over 50 words, did recognize them in both dialects (Best, Tyler, Kitamura, Notley, & Bundgaard-Nielsen, 2008). Additionally, there is evidence that toddlers in the early phase of word-learning attend to more phonetic detail in L1 segmental speech discrimination tasks than in L1 word learning tasks. Specifically, mispronounced words are readily accepted by these young children (14 months), while the same contrasts that differentiate the original versus mispronounced word are well discriminated by infants of 8-10 months, already before their first birthday (Stager & Werker, 1997). In contrast, word-recognition of 19-month-olds is clearly affected by phonetic mispronunciations, so that they reject familiar words when one of the phonemes in the word has been replaced by another to create a mispronounced token of that word (Swingley, 2003).

Together, prior results indicate that early L1 words are learned and recognised in their detailed dialect-specific phonetic form, rather than as structures built of higher order phonological properties, by very young children. Successful cross-dialect recognition of
otherwise familiar words, which reflects recognition of phonological constancy, appears to emerge only around the age of the ‘vocabulary spurt’ (see for example, Metsala & Walley, 1998), which occurs in the average child at approximately 18 months of age. Together, early attunement to ambient phonetic details and later attunement to the abstract phonological structure of familiar words may indicate that pressures from a rapidly increasing vocabulary drive the attunement to the higher order phonological structures in the language (Best & Tyler, 2007), that is, to encourage the child to phonologise the phonetic properties of the input.

On the premise that the drive to phonologise that arises from an increasing vocabulary is a key factor in developmental changes in L1 perception, PAM-L2 (Best & Tyler, 2007) postulates that experienced-based changes in L2 perception are directly influenced by the language learning history of the individual learner. According to this view, initial L2 perception is based on the phonetic similarity of the L2 and L1 phones, rather than the abstract organisation of the L2 phones. PAM-L2 further posits that L2 vocabulary size may be directly related to the development of L2 phonology, as a larger L2 vocabulary may lead to fossilisation in the perception of L2 phones and curtail further reattunement and rephonologisation. Here, we revise and extend this view to suggest that L2 reattunement and rephonologisation may actually be facilitated by an expanding L2 vocabulary in the beginning period of L2 acquisition, specifically in L2-immersion settings. It is suggested that this reattunement and rephonologisation happens in the same way in which the emergence of L1 phonological function appears to be driven by the vocabulary explosion in children around 18+ months (Metsala & Walley, 1998). Specifically, it is argued that the establishment of an initial L2 vocabulary may ‘exert forceful linguistic pressure’ (Best & Tyler, 2007, p. 32) on the learner to ‘rephonologise’ and attune to articulatory, phonetic, and phonological differences in the L2 that have previously been ignored in the L1. This linguistic pressure may be particularly strong when several contrastive L2 phones are initially assimilated into the same L1 category and confusion subsequently arises from the activation of several competing lexical items in the learner’s L2 vocabulary (Cutler & Otake, 2004; Cutler, Weber, & Otake, 2006). This reattunement of the perceptual systems ideally leads to the formation of phonological categories that correspond (more closely) to
those in the L2, though it has been amply established that this is not always the case due to L1-L2 interactions.

The PAM-L2 vocabulary-growth hypothesis predicts that learners with a larger L2 vocabulary will be further along in the L2 acquisition process and thus will have better reattunement and rephonologisation, such that they more consistently assimilate L2 phones to their L1 categories – or to new L2 categories - and more fully exploit their native L1 inventory in their early identification of L2 phones. By consistent assimilation we refer to two behavioural measures: firstly, a significantly higher number of non-native phones assimilated into native categories by learners with larger L2 vocabularies; secondly, a higher assimilation score (% of L2-target tokens assimilated to a given L1 category) for assimilated phones in the learners with larger vocabularies. PAM-L2 also predicts that learners who are further along in their L2 acquisition process will be more likely to successfully discriminate non-native contrasts that are difficult for naïve and less proficient speakers of the L1 whose L2 perception is still dominated by superficial phonetic similarities between the L2 phones and the L1 categories. According to PAM-L2, this is particularly the case for CG, UC, and UU contrasts, and less so for SC contrasts, which may be rarer for vowels than for consonants, as learners and native speakers alike remain sensitive to even small within-category vowel differences. Conversely, less proficient listeners are also less likely to show difficulties for TC contrasts, which even very inexperienced learners have already successfully teased apart simply by reference to their L1 vowel contrasts.

One recent study found support for the role of L2 vocabulary expansion in changes to L2 vowel perception (Chapter 5: Bundgaard-Nielsen, Best, Tyler, & Kroos, submitted; see also Bundgaard-Nielsen, Best & Tyler, 2008). In that study, native Japanese learners of Australian English (AusE) (N = 11), with less than 12 weeks of immersion (M = 7 weeks) in Australia, mapped all 18 AusE vowels (monophthongs /i:, i, e, e:, æ, ə, ø, ɔ, o, u, u:/ and diphthongs /θə, æθi, əθe, əθi, əθu, æθi/ (presented in a /hVba/ nonsense word in Sentence (S) and Citation (C) form) to their native phonological system. They were allowed to select from all possible Japanese mono- and bi-moraic vowel categories (/i, i:, e, e:, a, ã, u, ɔ:/, and all permissible pair-wise combinations of mono-moraic vowels
The participants also completed a vocabulary size test (Nation & Beglar, 2007), which indicated that they had an average English vocabulary of 6009 words (range = 4100-7800 words). Learners with larger L2 vocabularies (>6000 words; $M = 7200$) showed more consistent use of their L1 categories in identifying L2 vowels (13 consistently assimilated vowels: $M \% S$ identification = 85%; $M \% C$ identification = 80%), in comparison with learners with smaller vocabularies ($M = 5017$ words; 12 consistently assimilated vowels: $M \% S$ identification = 73%; $M \% C$ identification = 74%). The two subgroups, importantly, were otherwise comparable in terms of duration of L2 immersion in Sydney, numbers of years of English instruction in Japan, age at testing, and L1/L2 usage patterns at the time of testing.

To confirm the relationship between L2 segmental perception and L2 vocabulary size and to further detail this developmental trajectory, we present two studies that probe the underpinnings of that pattern more deeply and broadly. The first study substantially extends the previous study (Chapter 5: Bundgaard-Nielsen et al., submitted) of the relationship between L2 vocabulary size and L2 vowel assimilation and discrimination in speakers of Japanese who are acquiring AusE through immersion. As indicated by the results from the previous study, Japanese L2 learners of English are an extremely well-suited group to test because they (surprisingly) exploit a large number of L1 vowel categories and combinations in the assimilation of L2 English vowels. We posit that larger vocabularies will be correlated with better L2 perception in terms of better L2 vowel contrast discrimination, as well as more consistent L2 assimilation to L1 categories. Furthermore, we expect to find systematic relationships between assimilation patterns and discrimination, as predicted by PAM (e.g., Best, 1995).

The second study examines the developmental changes in L2 perception and vocabulary size in the first 6-8 months of immersion acquisition in a group of Japanese L2 English learners. We predict an increase in L2 vocabulary size as well as associated improvements in the percentage of correctly discriminated AusE vowel contrasts and in the consistency of the L2-L1 assimilations between the two testing times.
6.3 Experiment 1: Vocabulary Size and L2 Segmental Perception in Newly-Immersed Adult L2 learners

6.3.1 Method

6.3.1.1 Participants

The participant group consisted of 31 native speakers of Japanese (8 male; $M$ age = 23.8 years, range 19-34 years), who were studying English as a second language at an English college in Sydney or following a General Studies course at University of Technology or University of New South Wales in Sydney. All had spent less than three months in Australia ($M_{\text{stay}}$ = 6 weeks, range 1-12 weeks), but had received English instruction in Japan, typically by native speakers of Japanese ($M_{\text{English FL instruction}}$ = 9 years; range = 6-18 years), but with occasional input from native speakers of various dialects of English (Australian, American, British, Irish, and New Zealand dialects of English). Data from an additional 7 Japanese participants were discarded from the analyses as they had been in Australia for longer than 12 weeks at time of testing, but had failed to understand this restriction in the pre-test interview and only divulged the information in the post-session discussion of the answers they provided on their Language Background Questionnaire.

6.3.1.2 Stimuli

The stimuli used in this study were the citation-form disyllables (C) used in Bundgaard-Nielsen et al. (submitted): Eight male L1 Western Sydney Australian English (AusEWS) speakers, ($M$ age = 22.7 years) produced five randomised repetitions of the AusE stressed monophthongs ($\hat{\iota}$, i, e, e′, ə, æ, ə, e′, o′, u′, o/ and diphthongs /ɪə, əɪ, æe, ər, æu, æʊ/) in a first syllable stressed /hVba/ context in isolation. The /hVba/ context was chosen to minimise any effect of consonant co-articulation on the vowels (Strange, Weber, Levy, Shafiro, Hisagi, & Nishi, 2007). The speakers were instructed to use an informal speaking style, and speak at a normal conversational speaking rate. Three speakers were selected on the basis of auditory judgements of similar voice quality, reading style and speed by two phonetically trained experimenters, and three tokens of each vowel were selected for the perceptual experiment from each of those three speakers on the basis of similar intonation pattern and speaking rate. The average word duration was 507 ms (range: 416 ms for
/hube/ to 601 ms for /haebə/). The audio recording took place in a sound-attenuated room at the University of Western Sydney, using a Shure SM10A headset microphone, an LG laptop computer and an external soundcard (Edirol UA-25). The recordings were high-pass filtered in Cool Edit with a cut-off frequency of 75 Hz to eliminate a low-frequency rumble from the recording. This did not impair the intonation of the utterances, as none of the F0 contours for the selected tokens dipped below 75 Hz.

The average RMS intensity of the target vowel in each word was adjusted in Praat (Boersma & Weenink, 2008) to be equal to 70dB. Vowel onset was defined as the beginning of the first regular and recognisable pitch pulse, and vowel offset was defined as the cessation of regular pitch pulsing. Following previous research (e.g., Strange et al., 1998), vowel duration as well as F1, F2, and F3 values at 25%, 50%, and 75% of the target vowel of the first syllable were estimated. The formant values were estimated using the Praat command ‘To Formant (burg)’. Time-step was 2.5 ms, maximum number of formants 5, maximum formant frequency 5 kHz, with a 25 ms window length. Pre-emphasis was from 50 Hz. The obtained measurements are shown in Table 6.1. The formant values reported in Table 6.1 are consistent with the vowel data reported by Cox (2006), with the exception of the diphthong she transcribes as /æu/. In our data, this diphthong is produced with a higher (and slightly more fronted) first target than /æ/ and is therefore transcribed as /æt/ in Table 6.1.

Four AusE vowel contrasts (from the 18 AusE vowels) were selected for discrimination by the Japanese participants. These contrasts were selected on the basis of the assimilation results from the previous study (Chapter 5: Bundgaard et al., submitted), and were selected to test the discrimination predictions for four different PAM contrast types (see, e.g., Best, 1995): the Single Category (SC) contrast /iː/-/æ/, the Two-Category (TC) contrast /iː/-/æt/, the Uncategorised-Categorised (UC) contrast /æ/-/æt/, and the Uncategorised-Uncategorised (UU) contrast /æu/-/æ/. The previous study did not indicate any likely Category Goodness difference (CG) contrast for Japanese listeners. Stimuli were presented over headphones from a MacBook laptop computer using PsyScope X software.
Table 6.1. Citation vowel durations (ms), word duration (ms) and F1, F2, and F3 values (in Hz) at 25%, 50%, and 75% of the target monopthongs and diphthongs of Australian English (AusE).
6.3.2 Procedure

All participants were tested in a quiet, sound-attenuated experiment room at MARCS Auditory Laboratories at University of the Western Sydney.

The participants first heard a randomised presentation of all AusE WS disyllables in citation form (N = 324). They identified each AusE token’s stressed (first) vowel in terms of their L1 vowels and permissible combinations on a computer screen, using a grid of Japanese katakana symbols representing all possible short, long and combination-vowels in a /hV/ context: /i, i:, e, e:, a, a:, u, u:, o, o:, ie, ia, iu, io, ei, ea, eu, eo, ai, ae, au, ao, ui, ue, ua, uo, oi, oe, oa, ou/ and rated the goodness-of-fit of the AusE WS token to the L1 vowel they had selected, on a scale from 1 (poor) to 7 (excellent).

Following the assimilation experiment, all participants discriminated the AusE contrast pairs /i/-/iø/, /i/-/iø/, /i/-/iø/, and /iø/-/iø/ in an AXB task. Order of presentation of the four contrast pairs was counterbalanced across participants. Upon completion of the perceptual tasks, the participants completed a multiple-choice L2-English vocabulary size test (Nation & Beglar, 2007; See Chapter 5: Bundgaard-Nielsen et al., submitted).

6.3.3 Predictions

The general assimilation pattern of AusE vowels to Japanese vowel categories was expected to reflect the perceptual similarities between the learners’ two vowel systems, such that short AusE vowels (/i, e, e, a, u/) would be assimilated into perceptually similar single-mora Japanese vowel categories (/i, e, a, u/, respectively), while long AusE vowels (/i:, e:, o:, u/) would most likely be assimilated into perceptually similar bi-moraic Japanese vowel categories (/i:, e:, a:, o:, u/, respectively). Also, in parallel with the results from our previous study (Bundgaard-Nielsen et al., submitted, See Chapter 5), some AusE diphthongs (such as /ei, ae, o/) were expected to be readily assimilated into Japanese bi-moraic combinations (such as /ei, ai, o/, respectively).

More importantly for the present investigation of the association between L2 vocabulary size and L2 segmental perception, the L2-L1 assimilation consistency (in terms of the number of assimilated L2 vowels and average assimilation percentage) was expected to be
higher for those Japanese learners who had larger L2 vocabularies than for learners with smaller L2 vocabularies, as the former participants were expected to be at an earlier stage of their L2 perceptual reattunement and rephonologisation process than learners with larger L2 vocabularies. This prediction was expected to be borne out, even though the two Japanese subgroups did not differ on any other potentially relevant dimension (See Table 6.3, below).

With reference to PAM, the overall discrimination of the SC contrast /i:/-/ιa/ was expected to be near chance level. This expectation is further supported by recent evidence for a near-merger of these two vowels in Western Sydney AusE (See Appendices 1 and 4). For the spectrally similar, but durationally different, TC contrast /i:/-/ιt/, discrimination was predicted to be excellent. For the UC contrast /u:/-/υt:/, discrimination was expected to be good to excellent; we would expect excellent discrimination in the case where there is a clear native category boundary between the two phones, i.e., an L1 phonological difference, while we would expect lesser discrimination accuracy for a case in which the uncategorised vowel in the pair shares some perceptual similarity with the native phone to which the categorised phone in the UC pair is assimilated. Finally, discrimination of the UU contrast /œu/-/œy/ was expected to be moderate (reflecting some degree of overlap in the identification pattern for these two vowels in the previous study. See Chapter 5, as well as Table 6.2 below, summarising the identification patterns for the UU contrast).

The overlap in the identification patterns for /œu/-/œy/ was interesting, and provided an excellent opportunity to tease apart PAM/PAM-L2 and SLM. In particular, as these two vowels seem quite distinct, both in terms of acoustics and articulation, psychoacoustic theories such as SLM should predict that they should be very well discriminated by the learners. The identification patterns revealed by the previous study, however, highlighted that discrimination might not be straightforwardly predicted on the basis of psychoacoustics. Rather, and despite the measurable acoustic-phonetic differences between these vowels, the nonnative listeners still perceived them as similar to a largely overlapping array of native phonological categories, leading to the prediction of only modest, rather than very good, discrimination.
In parallel with the expectation for greater assimilation consistency for learners with a larger L2 vocabulary than those with a smaller L2 vocabulary, discrimination accuracy for learners with a larger L2 vocabulary was expected to be greater than for those with a smaller vocabulary. This vocabulary size effect was not expected to extend to the discrimination of the SC /ɪː/-/ɪə/ contrast, as PAM-L2, as well as the proposed extension of PAM-L2 to vocabulary-driven L2 reattunement and rephonologisation (See Chapter 3) predicts great difficulty in reattuning to differences between the phones in an SC contrast.

We also expect L2 vocabulary size to have a smaller effect on the TC /ʃ:/-/θ/, as this contrast type, according to PAM/PAM-L2 and the Vocab Model, is already well differentiated by even the least-experienced and smallest-vocabulary L2 learners.

**Table 6.2.** Perceptual assimilation pattern for the UU contrast /oː/-/ʊə/, adapted from Chapter 5. The table presents the top three selected L1 Japanese vowel categories (columns) for the two uncategorised AusE<sub>WS</sub> vowels (rows). Group refers to High Vocabulary learners (HV) and Low Vocabulary learners (LV) of the particular study (Chapter 5). The % columns indicate the total percent of each of the uncategorised AusE<sub>WS</sub> vowels which were identified as each of the top three L1 categories.

<table>
<thead>
<tr>
<th>AusE Vowel</th>
<th>Context</th>
<th>Group</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>/oː/</td>
<td>Citation</td>
<td>HV</td>
<td>42</td>
<td>/o/</td>
<td>32</td>
<td>/oː/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>47</td>
<td>/oː/</td>
<td>28</td>
<td>/oː/</td>
</tr>
<tr>
<td>/ʊə/</td>
<td>Sentence</td>
<td>HV</td>
<td>48</td>
<td>/uː/</td>
<td>13</td>
<td>/aː/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>26</td>
<td>/uː/</td>
<td>19</td>
<td>/oː/</td>
</tr>
<tr>
<td>/ʊː/</td>
<td>Citation</td>
<td>HV</td>
<td>41</td>
<td>/uː/</td>
<td>19</td>
<td>/auː/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LV</td>
<td>31</td>
<td>/aː/</td>
<td>10</td>
<td>/auː/</td>
</tr>
</tbody>
</table>
6.3.4 Results

6.3.4.1 Vocabulary Size

The vocabulary size test indicated that the 31 participants had a mean vocabulary size of 6452 words (range: 4600-8700 words). We split the participants into a High Vocabulary (HV) and a Low Vocabulary (LV) group with 6000 words as the threshold, following evidence that comprehension of a minimum of 6000-7000 English words is needed to sufficiently understand spoken conversational English (Nation, 2006). Significant vocabulary-related differences in L2 segmental perception for speakers with an L2 vocabulary above (HV) and below (LV) 6000 words had also been found in the previous experiment (Chapter 5: Bundgaard-Nielsen et al., submitted).

The LV L2 group consisted of 12 participants who had a mean vocabulary of 5250 words (range: 4600-5900), while the HV L2 group consisted of 19 learners who had an average L2 vocabulary of 7211 words (range: 6300-8700). A series of One-Way ANOVAs (see Table 6.3) revealed that this difference in vocabulary size was significant, while there were no significant differences between the two groups in terms of their Age of English Immersion, the age at which they had started to learn English in Japan, the number of years they had learned English in Japan, or the duration of their L2 immersion in Sydney, Australia.

Table 6.3. Vocabulary Group comparisons in mean Length of Stay (LOS) in Australia, Age of Immersion (AOI) in English, Age of Learning (AOL) onset, number of years of L2 study (Study), and L2 vocabulary size (Vocab).

<table>
<thead>
<tr>
<th>Comparison</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOS</td>
<td>1</td>
<td>10.856</td>
<td>1.139</td>
<td>.295</td>
</tr>
<tr>
<td>AOI</td>
<td>1</td>
<td>8.562</td>
<td>.485</td>
<td>.492</td>
</tr>
<tr>
<td>AOL</td>
<td>1</td>
<td>4.041</td>
<td>2.579</td>
<td>.119</td>
</tr>
<tr>
<td>Study</td>
<td>1</td>
<td>13.509</td>
<td>2.284</td>
<td>.142</td>
</tr>
<tr>
<td>Vocab</td>
<td>1</td>
<td>28269524.618</td>
<td>64.310</td>
<td>.000</td>
</tr>
</tbody>
</table>
6.3.4.2 Assimilation and Goodness Ratings

6.3.4.2.1 Assimilation Criterion of 50%

An L2 vowel was defined as ‘categorised’ if it was assimilated as an instance of a particular L1 category in more than 50% of instances, thus a fairly lenient criterion. We chose this lenient criterion in recognition of the complexity of the task presented to the Japanese learners of selecting, for each AusE vowel, from a grid of a total of 35 Japanese katakana symbols. Due to the high consistency between the participants’ individual assimilation patterns, we considered an L2 vowel assimilated if the average of all participants’ assimilations reached the 50% criterion, even if not every individual participant reached the assimilation criterion (following Chapter 5: Bundgaard-Nielsen et al., submitted). An assimilation matrix based on all 31 participants’ responses revealed that a total of 14 out of 18 AusE vowels (/ɪ, e, ɛ, ɔ, u, iː, ɪə, eː, ɛː, oː, uː, əe, əɪ, əʊ/) were successfully assimilated to Japanese vowel categories, while only four vowels (/ʒ, əe, əʊ, əu/) remained unassimilated (Uncategorized). The average assimilation score for each assimilated vowel ranged from 63% (for AusE /ʊ/) to 92% (for AusE /eː/), with an average assimilation score of 75%, for the 14 assimilated AusE vowels. The average Goodness Rating (i.e., the goodness of fit of the non-native phone to the native category) for those assimilated vowels was 5.12 (range = 4.75 - 5.46) out of a maximum of 7.

When the participants were divided into a HV and a LV group, systematic differences in their assimilation behaviour emerged. The HV group (see Table 6.4) successfully assimilated 15 AusE vowels (/ɪ, e, ɛ, ɔ, u, iː, ɪə, eː, ɛː, oː, uː, əe, əɪ, əʊ, əʊ/) with an average assimilation percentage of 77% (range = 53% for AusE /ɪ/ to 93% for AusE /eː/). The average Goodness Rating was 5.09 out of a maximum of 7. Only three vowels (/ʒ, əe, əu/) remained unassimilated by the HV group. The LV group (see Table 6.5) assimilated 14 AusE vowels to native Japanese categories (/ɪ, e, ɛ, ɔ, u, iː, ɪə, eː, ɛː, oː, uː, əe, əɪ, əʊ/) but not AusE /æə/), but their average assimilation percentage was significantly lower than that of the HV group, at only 68% (range = 52% for AusE /ɪ/ to 90% for AusE /eː/; paired t-test between HV and LV means: t(13) = 3.02, p = 0.01). The average Goodness Rating for the LV group was 5.03, which was not different to that of the HV group (paired t-test, t(13) = .589, ns).
Table 6.4. Assimilation pattern (percentage of each Australian English vowel assimilated to specific Japanese vowel categories, averaged over participants) and Goodness Ratings (on a scale of 1-7) for the High Vocabulary (HV) group with a 50% assimilation criterion. Grey indicates an Uncategorised L2 vowel.

<table>
<thead>
<tr>
<th>Category</th>
<th>/i:/</th>
<th>/ɪː/</th>
<th>/æː/</th>
<th>/æːl</th>
<th>/ɛː/</th>
<th>/æːl</th>
<th>/æːl</th>
<th>/oː/</th>
<th>/oːl</th>
<th>/uː/</th>
<th>/uːl</th>
<th>GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ɪː/</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.94</td>
</tr>
<tr>
<td>/ɪː/</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.53</td>
</tr>
<tr>
<td>/æː/</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.98</td>
</tr>
<tr>
<td>/æː/</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.37</td>
</tr>
<tr>
<td>/eː:/</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.84</td>
</tr>
<tr>
<td>/æː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.23</td>
</tr>
<tr>
<td>/æː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.15</td>
</tr>
<tr>
<td>/æːl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.14</td>
</tr>
<tr>
<td>/æːl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.39</td>
</tr>
<tr>
<td>/eː:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.92</td>
</tr>
<tr>
<td>/oː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.24</td>
</tr>
<tr>
<td>/oː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.99</td>
</tr>
<tr>
<td>/oːl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.51</td>
</tr>
<tr>
<td>/uːl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.38</td>
</tr>
<tr>
<td>/uːl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.76</td>
</tr>
<tr>
<td>/æːl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɔːl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6.5. Assimilation pattern (percentage of each Australian English vowel assimilated to specific Japanese vowel categories, averaged over participants) and Goodness Ratings (on a scale of 1-7) for the Low Vocabulary (LV) group with a 50% assimilation criterion. Grey indicates an Uncategorised L2 vowel.

<table>
<thead>
<tr>
<th>Category</th>
<th>/i/</th>
<th>/ɪ:/</th>
<th>/æ/</th>
<th>/ʌ/</th>
<th>/ɑ:/</th>
<th>/ao:/</th>
<th>/æ:/</th>
<th>/o:/</th>
<th>/ʊ:/</th>
<th>GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.06</td>
</tr>
<tr>
<td>/ɪ:/</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.29</td>
</tr>
<tr>
<td>/æ/</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.97</td>
</tr>
<tr>
<td>/ʌ/</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.19</td>
</tr>
<tr>
<td>/ɑ:/</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.72</td>
</tr>
<tr>
<td>/æ:/</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.91</td>
</tr>
<tr>
<td>/o:/</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.45</td>
</tr>
<tr>
<td>/ʊ:/</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.21</td>
</tr>
<tr>
<td>/ʌ:/</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.55</td>
</tr>
<tr>
<td>/ɑ:/</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.21</td>
</tr>
<tr>
<td>/æ:/</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.78</td>
</tr>
<tr>
<td>/o:/</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.28</td>
</tr>
<tr>
<td>/ʊ:/</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.11</td>
</tr>
<tr>
<td>/ʌ:/</td>
<td></td>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.74</td>
</tr>
</tbody>
</table>
6.3.4.2.2 Assimilation Criterion of 70%

Recognizing the lenient nature of the 50% assimilation criterion used above differs from the approach in many other studies (Strange et al., 1998, for instance, used modal categories), we reanalysed the assimilation pattern of the two Japanese learner groups with a more stringent 70% assimilation criterion.

Applying the more stringent 70% assimilation criterion (see Tables 6.6 and 6.7), the group differences in the number of L2 AusE vowels successfully assimilated is much more striking, with the HV group assimilating 12 of the 18 vowels \( (M_{\text{assimilation}} = 82\%, \text{ range } 74\% \text{ for AusE } /u:/ \text{ to } 93\% \text{ for AusE } /e:/) \), but the LV group only successfully assimilating 6 AusE vowels \( (M_{\text{assimilation}} = 78\%, \text{ range } 72\% \text{ for } /u:/ \text{ to } 90\% \text{ for } /e:/) \).
Table 6.6. Assimilation pattern (percentage of each Australian English vowel assimilated to specific Japanese vowel categories, averaged over participants) and Goodness Ratings (on a scale of 1-7) for the High Vocabulary (HV) group with a 70% assimilation criterion. Grey indicates an Uncategorised L2 vowel.

<table>
<thead>
<tr>
<th>Category</th>
<th>M</th>
<th>/i:ə/</th>
<th>/e:/</th>
<th>/ɛː/</th>
<th>/æ/</th>
<th>/ɛ/</th>
<th>/ə/</th>
<th>/æ:/</th>
<th>/ɪ/</th>
<th>/ɛ/</th>
<th>/i:ɪ/</th>
<th>GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iː:/</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.94</td>
</tr>
<tr>
<td>/ɪː:/</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.53</td>
</tr>
<tr>
<td>/æː:/</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.98</td>
</tr>
<tr>
<td>/ɛː:/</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.37</td>
</tr>
<tr>
<td>/æːː/</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.84</td>
</tr>
<tr>
<td>/æː/</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.23</td>
</tr>
<tr>
<td>/æːː/</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.14</td>
</tr>
<tr>
<td>/æːː/</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.39</td>
</tr>
<tr>
<td>/æːː/</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.92</td>
</tr>
<tr>
<td>/æːː/</td>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.24</td>
</tr>
<tr>
<td>/æːː/</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.51</td>
</tr>
<tr>
<td>/uː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/uːː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>74</td>
</tr>
<tr>
<td>/æːː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.76</td>
</tr>
</tbody>
</table>
Table 6.7. Assimilation pattern (percentage of each Australian English vowel assimilated to specific Japanese vowel categories, averaged over participants) and Goodness Ratings (on a scale of 1-7) for the Low Vocabulary (LV) group with a 70% assimilation criterion. Grey indicates an Uncategorised L2 vowel.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/iː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>73</td>
</tr>
<tr>
<td>/ɪ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.06</td>
</tr>
<tr>
<td>/æː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>85</td>
</tr>
<tr>
<td>/æ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.97</td>
</tr>
<tr>
<td>/æː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>77</td>
</tr>
<tr>
<td>/ɛː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.91</td>
</tr>
<tr>
<td>/ɛː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>90</td>
</tr>
<tr>
<td>/ɛː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.55</td>
</tr>
<tr>
<td>/oː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>73</td>
</tr>
<tr>
<td>/oː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.28</td>
</tr>
<tr>
<td>/uː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>72</td>
</tr>
<tr>
<td>/uː/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.74</td>
</tr>
</tbody>
</table>

89
6.3.4.3 Discrimination

On the basis of the assimilation patterns from the previous study, and confirmed by the pattern of assimilation in the present study, the contrast /iː/-/ɪə/ was predicted to be a SC assimilation with very poor discriminability. Similarly, the contrast /iː/-/ɪə/ was confirmed to be TC assimilation and discrimination was thus expected to be excellent. The contrast /æː/-/ɑː/ was also confirmed to be an UC assimilation by the assimilation pattern identified in this study and discrimination expected to be good, and, finally, the assimilation pattern for both subgroups of Japanese learners in this study revealed that the predicted UU assimilation of /œːl/-/oː/ was actually a UC contrast for them (henceforth UU/UC, as the long vowel /oː/ reached the 50% assimilation criterion and was categorised as long Japanese /oː/). Discrimination for this contrast was expected to be fair to good. We chose to apply this 50% criterion, despite observations that native-like performance is often considered to be discrimination above 90%, for two reasons: One, we suspect that a 90% (or even 70% criterion) may be too strict for vowel research, due to the continuous nature of vowel perception; and two, because imposing a more stringent criterion may lead to even native speakers appearing to have difficulty systematically identifying vowels in their own language (See Appendix 1: Bundgaard-Nielsen, Best and Tyler, submitted).

As predicted by the perceptual assimilation patterns laid out by PAM, the combined results from the 31 participants revealed good discrimination of the TC contrast /iː/-/ɪə/ (79% correct discrimination), the UC contrast /æː/-/ɑː/ (83% correct discrimination), and fair discrimination of the UU/UC contrast /œːl/-/oː/ (76% correct discrimination). The discrimination pattern for /œːl/-/oː/, however, was not readily compatible with predictions based on psychoacoustic models, such as SLM, which would have expected these acoustically rather different vowels to be discriminated easily. The discrimination for the predicted SC assimilation /iː/-/ɪə/ was instead at chance level (50% correct discrimination), as PAM-L2 predicted based on the UU assimilation pattern that involved highly overlapping L1 vowel categorizations for this pair of English vowels.

A 2x(2) ANOVA was conducted on the percent correct discrimination scores. There was one between-group factor, L2 vocabulary size, with two levels (HV and LV) and one
within-group factor, contrast, with four levels (SC, TC, UC, and UU/UC). An alpha level of .05 was used for all analyses unless specified. Mean percent correct discrimination scores, split by L2 vocabulary size and by contrast, are presented in Figure 6.1. There were significant main effects of contrast, $F(3,87) = 78.71$, $p < .001$, $\eta^2 = .73$, and L2 vocabulary size, $F(1,29) = 4.44$, $p = .044$ $\eta^2 = .13$, but there was no significant interaction, $F(3,87) = 0.90$, ns.

To test for differences between discrimination of the four contrasts, we conducted six post-hoc paired t-tests with a Bonferroni-adjusted alpha level of .008. As there was no interaction between contrast and L2 vocabulary size in the 2x(2) ANOVA, the two vocabulary groups were combined for these analyses, which are presented in Table 6.8. This effect of contrast was due to the difference in discrimination scores between SC /iː/-/tÆ/ and the three other contrasts TC /iː/-/t\textipa{3}/, UC /\textipa{3}:/-/u\textipa{3}/, and UU/UC /\textipa{3}ː/-/ɔ/; as well as a difference in the discrimination of the UC /\textipa{3}:/-/u\textipa{3}/ and UU/UC /\textipa{3}ː/-/ɔ/ contrasts, but not for the TC /iː/-/t\textipa{3}/ versus UC /\textipa{3}:/-/u\textipa{3}/ contrast, ns.

**Table 6.8.** Paired t-tests of the significance (Bonferroni-corrected) of discrimination difference of the four Australian English vowel contrasts SC, TC, UC, and UC.

<table>
<thead>
<tr>
<th>Contrast</th>
<th>$t$ (30)</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC-TC</td>
<td>13.45</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SC-UC</td>
<td>13.25</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SC-UC</td>
<td>13.83</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TC-UC</td>
<td>1.50</td>
<td>=.14</td>
</tr>
<tr>
<td>TC-UC</td>
<td>1.51</td>
<td>=.14</td>
</tr>
<tr>
<td>UC-UC</td>
<td>3.05</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>
Figure 6.1. Discrimination accuracy for the contrasts SC (ɪː/ɪɛ/), TC (ɪː/ɪʌ/), UC (ɪː/ɪʊ/), UU/UC (ɪəʊ/-
ɪə/), overall discrimination score (Overall) as well as overall discrimination score without the SC contrast 
(Overall-SC) for the two vocabulary size groups, HV and LV. Error bars reflect S.E.M.

6.3.5 Discussion

As expected, the overall assimilation pattern of AusE vowels to Japanese vowel categories 
reflected the perceptual similarities between the two vowel systems, such that short AusE 
vowels were assimilated to single mora Japanese vowels of ‘similar’ spectral qualities, 
while long AusE vowels were assimilated to bi-moraic Japanese vowels of ‘similar’ 
spectral qualities, and some AusE diphthongs were assimilated to bi-moraic combinations 
of two Japanese mono-moraic vowels. The general assimilation pattern thus reflects 
previous findings with Japanese learners of AusE (Chapter 5: Bundgaard-Nielsen et al, 
submitted), and, to some extent, also findings with native Japanese learners of American 
English (Strange, Akahane-Yamada, Kubo, Trent, Nishi, & Jenkins 1998), where L2-L1 
assimilation was also partially predicted by spectral similarities. That previous study, 
however, had not provided the learners with the full set of native single, bi-moraic, and bi-
moraic combinations available to the participants in the current study, perhaps limiting
their responses. Finally, Strange et al. also found that their Japanese learners did not typically use bi-moraic native categories for American English vowels presented in words in citation, but only in sentence presentation. This pattern was not observed for Japanese learners of AusE in the present study, in which the HV group assimilated eight out of 18 of the AusE vowels in citation-form nonsense disyllables, to Japanese bi-moraic vowels or bi-moraic combinations, nor in the previous study (Chapter 5: Bundgaard-Nielsen et al, submitted) in which AusE vowels were presented in disyllabic nonsense words both in citation and in sentence-embeddings.

One noteworthy difference in the overall assimilation pattern presented here and in the previous study is that AusE /o:/ was successfully identified as Japanese /o:/ in this study (HV 68%, LV 62% assimilation), while it only approached assimilation in the previous study (see previous study’s findings in Table 6.9).

**Table 6.9.** Most frequently selected Japanese vowel category, in percent, across participants, for assimilation of AusE /o:/ in both Citation and Sentence presentation in Bundgaard-Nielsen et al., submitted (Chapter 5).

<table>
<thead>
<tr>
<th>Context</th>
<th>Group</th>
<th>Category</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV</td>
<td>/ou/</td>
<td>42%</td>
<td></td>
</tr>
<tr>
<td>Citation</td>
<td>LV</td>
<td>/o:/</td>
<td>47%</td>
</tr>
<tr>
<td></td>
<td>HV</td>
<td>/o:/</td>
<td>42%</td>
</tr>
<tr>
<td>Sentence</td>
<td>LV</td>
<td>/o:/</td>
<td>44%</td>
</tr>
</tbody>
</table>

With the 70% assimilation criterion, however, AusE /o:/ is no longer categorised in either study. A likely explanation for the difference in results between the present study and the previous is a more advanced perceptual reattunement and rephonologisation process in the learners of the present study, possibly due to the reported larger average L2 vocabulary of the learners in this study, due to random sampling variations. This is compatible with the predictions of PAM-L2 that learners with larger L2 vocabularies will perform better or
more consistently on L2 segmental assimilation tasks and better on L2 segmental discrimination tasks than learners with smaller L2 vocabularies.

The clear difference in assimilation consistency for the LV group and the HV group, and the difference in the number of L2 vowels assimilated to L1 categories by the HV and LV groups, is consistent with the PAM-L2 hypothesis that L2 vocabulary expansion is correlated with changes to L2 segmental perception, reflecting better reattunement and rephonologisation. Applying the more stringent 70% assimilation criterion only enhances this observed difference by greatly decreasing the number of categorised AusE vowels for the LV group from 14 categorised vowels to only six, while only decreasing in the number of categorised vowels for the HV group from 15 to 12.

The overall pattern of discrimination accuracy for the four contrasts in the discrimination task reflected our PAM-based predictions for the perceived similarity of the AusE L2 phones relative to the L1 Japanese vowel inventory. Again, the HV and LV group difference in overall discrimination accuracy was associated with L2 vocabulary size, supporting the PAM-L2 hypothesis about the role of L2 vocabulary expansion in L2 segmental reattunement and rephonologisation.

In summary, this experiment confirmed and extended the relationship between L2 vocabulary size and L2 segmental perception, in which L2 vocabulary may be seen as connected to the reattunement of the perception of L2 phones. The study also shows that L2 assimilation and relative ease of discrimination of the chosen L2 contrasts reflect the perceived similarities between the L1 and L2, as predicted by the Perceptual Assimilation Model. These general patterns remain consistent as L2 vocabulary and segmental perception develop, in line with many results confirming a pervasive effect of the perceptual relationship between L1 and L2 phones even in advanced learners. However, in order to systematically examine change in L2 vocabulary size and L2 vocabulary in the first half-year of L2 immersion, we also conducted a longitudinal experiment with native Japanese learners of English.
6.4 Experiment 2: Change in L2 Vowel Perception in L2 Learners in the First Six Months of L2 Immersion

The present experiment examines the role of vocabulary size on segmental perception in L2 learners over the course of the first six to eight months of L2 immersion. It tests two contrasting hypotheses. The first hypothesis is based on previous observations (see for instance Tsukada et al., 2005) that extended L2 exposure and use, and limited L1 use, leads to improved L2 perception in the early phase of adult L2 acquisition. This hypothesis predicts that L2 learners, irrespective of their L2 vocabulary size, are likely to improve their L2 perception in the first 6-12 months of L2 immersion.

The second hypothesis is consistent with PAM-L2, and in particular with the Vocab Model (see Chapter 3). It instead predicts that improvement in L2 perception is associated with L2 vocabulary expansion in L2 learners with small L2 vocabularies. This hypothesis predicts the complementary outcome that, despite increased immersion duration, L2 learners with larger (>6000) L2 vocabularies at the beginning of immersion will not improve their L2 perception.

6.4.1 Method

6.4.1.1 Participants

All 31 participants from Experiment 1 were invited to return to UWS for a second testing session approximately 6 months after their arrival in Australia, and at least three months after the first testing session. Most of the participants from Experiment 1, however, had returned to Japan after 1-5 months of immersion and so could not return for a second test session. A total of 10 remaining Japanese participants (3 male; $M_{\text{age}} = 23.2$ years) agreed to participate in a second testing session. All participants were studying either English as a second language at one of the English L2 colleges in Sydney, or were first year students of General Studies courses at University of Technology, Sydney, or University of New South Wales in Sydney. One of these 10 participants was discarded from our analyses due to not consistently following test instructions during the second test session.
The nine remaining Japanese participants had an average vocabulary size of 6722 words (range: 5200-8100) in the first testing session, when they had spent an average of 5.45 weeks (range: 4-8 weeks) in Sydney. Seven of the nine participants had vocabularies above 6000 words at that first test session, placing them in the HV group of study 1.

At the time of the second testing session, the nine participants had spent an average of 27.8 weeks (range: 22-34 weeks) in Australia and had an average L2 vocabulary size of 7400 words (range: 5600-8700), with only one of the nine participants still having an L2 vocabulary under 6000 words. This reflected a significant average vocabulary gain of 678 words in just over 22 weeks of additional L2 immersion (paired samples t-test of the difference in L2 vocabulary size: $F(1, 8) = 3.516, p = 0.008$).

6.4.1.2 Stimuli and Procedure

The stimulus materials used in Experiment 2 were identical to those in Experiment 1. Similarly, all testing was conducted in a quiet room at MARCS Auditory Laboratories, University of Western Sydney, with the exception of the second test sessions for two participants which were conducted in quiet study room at the UTS library in Sydney.

6.4.2 Predictions

We expected the general L2 assimilation pattern to reflect the perceptual similarities and differences of the learners’ specific L1-L2 combination at both testing times. The predictions were the same as for Experiment 1.

In parallel with the predictions for the HV-LV group differences in Experiment 1, we expect poorer assimilation consistency for learners at test Time 1 than at test Time 2, when we expect the learners to have acquired a larger L2 vocabulary and to have further re-attuned and rephonologised. This effect of extended immersion and furthered L2 vocabulary acquisition and L2 vowel rephonologisation may not extend to the discrimination of the SC /i:/-/ɪə/ contrast, as PAM predicts great difficulty in reattuning to differences between the phones in an SC contrast, even after extended exposure. We also expect the additional time in Australia to have a small effect on the TC /iː/-/ɪə/, as this
contrast type, according to PAM, is already well differentiated by even beginning L2 learners.

We further expected the discrimination of the four L2 contrasts to follow the predictions outlined in PAM, such that the SC contrast /i:/-/æ/ was expected to be at chance level; that discrimination of durationally different, but spectrally similar, TC contrast /i:/-/u/, discrimination would be excellent; and that discrimination of the UC contrast /ɔ:/-/u:/ would be moderate to good. Finally, the discrimination of the predicted UU contrast /ɛu/-/o:/ was expected to be either poor to moderate if both vowels were un-categorised by this particular group of learners, or moderate to good, if it was indeed revealed to be another UC contrast, as was the case in Experiment 1 of the present paper with the 50% assimilation criterion. We expected that the average L2 vocabulary size of the participants of the present study to predict the classification of the /ɛu/-/o:/ contrast, so that a high average L2 vocabulary would be associated with the vowel /o:/ reaching the assimilation criterion of 50%, and perhaps even the more demanding 70% criterion.

As in Experiment 1 reported here, we expected that vocabulary size would be correlated with L2 segmental perception. Specifically, we expected that the average L2 vocabularies of the Japanese learners would increase over the course of their immersion, and that this vocabulary increase would lead to further L2 segmental reattunement and rephonologisation, reflected in more consistent L2 vowel assimilation to native categories and better discrimination of three of the four L2 contrasts selected for this experiment (TC /i:/-/u/, UC /ɔ:/-/u:/, and UU/UC /ɛu/-/o:/). Again, we did not expect improvement in the discrimination of the SC contrast /i:/-/æ/, as even native speakers of AusE fail to discriminate this contrast, despite significant acoustic differences in the vowel offglide (See Appendices 1 and 4).
6.4.3 Results

6.4.3.1 Assimilation and Goodness Rating

6.4.3.1.1 Time 1

Using the lenient 50% assimilation criterion for the averaged participants, the responses from these nine participants at the first test session revealed a total of 16 of 18 AusE vowels (/i, e, æ, e, ɔ, u, iː, ɪə, eː, ɔː, uː, ɒe, æɪ, æɝ, ʌɪ/) were successfully assimilated to Japanese vowel categories at the first testing session, while only two vowels (/ɪ, əʊ/) remained unassimilated (See Table 6.10 for assimilation pattern and Goodness Rating). The average assimilation score in percentages ranged from 57% (for /æɪ/, which was uncategorised in Experiment 1 reported above) to 92% for /ɛː/, /ɪəl/, and /ɛl/, with an average assimilation score of 81%, and an average Goodness Rating of 5.27. The assimilated vowels also included including AusE /ɔːl/, again changing the PAM contrast classification for /æʊl/ to /ɔːl/ from an UU to a second UC contrast as in Experiment 1.

Using the more stringent 70% assimilation criterion, a total of 14 AusE vowels (/i, e, e, ɔ, u, iː, ɪə, eː, ɔː, uː, ɒe, æɪ, ʌɪ/) were successfully assimilated to Japanese vowel categories (again, see Table 6.10), with an average assimilation score of 84% (range: 70% for /ɛː/ to 92% for /ɛː/, /ɪəl/, and /ɛl/). The average Goodness Rating was 5.35 out of 7.

6.4.3.1.2 Time 2

Applying the 50% criterion to the results from the second test session, an assimilation matrix based on the responses from the nine participants at the second test session revealed a total of 15 of 18 AusE vowels (/i, e, æ, e, ɔ, u, iː, ɪə, eː, ɔː, uː, ɒe, æɪ, ʌɪ/) were assimilated to Japanese vowel categories (see Table 6.11 for Assimilation pattern and Goodness Rating). The average assimilation score was 82% (range: 53% for /oː/ to 96% for /ɛ/), with an average Goodness Rating of 5.24 out of a maximum of 7. Three AusE vowels (/ʃ, æʃ, əʊ/) were not categorised, including AusE /æʊl/ which was categorised as Japanese /β/ (57%) in Experiment 1.

A reanalysis using the more stringent 70% assimilation criterion, however revealed only a total of 12 AusE vowels (/i, e, e, ɔ, u, iː, ɪə, eː, ɔː, uː, ɒe, æɪ, ʌɪ/) to be successfully
assimilated to Japanese categories at the second testing session with an average assimilation score of 87% (range: /e/ 73% to 96% for /e/) and an average GR of 5.26 out of 7.
Table 6.10. Assimilation pattern and Goodness Rating (GR) at Time 1 for the nine participants of Experiment 2. Light grey indicates an uncategorised L2 vowel using a 50% assimilation criterion. Dark grey indicates an uncategorised vowel using a 70% assimilation criterion.

<table>
<thead>
<tr>
<th>Category</th>
<th>/l/</th>
<th>/l:/</th>
<th>/el/</th>
<th>/el:/</th>
<th>/o/</th>
<th>/o:/</th>
<th>/o:/l</th>
<th>/o:/l</th>
<th>/u:/</th>
<th>/u:/l</th>
<th>/u:/l</th>
<th>GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>/l/</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/l:/</td>
<td>85</td>
<td>5.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/el/</td>
<td>92</td>
<td>5.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/el:/</td>
<td>92</td>
<td>5.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/e:/l</td>
<td>70</td>
<td>4.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/e:/l</td>
<td></td>
<td></td>
<td>57</td>
<td>5.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/e:/l</td>
<td>91</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a:/l</td>
<td></td>
<td></td>
<td>56</td>
<td>4.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a:/l</td>
<td></td>
<td></td>
<td>84</td>
<td>5.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/e:/l</td>
<td></td>
<td></td>
<td>87</td>
<td>5.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/e:/l</td>
<td></td>
<td></td>
<td>92</td>
<td>5.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/e:/l</td>
<td></td>
<td></td>
<td>88</td>
<td>5.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/o:/l</td>
<td></td>
<td></td>
<td>76</td>
<td>4.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/o:/l</td>
<td></td>
<td></td>
<td>90</td>
<td>5.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/o:/l</td>
<td></td>
<td></td>
<td>77</td>
<td>5.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/u:/l</td>
<td></td>
<td></td>
<td>75</td>
<td>4.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/u:/l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/u:/l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/u:/l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

100
Table 6.11. Assimilation pattern and Goodness Rating at Time 2, Experiment 2. Light grey indicates an uncategorised L2 vowel using a 50% assimilation criterion. Dark grey indicates an un categorised vowel using a 70% assimilation criterion.

<table>
<thead>
<tr>
<th>Category</th>
<th>/l/</th>
<th>/lː/</th>
<th>/əl/</th>
<th>/əːl/</th>
<th>/ol/</th>
<th>/ɒl/</th>
<th>/ʌl/</th>
<th>GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>/l/</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.78</td>
</tr>
<tr>
<td>/lː/</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.6</td>
</tr>
<tr>
<td>/əl/</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.08</td>
</tr>
<tr>
<td>/əːl/</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.71</td>
</tr>
<tr>
<td>/lː/</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.05</td>
</tr>
<tr>
<td>/əl/</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.42</td>
</tr>
<tr>
<td>/əːl/</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.88</td>
</tr>
<tr>
<td>/ɒːl/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɒl/</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.39</td>
</tr>
<tr>
<td>/el/</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.45</td>
</tr>
<tr>
<td>/eːl/</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.97</td>
</tr>
<tr>
<td>/ɒl/</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.49</td>
</tr>
<tr>
<td>/ɒːl/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɒːl/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɒːl/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A series of paired samples \( t \)-tests revealed that the difference in the average assimilation scores between test time 1 and 2 was non-significant using the 50% assimilation criterion and excluding the score from AusE /æə/ which was not categorised at both testing sessions: \( t(14) = .431, ns \), as was the difference in Goodness Rating (paired samples \( t \)-test: \( t(14) = .561, ns \)). Likewise, the difference in the average assimilation scores between test time 1 and 2 was non-significant using the 70% assimilation criterion (paired samples \( t \)-test, excluding the scores from AusE /oː/ and /uː/ which were not categorised at both testing sessions: \( t(11) = .537, ns \)). The difference in Goodness Rating of the vowels that were categorised using the 70% criterion was also non-significant (paired samples \( t \)-test between times 1 and 2: \( t(11) = 1.961, ns \)).

### 6.4.3.2 Discrimination

The results from the discrimination tasks at test time 1 (See Figure 6.2) revealed an SC /ɪː-/ /ʊə/ discrimination accuracy of 51%, a TC /ɪː-/ /ɜː/ discrimination accuracy of 87%, a UC /ɔː-/ /uː:/ discrimination accuracy of 88%, and finally a UU/UC /θʌ/-/oː/ discrimination accuracy of 77%. Results from the second test session revealed an SC /ɪː-/ /ʊə/ discrimination accuracy of 52%, a TC /ɪː-/ /ɜː/ discrimination accuracy of 86%, a UC /ɔː-/ /uː:/ discrimination accuracy of 92%, and a UU/UC /θʌ/-/oː/ discrimination accuracy of 76%. The average discrimination accuracy was 76% at both testing times. When we excluded the SC vowel contrast /ɪː-/ /ʊə/, which is indiscernible even to native speakers of AusE for whom it has apparently undergone a merger (see Appendix 1: Bundgaard-Nielsen, Best, Kroos, & Tyler, submitted), the discrimination average was an impressive 84% at both testing times.
A (2)x(2) ANOVA was conducted on the percent correct discrimination scores. There were two within-group factors, contrast, with four levels (SC, TC, UC, and UU/UC), and time (Time 1, Time 2). An alpha level of .05 was used for all analyses unless specified. Mean percent correct discrimination scores, split by time and by contrast, are presented in Figure 6.2. There were significant main effects of contrast, $F(2,24) = 46.98$, $p < .001$, $\eta^2 = .85$, but not of time, $F(1,8) = 0.95$, $ns$, and there was no significant interaction, $F(3,24) = 1.20$, $ns$. To test for differences between discrimination of the four contrasts, we conducted six post-hoc paired t-tests with a Bonferroni-adjusted alpha level of .008. The effect of contrast was due to the difference in the average of the Time 1 and Time 2 discrimination scores between SC $\text{[ɪ:]-[tɛ]}$ and the three other contrasts TC $\text{[ɪ:]-[l]}$, UC $\text{[ɛ]:-[u]}$, and UU/UC $\text{[ɬu]:-[l]}$, as well as a difference in the discrimination of the TC $\text{[ɪ:]-[l]}$ and UU/UC $\text{[ɬu]:-[l]}$ contrasts (see Table 6.12). A t-test was conducted on the average L2 vocabulary sizes of the learners at Test Time 1 and 2, showing a marginal increase, $t(8) = 3.52$, $p = 0.08$. 

Figure 6.2. Discrimination accuracy for the SC ($\text{[ɪ:]-[tɛ]}$), TC ($\text{[ɪ:]-[l]}$), UC ($\text{[ɛ]:-[u]}$), and UU/UC ($\text{[ɬu]:-[l]}$) contrasts and overall discrimination accuracy including SC (Overall), as well as overall accuracy excluding SC (Overall-SC). Error bars reflect S.E.M.
Table 6.12. Paired t-tests of the difference in the discrimination average of the four AusE contrasts at the two test sessions. * indicates a statistically significant difference.

<table>
<thead>
<tr>
<th>Contrast</th>
<th>t(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC-TC</td>
<td>12.41*</td>
</tr>
<tr>
<td>SC-UC</td>
<td>10.88*</td>
</tr>
<tr>
<td>SC-UC2</td>
<td>7.90*</td>
</tr>
<tr>
<td>TC-UC</td>
<td>.74</td>
</tr>
<tr>
<td>TC-UU/UC</td>
<td>3.73*</td>
</tr>
<tr>
<td>UC-UU/UC</td>
<td>2.90</td>
</tr>
</tbody>
</table>

6.4.3.3 Discrimination in LV-Participants

In order to further understand the role of L2 vocabulary size in L2 perception, we took a closer look at the differences between the performance of the two LV-participants and the seven HV-participants in Experiment 2. The results indicate that the two LV-participants did improve in their L2 perception over time, while this was not the case for the seven participants who had L2 vocabularies above 6000 words already at the first testing session. Figure 6.3 below indicates (with ellipses) which L2 contrasts the two LV-participants improved on between the two test sessions. Figure 6.3 also indicates that the LV-participants improved in their overall performance both when the SC contrast (for which no improvement was expected over time) was included and when it was excluded in the calculation. Figure 6.4 below indicates that the HV-participants appeared to improve only in their discrimination of the UC and UU/UC contrasts (which was indeed an UC contrast for these learners, just as it had been for the HV learners in Experiment 1), and displayed no improvement in their overall performance between the two testing times.
Figure 6.3. Closer look at the change in discrimination for the four contrast types (SC, TC, UC, and UU) as well as the overall discrimination score (Overall) and the overall discrimination score excluding the SC contrast between the two testing times for the two LV-participants. Improvement in performance between the two testing times is indicated by ellipses. Error bars reflect S.E.M.
6.4.4 Discussion

The results of Experiment 2 revealed very consistent and successful L2 vowel assimilation by the group of nine Japanese learners at both testing times; successful discrimination of three of the four L2-vowel contrasts of interest, also at both testing times; and a gain in average vocabulary size of almost 700 words, or approximately 30 words per week. The results reflect a persistent effect of the L1 background of the learners on their L2 perception similar to that evident in Experiment 1, and in the previous study (Chapter 5: Bundgaard-Nielsen et al., submitted), but the results here differ from the first study in that the learners of the present paper had larger L2 vocabularies, better L2 discrimination and more consistent L2 assimilation behaviour. This indicates that, for the learners tested here, significant adjustment of their L2 perception had already taken place before the first testing time, alongside their L2 vocabulary expansion during initial English immersion. The closer
look at the differences in the results for the two LV-participants and the seven HV-participants further support the hypothesised role of L2 vocabulary acquisition in L2 perception: while the LV-participants appeared to have improved in the discrimination of two L2 contrasts (UC and UU/UC), as well as in their overall discrimination accuracy, the HV-participants, who had enjoyed a similar extended immersion period, only improved in their discrimination of the UC contrast, and not in their overall discrimination performance.

While the results are compatible with the Vocab Model predictions, they are not compatible with the general predictions of improved L2 perception with increased L2 immersion duration, as there was no evidence of improvement in the learners’ L2 vowel perception from testing time 1 to testing time 2. This lack of improvement is likely due to this particular group of learners having relatively large L2 vocabularies and exhibiting very successful L2-L1 vowel assimilation and high levels of discrimination accuracy even at the beginning of immersion in Australia. Indeed, the learners performed on par with a group of native Australian English speakers tested on those same contrasts in a previous study (Appendix 1: Bundgaard-Nielsen, Best, Kroos, & Tyler, submitted), indicating that their L2 perception even at test time 1 (Experiment 1: $M = 7$ weeks of immersion) was most likely sufficient for their day-to-day interactions in English.

The lack of perceptual improvement from an additional 22 weeks of L2 immersion, and the speed with which the L2 learners have reached such a high level of reattunement to the L2 raises questions of the rate (and timing within the L2 learning process) of perceptual re-organisation. While recent research has indicated that most changes to L2 perception and production occur in the first year of immersion, few have suggested that reattunement takes place in just a few weeks of immersion. Importantly, however, all but two$^6$ of the nine participants had a vocabulary well beyond the 6000 word-threshold already at test time 1.

---

$^6$ The two LV-participants were 21 and 31 years old; had been in Australia for 2 and 3 weeks at the first testing session and 24 and 22 weeks at the second test session; had 6 and 8 years of English instruction in Japan from the ages of 12 and 13; and increased their L2 vocabulary by 400 and 1100 words, respectively.
Besides rapid reattunement and rephonologisation, the present findings may indicate that, while the learners’ vocabularies continue to expand during the period of their L2 immersion, changes to their L2 segmental perception have already slowed down as they have reached a level of L2 perceptual performance that allows them to function (without too many segmental confusions arising from accented perceptions) in their L2 language environment.

Finally, while there is no doubt that the Japanese learners of this experiment had existing written vocabularies in English before they arrived in Australia, we doubt that perceptual reattunement to spoken Australian English vowels had already started in Japan, as their L2 instruction in Japan was mainly in written form, and typically given by native speakers of Japanese. We do, however, expect that the learners may have started to re-attune to Japanese accented English vowels before their arrival in Australia, which may have further complicated the task at hand for the learners once they were exposed to native Australian English vowels.

6.5 General Discussion

Experiment 1 replicated and extended our previous work on the role of L2 vocabulary size in L2 vowel perception (Chapter 5: Bundgaard-Nielsen et al., submitted). The results show a significant difference in both the consistency of the L2 to L1 vowel assimilation pattern, as in the previous paper, but also differences in L2 vowel discrimination accuracy in L2 learners with larger versus smaller L2 vocabularies, novel in the present paper. The present report also presents a picture of an unexpectedly high level of competence, in particular with the 50% assimilation criterion, in Japanese learners of Australian English when they are presented with the full L2 vowel system for assimilation while provided with the full set of possible Japanese vowels (single and bi-moraic) and dynamic vowel combinations as assimilation choices. This approach highlights the perceptual flexibility of L2 learners, as they are not restricted by only a small set of pre-selected L1 and/or L2 categories in the experimental set-up.

The present research also suggests that future research could benefit by developing new methods to assess L2 rephonologisation. If indeed learners shift their attention from
surface L2-L1 phonetic similarities to the higher-order phonological properties of their L2, simply applying the task of assimilating L2 phones to L1 categories may not allow advanced/competent L2 learners to adequately display their L2 phonological competence (in particular when the L1 contains a very limited number of categories and the learner has successfully established a new L2 category), though they may well be able to continue to assimilate the L2 phones to L1 categories on the basis of surface phonetic similarities, if that is the only option provided to them in the laboratory task.

Experiment 2 extended those findings to a longitudinal examination of the effect of immersion duration on L2 vocabulary size and L2 vowel perception. The results indicate that while the L2 vocabularies of L2 learners in immersion situations may increase during the first 6-8 months of their immersion, this is not necessarily reflected in improved L2 vowel perception.

While not indicating improved L2 perception at the second testing time by the whole longitudinal group, the results from Experiment 2 are nonetheless consistent with the findings in both Experiment 1 and our previous study (Chapter 5: Bundgaard-Nielsen et al, submitted), as the learner group in Experiment 2 exhibited a large L2 vocabulary of more than an average 6000 words even at the first testing time, and therefore should already have progressed significantly in their reattunement to the L2 vowels and subsequent phonological reorganisation by that first time. If the group of learners had had smaller vocabularies at the first testing time, we suspect that their L2 assimilation and discrimination performance at the first testing time would have been poorer, and that they would have thus have been better able to exhibit a pattern of improved assimilation accuracy and higher discrimination accuracy at the second testing time, consistent with the vocabulary-size perceptual differences we found in Experiment 1 and our previous study. This expectation is also supported by our interpretation of the performance of the two LV-participants, compared to that of the HV-participants, in the second experiment (See Figure 6.3 and 6.4) who did have small L2 vocabularies at the first testing session.

The results from both experiments are compatible with the PAM-L2 prediction that the discrimination of SC and TC contrasts should not improve with extended exposure to the
L2 contrast. The results also suggest that similar language acquisition processes may underlie late L2 acquisition as they do in L1 acquisition in infants. And though more data is needed on early-phase L2 development, the two studies do suggest a close link between L2 vocabulary acquisition and perceptual reattunement and reorganisation in the learner, until the learner’s segmental perception has reached a level at which everyday communication in the L2 is not significantly hindered by misperceptions based on the L1 phonetic and phonological systems of the learner. Of course, it is possible that L2 vowel perception and rephonologisation and L2 vocabulary expansion are parallel but unrelated phenomena, though we find this highly unlikely as only vowels that are contrastive in the lexicon are also contrastive in terms of their phonological status. The acquisition of the higher-order phonological organisation of any given language must thus be closely related to the acquisition of lexical items, though of course there is no upper limit to the number of L2 words that a learner may acquire above the most necessary vocabulary for successful verbal navigation in the L2 community.

In addition, the findings from the two studies confirm the existing understanding that the L1 background of a learner shapes his or her L2 perception, and the PAM predictions for non-native contrast discrimination for the different contrast types SC, TC, and UC.

The studies raise a number of important points in relation to the existing literature on L2 phonetic- and phonological acquisition and the developmental trajectory of phonetic/phonological re-organisation in late L2 acquisition. While it has been suggested that the most significant changes in L2 perception may occur within the first year of L2 learning, the present studies indicate that this perceptual change may indeed happen in just 12 weeks of immersion. However, it is not clear whether the rapid perceptual readjustments displayed by the learners reflect a general pattern even for late L2 learners, or if the particular L1-L2 pair or the typical L2 learning history of the participants in the present experiments partly explains the finding. In particular, it is possible that the L2 English exposure that the Japanese learners did have prior to their arrival in Australia provided them with a ‘head start’ in their reattunement and rephonologisation process, as it most likely did in terms of their L2 vocabulary size. This speculation must, however, be qualified, as the learners typically had mainly written English instruction and were taught
by native Japanese foreign language (FL) teachers, and thus may have attuned to Japanese accented English more so than to native English. Importantly, however, the HV and LV groups did not differ in the amount of English FL instruction they had received prior to their AusE immersion. Future research must address the question of the role of L2 vocabulary acquisition in L2 rephonologisation in learners who are truly naïve upon their arrival in the L2 language environment.

Another point to consider is that the present research also focused solely on vowel perception. Further research would do well to also focus on possible differences in reattunement and rephonologisation for vowels and consonants. It is possible that the perceptual differences (and underlying articulatory differences) in vowels and consonants lead to different acquisition trajectories, such that vowels, for which perception is continuous rather than categorical and category discrimination remains possible to some extent, afford a more rapid re-adjustment than consonants, which are categorically perceived and for which within-category discrimination is generally much more difficult (See for instance, Strange, 1998a; 1998b). We suggest that L2 consonant acquisition may exhibit a later reattunement and rephonologisation trajectory, parallel to the observed pattern of timing differences in L1 perceptual attunement in infancy, where native-language attunement to consonants occurs approximately 4 months after it is evident for native vowels. Finally, future research must determine how the rapid perceptual reattunement and rephonologisation affects production of L2 vowels and consonants.

In conclusion, the present paper supports the PAM-L2 prediction that L2 vocabulary acquisition is intimately linked to improvement in L2 segmental perception, with a further indication that the perceptual reattunement and rephonologisation plateaus once the learner has reached an L2 vocabulary that allows him or her to function adequately in the L2. The findings also bring out similarities between infant L1 acquisition and late L2 acquisition, and thus are also consistent with the propositions of both PAM/PAM-L2 and SLM that even late L2 acquisition is analogous to L1 acquisition in that the learning mechanisms employed in L1 acquisition are available to the late learner of a second language. Finally, the experiments lead to a series of unanswered questions about how L2 vocabulary may
effect change in the perception of L2 consonants, and also questions about the role of L2 vocabulary expansion in change in L2 production.
6.6 References


7 SECOND LANGUAGE LEARNERS’ VOCABULARY EXPANSION IS ASSOCIATED WITH IMPROVED L2 VOWEL INTELLIGIBILITY

Submitted to Journal of the Acoustical Society of America.

7.1 Abstract

This investigation extends the PAM-L2 (Best & Tyler, 2007) proposition that a large L2 vocabulary may curtail improvements to L2 segmental perception by introducing the Vocabulary-Tuning Model of L2 Rephonologisation (Vocab Model), which proposes a facilitating effect of the early phase of adults’ second language (L2) vocabulary expansion on L2 production and suggests that early improvements in L2 segmental production may be positively associated with an expanding L2 vocabulary. To test the Vocab Model, we conducted two experiments. The first tested the prediction that learners who differ only in the size of their L2 vocabularies will also differ in their L2 vowel intelligibility. The second was a longitudinal examination of the production of L2 vowels by a group of L2 learners, most of whom had relatively large L2 vocabularies at only a few weeks of L2 immersion. It compared their production accuracy at a time shortly after arrival in Australia with their productions after 6-8 months of L2 immersion in Australia. The results supported the predicted positive association between L2 vocabulary size and L2 vowel intelligibility in a group of newly-immersed adult Japanese learners of Australian English (AusE) and thus also supported the Vocab Model prediction that early-phase L2 vocabulary expansion leads to improved L2 production. At the same time, the results supported the additional Vocab Model premise that extended immersion does not necessarily lead to improved intelligibility when the L2 learners already have L2 vocabularies that serve them as adequate for basic conversation (i.e., > 6000 words, see Nation, [2006]) very early in their L2 immersion.
7.2 Introduction

Adult Second Language (L2) learners typically speak their L2 with an accent (Flege, Munro, & MacKay, 1995; Flege, 2002; Munro, Flege, MacKay, 1996). The nature of this accent is connected to the relationship between the learner’s native language (L1) and the L2, both in terms of the perceptual similarities between the phones of the L2 and the L1 (see for instance Bohn & Flege, 1992), and in terms of the size of the L2 phonemic inventory relative to the L1 phonemic inventory (Iverson, & Evans, 2007). The strength of the L2 accent is also partly determined by the amount of experience that a learner has had with the L2 (Best & Strange, 1992; Bohn & Flege, 1992; Flege, Bohn, & Jang, 1997; MacKain, Best, & Strange, 1981), as well as his or her relative L1/L2 usage patterns (Flege, Frieda, & Nozawa, 1997; Piske, MacKay, & Flege, 2001), and the age at which the learner has acquired the L2 (Flege, Munro, & MacKay, 1995; Flege, Schirru, & MacKay, 2003; Flege, Yeni-Komshian, & Liu, 1999; MacKay, Meador, & Flege, 2001; Munro, Flege, & MacKay, 1996; Oyama, 1976; Piske, Flege, MacKay, & Meador, 2002; Seliger, Krashen, & Ladefoged, 1975; Yamada, 1995; Yeni-Komshian, Flege, & Liu, 2000).

Despite this wealth of knowledge about the factors that influence L2 speech accent, theories dealing specifically with L2 production, such as the Speech Learning Model (SLM; Flege, 1995), have been slow to put forward testable hypotheses for what drives the clearly observable improvement from naïve to experienced L2 production. Rather, the focus has been on differences in the ‘learnability’ of L2 phones based on their perceived similarity to L1 categories (Flege, 1987; Aoyama et al., 2004), and the detailing of the improved production observed in the relative patterns of intelligibility, i.e., how well native listeners are able to identify accented phones as members of the intended category) at the two most extreme points of L2 acquisition - naïve or inexperienced L2 speakers, versus highly experienced L2 speakers (see for instance Flege, Bohn, & Jang, 1997; Flege, Takagi, & Mann, 1995; Larson-Hall, 2006).

While theories of speech perception and production generally and sensibly assume that a close relationship exists between changes in L2 production and L2 perception. And although the perceptual training has been found to have an beneficial effect also on production (see for instance, Bradlow et al., 1997), theories more focused on cross-
language speech perception, such as the Perceptual Assimilation Model (PAM; Best, 1994; 1995), have likewise avoided the issue of what drives the change from naïve to experienced L2 perception. Instead, the main focus has been on how the native phonological inventory (and the native phonetic realisations of those phones) affects the perception of non-native phones by naïve listeners. In particular, PAM in its original form focussed on the patterns of discrimination of pairs of non-native phones that map onto native phonological categories in different ways, as a means to infer the organisation of the native phonological inventory.

Only recently, through the extension of PAM to L2 acquisition (PAM-L2; Best & Tyler, 2007), has a model been developed that includes predictions regarding which aspects of L2 use and exposure drive L2 phonetic and phonological change in perception and production.

7.2.1 PAM-L2: An Account of L2 Perception and Production

PAM-L2 assumes (as does SLM) that the language learning mechanisms of childhood L1 acquisition remain accessible throughout life, and, by extension, that the mechanisms involved in L1 attunement and phonologisation, i.e., the tuning-in to the higher-order phonological structure of one’s native language, are also employed in the L2 acquisition requirement of reattunement of the perceptual systems and subsequent rephonologisation, i.e., the establishment of an L2 phonological system, by modification or addition to the L1 phonological system of the learner.

According to PAM-L2, L2 phones are first perceptually assimilated into or dissimilated from already existing L1 categories on a phonetic level, i.e., on the basis of their language specific realisation. PAM-L2 further proposes that continued L2 vocabulary acquisition may subsequently lead an L2 learner to attune to higher-order phonological distinctions in the L2 that may differ systematically from those in the L1. This phonetic and subsequently phonological attunement, however, only happens optimally during the very beginning phase of L2 acquisition before a learner ‘settles’ on a version of the L2 phonology, as he or she uses the language more and more routinely. The L2 phonology is, however, likely to still be accented and reflect the L1 phonology of the learner. In other words, PAM-L2 suggests that improvement in L2 segmental perception may slow down very early in L2
learning, while the L2 vocabulary is still fairly small. Furthermore, an even modestly larger beginning-phase L2 vocabulary may pressure the learner to settle on a particular L1-influenced version of the L2 phonology in order to build and maintain a large enough L2 vocabulary for adequate communication, curtailing further phonetic and phonological learning.

True to its theoretical foundations in Direct Realism (Gibson, 1966; 1979) and Articulatory Phonology (Goldstein & Fowler, 2003; Browman & Goldstein; 1989; 1990; 1992), PAM/PAM-L2 suggests that the directly perceived gestural information in speech (i.e., the multimodal speech information) available from other speakers is also available, along with proprioception, from the listener herself during her own speech production. As no translation is needed between what is perceived and what is produced by the self, this leads to the general prediction that production will follow perception (unless the speaker receives specific articulatory training), as the speaker will perceive the articulatory gestural basis of his or her own productions and the productions of others, and subsequently fine-tune the gestures involved in particular speech acts.

This lag in the development of speech production skills relative to speech perception development has been consistently identified in L1 learning toddlers (Benedict, 1979; Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998) and in L2 learning adults (Fan, 2000; Laufer, 1998; Laufer & Paribakht, 1998; Webb, 2008). In comparing the development of L1 and L2 production at least two factors must be taken into consideration: firstly, there are differences between the cognitive abilities and motor control of a child L1 learner and those of an adult L2 learner (Reznick & Goldfield, 1992; Harris & Chasin, 1999), and secondly, the adult L2 learner may experience additional interference from the articulatory habits developed to suit their L1. Despite these two factors, however, the transfer of L2 perceptual learning to (more delayed) L2 production has been documented (e.g., Bradlow et al., 1997; Akahane-Yamada, Tohkura, Bradlow, & Pisoni, 1996), as would be expected.

The expectation of perceptual change (i.e., the achievement of more native-like L2 perception) occurring before changes in L2 production sets PAM apart from the Motor
Theory of Speech Perception (MT: Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985) with which PAM is often compared, due to both theories’ focus on speech gestures rather than speech acoustics. While not pertaining to L2 speech perception/production in particular, MT assumes that the fundamentals of speech are neuro-motor commands that correspond – in a one-to-one fashion – to the intended phonemes, and that perception and production (even in an L2) must be conceptualised as flipsides of the same coin: production is perception in reverse, and vice versa. This leads to the prediction that perception and production develop simultaneously, or even with a production lead, contrasting with the PAM/PAM-L2 prediction that perceptual changes lead, and changes in production lag.

7.2.2 L2 Vocabulary Size Matters

We extend PAM-L2 by introducing the Vocabulary-Tuning Model of L2 Rephonologisation, or Vocab Model for short, which focusses on the facilitating effect of L2 vocabulary expansion in the early phase of L2-perceptual attunement, rephonologisation and L2 production, rather than simply on the curtailing effect of a larger L2 vocabulary. The Vocab Model suggests that improvements in L2 segmental perception and production during the early stages of L2 immersion may be positively associated with an expanding L2 vocabulary, as the need to decipher and comprehend L2 speech rapidly guides the learner to tune into the phonological system of that particular language, rather than continue to perceive L2 speech on the basis of its superficial phonetic similarities (and dissimilarities) to the L1 (Cutler & Otake, 2004; Cutler, Weber, Otake, 2006).

The proposed positive association between a larger L2 vocabulary and both L2 perception and L2 production is commensurate with research detailing the close relationship between initial L1 vocabulary development and both segmental perception and production in infants and toddlers. For instance, a positive association has been demonstrated between a child’s ability to discriminate native speech sounds at six months and the size of his/her vocabulary (receptive and productive) in the second year of life (Tsao Liu & Kuhl, 2004), suggesting that successful segmental perception is essential to vocabulary development. Likewise, children with smaller productive vocabularies perform more poorly than children with larger vocabularies in terms of accuracy in the production of low-frequency
L1 two-phoneme combinations (Edwards, Beckman, & Munson, 2004), indicating that vocabulary size is closely related to successful speech production as well. Similarly, late talkers between 18-33 months-of-age perform worse in tests of phonetic complexity i.e., they produce phonetically complex utterances in a manner less intelligible than typically developing toddlers (Thal, Oroz, & McGaw, 1995), while children with larger vocabularies have been found to outperform children with smaller vocabularies in non-word repetition tasks (Bowey, 2001). Taken together, these studies indicate strong links among successful speech perception, vocabulary size, and speech production. And while the association may be somewhat bidirectional, infant speech perception does seem to develop prior to the corresponding segmental production abilities (see for instance Whalen, Levitt, & Goldstein, 2007; Whalen, Levitt, & Wang, 1991; Rvachew, Alhaidary, Mattock, & Polka, 2008).

While the hypothesised positive association between L2 vocabulary expansion and improved L2 vowel production has yet to be tested, two experiments (see Chapters 5 and 6) have recently tested the related hypothesis that L2 vocabulary size is positively associated with L2 segmental perception (e.g., L2 vowel identification and discrimination) in adult L2 learners. In both experiments, Japanese L2 learners of English with larger (>6000 words, which has been proposed to be the necessary L2 vocabulary size for participation in normal L2 conversation, see Nation [2006]) English vocabularies more consistently identified L2 vowels than learners with smaller L2 vocabularies (<6000 words). Indeed, L2 learners with larger vocabularies identified more vowels as consistently belonging to a specific L1 category in more than 50% of presentations, exhibited a higher average assimilation score (percentage of vowels identified as that one category) for those vowels, and used a smaller set of L1 categories to label the set of L2 phones (Bundgaard-Nielsen et al, submitted-a; Bundgaard-Nielsen, Best & Tyler, submitted-b).

One study (see Chapter 6: Bundgaard et al, submitted-b) also examined the discrimination of a number of PAM contrast assimilation pairs (Single Category contrast /iː/-/øː/, Two
Category contrast /i/-/ɪ/, Uncategorised-Categorised contrast /ɜ/-/ʊ/, and Uncategorised-Uncategorised contrast /əʊ/-/o;\(^7\)). Consistent with the finding that learners with large L2 vocabularies more consistently categorise L2 phones in terms of their L1 categories, the discrimination study found that Japanese learners with larger L2 vocabularies were better at discriminating the predicted subset of selected L2 contrasts (i.e., better discrimination of UU and/or UC assimilation types) than those with smaller L2 vocabularies.

In the following experiments, we turn to L2 speech production, testing the predicted positive association between L2 vocabulary size and L2 vowel intelligibility in a group of newly-immersed adult Japanese learners of Australian English (AusE) in two ways. Firstly, we test the prediction that learners who differ only in the size of their L2 vocabularies (and not in other relevant measures such as amount of L2 experience, years of learning, or duration of immersion), will also differ in their L2 vowel production intelligibility (Experiment 1). Secondly (Experiment 2), we examine the production of the same L2 vowels by a subset of the participants from Experiment 1 in a longitudinal study, comparing their production accuracy at a time shortly after arrival in Australia with that after 6-8 months of L2 immersion in Australia. The execution of this study is important as it allows for the testing of the assumption that L2 immersion in the early stages of L2 acquisition invariably leads to improved L2 production, even when the L2 learners have larger L2 vocabularies at a very early point during their L2 immersion, as an alternative to the expectations of PAM-L2 that any improvements will be restricted to only certain types of perceptual assimilation (UU, UC, CG), and of the Vocab Model that learners with larger L2 vocabularies may actually show less improvement in vowel intelligibility than those with smaller vocabularies, at the same points in L2 learning.

We examined Japanese learners of (Australian) English for several reasons. This particular language combination is traditionally considered one that is fairly taxing on the learner, as the very limited set of L1 vowel categories -- Japanese has only five phonologically

---

\(^7\) This contrast was found to be a second Uncategorised-Categorised contrast for L2 learners with larger L2 vocabularies in Chapter 6.
distinct vowels, in long/short pairs: /i, i, e, eː, a, aː, o, oː, u, uː/ -- does not seem to provide the learner with very many options for assimilation of the total array of 18 Australian English vowels (monophthongs /iː, i, e, eː, ə, æ, e, eː, o, oː, u, uː, ɔ/ and diphthongs /ɪə, əɪ, ɪə, əʊ, ʌə, ʌɪ, ʌə, æə/)⁸. This requires Japanese L2 learners of English to divide their existing phonetic and phonological space more finely in order to accommodate all the L2 vowels. Studying Japanese learners of Australian English also allows us to relate our findings to thorough examinations of the typical Japanese perception of Australian English vowels rather than simple contrastive analyses, i.e., which involve direct comparison of the phonological elements of the L2 relative to those of the L1 in order to predict which ones will be problematic to a given L2 learner on the basis of his/her native language background (see Chapter 5: Bundgaard-Nielsen et al, submitted-a; and Chapter 6: Bundgaard-Nielsen et al., submitted-b). We will thus focus on the seven Australian English vowels /iː, i, ɪə, ʊː, əu, ɔː/ that make up the four PAM contrasts studied in detail in the previous perceptual study (See Chapter 6: Bundgaard-Nielsen et al., submitted-b). These vowels also form a suitable test set for a production study, as they vary in duration, spectral qualities, and in the amount of dynamic change during production, in ways that are not found in Japanese.

7.3 Experiment 1

Experiment 1 focuses on the production of the Australian English vowels /iː, i, ɪə, ʊː, əu, ɔː, ɔː/ by Japanese learners who differ in L2 vocabulary size, but not in their L2 acquisition history or the amount of immersion experience they have had with their L2.

---

⁸ The results from the recent perceptual experiments indicate a much greater perceptual flexibility in Japanese learners than have previously been predicted on the basis of contrastive analyses (see Chapters 5-6: Bundgaard-Nielsen et al, submitted-a, submitted-b).
7.3.1 Method

7.3.1.1 Japanese Productions of L2 English Vowel Targets

The stimuli used in this experiment were citation-form bi-syllabic nonsense words designed to elicit five AusE stressed monophthongs (ɨː, ɪ, ə, ʊː) and two diphthongs (æə, eu). These specific vowels were selected on the basis of the perceptual assimilation patterns observed in our preceding study, which were used in the contrasts tested in our discrimination task (Bundgaard-Nielsen et al, submitted-b). The bi-syllabic nonsense words were in the form of first-syllable stressed /hVbə/. This context was chosen to minimise any effect of consonant co-articulation on the vowels (Strange, Weber, Levy, Shafiro, Hisagi, & Nishi, 2007). Furthermore, this structure is phonotactically legal in both English and Japanese, and thus has the benefit of not causing problems for the Japanese speakers or the Australian English listeners. The nonsense words were produced in isolation by 26 (7 male; 19 female) native Japanese learners of English as an L2 (M_{age} = 23.1 years, M_{stay in Australia} = 5.4 weeks, M_{Foreign Language (FL) English instruction} = 8.3 years).

All Japanese speakers produced five repetitions of each vowel. They were elicited from the speakers by having them read aloud randomised lists containing the target nonsense words heeba, hibba, horba, whoba, herba, hereba, hoeba. The speakers were instructed to use an informal speaking style at a normal conversational speaking rate. Tokens number 2, 3, and 4 of the five recorded tokens were selected from each of the 26 speakers when possible for use in the perceptual intelligibility experiment with native Australian English (AusE) listeners. A small number of tokens were eliminated from the study as they were clearly identifiable as vowel misreadings (i.e., produced in a manner atypical for that individual speaker, or in a manner that did not conform to English phone-to-grapheme transformation, such as long Japanese /eː/, rather than English /ɪː/, for the spelling ee). They were replaced with tokens 1 and/or 5 when possible, though some speakers consistently misread certain tokens and thus provided no alternatives. Tokens with mispronounced consonants were not considered misread, as L2 consonants were not the focus of this study. Typical consonantal errors were the substitution of /f/ for /θ/ before /aː/ and /oː/, and bilabial fricative /β/ for /b/ before the word-final /aː/. These particular substitutions reflect phonotactic constraints in Japanese, as the so-called /h/-kana series /ha/, /he/, /hi/, /ho/, /hu/ is realised with a
palatalised [ʃ] preceding /l/, and a voiceless bilabial fricative (rather similar to a labiodental
/l/) preceding back vowels /o/ and /u/. Similarly, Japanese does not have a phonemic /ʍ/ but
sometimes realises Japanese /b/ as /β/, which is similar to English /ʍ/. The Japanese
substitutions also included errors which seem to have no basis in L1 phonotactics, such as
/ʔ/ and /b/ for /h/ before /u:/ and /o:/.

Audio recording took place in a sound-attenuated room at MARCS Auditory Laboratories
at the University of Western Sydney, using a Shure SM10A headset microphone, an LG
laptop computer and an external soundcard (Edirol UA-25). The intensity of all words was
adjusted in Praat (Boersma & Weenink, 2008) so that the average RMS intensity of the
target vowel was equal to 70dB. The average vowel durations, F1, F2, and F3 at 25%,
50%, and 75% for each target vowel for the speakers are presented in Appendix 7.A (male
speakers) and 7.B (female speakers).

The participants also completed a multiple-choice L2-English vocabulary size test (Nation
& Beglar, 2007), revealing that they had an average L2 vocabulary of 6392 words (range:
4600-8300 words).

7.3.1.2 Native Australian English Listeners
The listeners were 8 Australian-born, native speakers of Australian English (3 male; 5
female, M_{age} = 25.1 years). They did not speak any other language fluently. One participant
reported knowing some Croatian; two participants had limited knowledge of French; one
of Japanese; and one participant had limited knowledge of French and German. None had
ever lived in a non-English speaking country. All had normal hearing and normal/corrected
to normal vision.

7.3.1.3 Procedure
All AusE listeners were tested in a quiet room at MARCS Auditory Laboratories at
University of Western Sydney, Australia. A total of 529 randomised individual tokens
were presented over KOSS UR/20 headphones from a MacBook laptop computer using
PsyScope. The participants identified each token’s stressed (first) vowel in terms of on-
screen spellings of the 18 bisyllabic target AusE nonsense words. These target words
(heeba, hibba, hebba, hairba, habba, hubba, harba, hobba, horba, hoobba, whoba, herba, hereba, hayba, hightba, howba, hoiba, hoeba) were created to allow the listeners to choose from the complete set of 18 AusE monophthongs and diphthongs (monophthongs /i/, /ɪ, e, ē, ə, ɐ, e, ɛ, ɔ, u, u/ and diphthongs /æi, əe, əi, eu, æʊ:/ Cox, 2006). Prior to the testing session, each participant was familiarised with the target words.

7.3.2 General Predictions

The predictions for this study were of two types; namely, specific predictions for the role of L2 vocabulary size in the overall intelligibility of L2 vowels; and a general expectation that the intelligibility of the L2 vowels will reflect L1 phonetic/phonological interference. On the basis of PAM-L2 and the Vocab Model, we predicted that L2 vocabulary size would be associated with L2 vowel intelligibility, such that L2 learners with larger L2 vocabularies would produce more intelligible (i.e., correctly-identifiable) L2 vowels than learners with smaller L2 vocabularies, even when those subgroups are otherwise comparable in terms of immersion duration and FL instruction in the L2. To operationalise L2 intelligibility, we proposed two complimentary measures: 1) the number of L2 vowels produced in a manner so that they could be reliably identified as intended by native listeners; and, 2) an overall intelligibility score (e.g., average percentage of the L2 vowels that are produced in an intelligible manner).

7.3.3 Results

7.3.3.1 General Intelligibility

The overall results from the identification task are presented in Table 7.1. An L2 vowel was defined as intelligible to the native listeners if it was identified as belonging to the intended L1 category in more than 50% of instances. This 50% criterion follows the approach of previous studies on L2 perception (Best, Faber, & Levitt, 1996; Bundgaard-Nielsen et al., submitted-a [see Chapter 5]; Bundgaard-Nielsen et al., submitted-b [see Chapter 6]).

We included the vowels (selected tokens as described above) of all speakers in the confusion matrix, despite not all speakers producing each of the seven vowels. We
consider this approach sound, as vowels were only excluded (resulting in missing datapoints) on the basis of misreading, that is, when the non-native speakers did not follow English rules for grapheme-to-phoneme transformation. No tokens were excluded because of an unusual realisation that was not based on problems with English orthography. In addition, the average intelligibility score for each vowel was calculated on the basis of the total number of intended productions of that vowel used in the stimulus presentations, and thus not influenced by variation in the number of tokens produced for each vowel.

The confusion matrix reveals that only three (/ɪː, ɪə, ʊ/) of the seven produced vowels (/ɪː, ɪ, ɪə, ʊː, əʊ, ɜː/) were produced in a manner that was intelligible to the listeners, according to our (lenient) 50% criterion (note that chance in this task would be 1/6 or 16.7%), with scores for two vowels (/ɪə/ and /ɪː/) at 51% and 52%, respectively, and /ʊ/ identified in 65% of instances. The vowels /ɪ, ʊː, əʊ, ɜː/ were not produced in a manner that was intelligible to the native listeners. The vowel /ɔː/ was identified as intended in 38% of instances, while /ɜː/ was only identified as intended in 29% of instances, and as /ɪː/ in 44% of instances. Poorest of all was the identification score for /ɔː/ at a mere 5%. Interestingly, the vowel /əʊ/ was perceived as AusE /əʊ/ 50% of the time, reflecting systematic misproduction and misidentification of this particular vowel.
Table 7.1. Overall identification pattern, in percent, summed across all participants. The intended L2 vowels are presented in the leftmost column, while the top row indicates what the native listeners selected/perceived from among all 18 AusE vowel choices. Bold-face with light gray cell indicates the percentage of each vowel identified as intended. Darker gray cells indicate greater mis-identification as a non-intended vowel than as the correct target.

<table>
<thead>
<tr>
<th>Intended vowel</th>
<th>Native listener identification pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/ɪ:/</td>
</tr>
<tr>
<td>/ɪ:/</td>
<td>52</td>
</tr>
<tr>
<td>/ɪ/</td>
<td>44</td>
</tr>
<tr>
<td>/ɪə/</td>
<td>20</td>
</tr>
<tr>
<td>/ɪəl/</td>
<td>2</td>
</tr>
<tr>
<td>/ɪəl/</td>
<td></td>
</tr>
<tr>
<td>/ɪəl/</td>
<td>1</td>
</tr>
<tr>
<td>/ɪəl:/</td>
<td></td>
</tr>
<tr>
<td>/ɪəl:/</td>
<td>2</td>
</tr>
</tbody>
</table>

7.3.3.2 Analysis by Vocabulary Size

To test the hypothesis that L2 vocabulary size is associated with improved L2 intelligibility we split the 26 Japanese speakers ($M_{\text{vocabulary}} = 6392$) into high (HV) and low L2 vocabulary (LV) groups with 6000 words defined as the cut-off, following Chapter 5: Bundgaard-Nielsen et al. (submitted-a) and Chapter 6: Bundgaard-Nielsen, et al. (submitted-b). A series of independent t-tests revealed that the HV and LV groups (HV: $n = 15$, $M_{\text{vocabulary}} = 7213$; LV: $n = 11$, $M_{\text{vocabulary}} = 5272$) differed in the size of their L2 vocabularies ($t(24) = 2.62, p < .001$), but did not differ in terms of years of English study in Japan ($M_{\text{HV}} = 8.8$ years, $M_{\text{LV}} = 7.5$ years, $t(24) = 1.816, p = .082$), nor in their length of residence in Australia on the day of testing ($M_{\text{HV}} = 5.9$ weeks, $M_{\text{LV}} = 4.6$ weeks, $t(24) = .986, p = .334$).
The intelligibility matrices for the LV and HV groups are presented in Tables 7.2 and 7.3. The results show that the intelligibility of the learners was related to their L2 vocabulary size. While the general pattern of intelligibility was similar for the LV (see Table 7.2) and HV (see Table 7.3) groups, the two groups differed on both proposed measures of consistency; namely, number of vowels produced in an intelligible manner, and the average intelligibility score (i.e., average of the percent of each of the seven vowels identified as intended by the native listeners). A tally of the differences between the two groups is tallied and presented in Table 7.4.

**Table 7.2.** Overall identification pattern, in percent, summed across LV participants. The intended L2 vowels are presented in the leftmost column, while the top row indicates what the native listeners selected/perceived from among all 18 AusE vowel choices. Bold-face with light gray cell indicates the percentage of each vowel identified as intended. Darker gray cells indicate greater mis-identification as a non-intended vowel than as the correct target.

<table>
<thead>
<tr>
<th>Intended vowel</th>
<th>Native listener identification pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/ɔ:ə/ /ɨ/ /ɛ/ /ɛ:/ /æ/ /æ:/ /ʌ/ /ʊ/ /ʌ:/ /ʊ:/ /ʌ̯/ /ʊ̯/ /ʌ̞/ /ʊ̞/</td>
</tr>
<tr>
<td>/ɔ:/</td>
<td>50 18 6 17 2 3 1 3</td>
</tr>
<tr>
<td>/ɨ/</td>
<td>48 26 9 13 1 1 1 1</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>23 11 46 7 5 1 4</td>
</tr>
<tr>
<td>/ɛ:/</td>
<td>3 1 2 12 33 6 3 19 20 1</td>
</tr>
<tr>
<td>/æ/</td>
<td>1 1 8 4 5 2 13 63 4</td>
</tr>
<tr>
<td>/æ:/</td>
<td>3 2 1 7 2 6 15 42 1 8 10</td>
</tr>
<tr>
<td>/ʌ:/</td>
<td>3 7 2 5 21 2 5 38 2 4 1 5</td>
</tr>
</tbody>
</table>
Table 7.3. Overall identification pattern, in percent, summed across HV participants. The intended L2 vowels are presented in the leftmost column, while the top row indicates what the native listeners selected/perceived from among all 18 AusE vowel choices. Bold-face with light gray cell indicates the percentage of each vowel identified as intended. Darker gray cells indicate greater mis-identification as a non-intended vowel than as the correct target.

<table>
<thead>
<tr>
<th>Intended vowel</th>
<th>Native listener identification pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/l:/</td>
</tr>
<tr>
<td>/l:/</td>
<td>53</td>
</tr>
<tr>
<td>/h:/</td>
<td>42</td>
</tr>
<tr>
<td>/a:/</td>
<td>17</td>
</tr>
<tr>
<td>/e:/</td>
<td>1</td>
</tr>
<tr>
<td>/o:/</td>
<td></td>
</tr>
<tr>
<td>/u:/</td>
<td>1</td>
</tr>
<tr>
<td>/e:/</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 7.4. Difference in the intelligibility score for the HV group versus the LV group. The leftmost column indicates the intended vowels, while the percent correctly identified vowels are presented horizontally. The reported score is calculated as the difference in the mean percent of a vowel identified as intended between the HV group and the LV group. The score thus reflects the increase in identifiability between the vowels produced by the LV and the HV groups, respectively.

<table>
<thead>
<tr>
<th>Intended vowel</th>
<th>/i:/</th>
<th>/ɪ/</th>
<th>/ɪə/</th>
<th>/ə/</th>
<th>/u:/</th>
<th>/ʊ/</th>
<th>/ɔ:/</th>
<th>/æ:/</th>
<th>/ɜ:/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɪ/</td>
<td></td>
<td>+5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɪə/</td>
<td></td>
<td></td>
<td>+9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ə/</td>
<td></td>
<td></td>
<td></td>
<td>+9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/u:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ʊ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɔ:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æ:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɜ:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

In terms of the number of L2 vowels that were identified as intended, the HV group successfully produced the vowels /i/, /ɪə, u:/ (at 53%, 55%, and 67% respectively), while the LV group did not reach the 50% intelligibility criterion for the vowel /ɪə/ (only 46% of /ɪə/ tokens identified as intended by the native listeners), and thus only produced /i:/ and /u:/ (at 50% and 63%, respectively) as intended. Another difference between the HV and LV group was in the production of AusE target /əu/, which was produced by the HV group in a manner that led the native judges to identify this target as /ɔ/ in 56% of instances, but which did not reach 50% identification as that (incorrect) vowel for the LV group.

For the analysis of overall intelligibility, we chose a different approach to the inclusive approach presented above. We thus excluded participants who did not produce all seven vowels from our analyses. We chose this approach because the identity of missing vowel tokens varied from speaker to speaker (rather than representing general difficulties in
common with a subset of the L2 vowels). Including speakers with an a variety of missing vowels would potentially influence individual intelligibility averages drastically so that some speakers might appear to have surprisingly high intelligibility averages due to not having produced vowels that would typically not be intelligible (such as /æ:/), whereas other participants would appear to be less intelligible than their peers due to not having produced vowels that would typically be relatively intelligible (such as /iː, ɪə, ʊː/). In total, we excluded six HV speakers and four LV speakers from the analysis.

Analysis of the background information for the remaining nine HV speakers (HV9) and seven LV speakers (LV7) followed the pattern for the larger HV and LV groups: HV9 and LV7 differed significantly in terms of their L2 vocabulary size (HV9 = 7288, LV7 = 5157; \( t(14) = 6.283, p < .001 \)), but not in terms of the duration of their stay in Australia (\( t(14) = .185, p = .856 \)), or the number of years that they had studied English in Japan (\( t(14) = 1.370, p = .192 \)). Importantly, HV9 and LV7 differed significantly in the overall intelligibility score: HV9 = 55% intelligible, LV7 = 44% intelligible; \( t(14) = 2.595, p = .021 \).

### 7.3.4 Discussion

While the specific acoustic characteristics of Japanese-accented Australian English vowels (or a subset of AusE vowels) have been systematically described previously (i.e., Tsukada, 1999; Ingram & Park, 1997), this is the first study of the perception of Japanese-accented AusE vowels by native listeners of the L2.

As expected, the general intelligibility pattern revealed that the L2 vowels produced by the Japanese AusE learners were heavily accented (i.e., not identifiable consistently or even reliably as they were intended by native listeners). This accent appears to reflect the phonological inventory of the learners, both in terms of the spectral qualities of the phones and their duration. To exemplify, the Japanese learners were unable to produce AusE /æ:/ and /əʊ/ in an intelligible manner, indicating that they struggled with central AusE vowels, which have no Japanese counterparts. Similarly, the learners had problems producing AusE /t/ in a manner that was identifiable, possible due in part to the durational difference between the (spectrally similar) vowels being different to that between single and bi-
moraic Japanese vowels, and in part due to the Japanese learners not producing crucial dynamic differences between the vowels, e.g., an onglide for AusE /iː/ (see Cox, 1999; 2006).

Also as expected, HV and LV Japanese learners differed in their L2 vowel production on both proposed measures, namely the number of L2 vowels they produced in an intelligible manner, and their overall intelligibility score. The results thus support the PAM-L2/Vocab Model hypothesis that L2 vocabulary development is positively associated with change in L2 production accuracy, even in learners who have spent a very limited time in an L2-speaking environment. The experiment however does not allow the testing of the PAM-L2/Vocab Model prediction that a large vocabulary by the first few weeks of immersion may curtail further improvements in L2 vowel intelligibility during the next few months of immersion.

In the following experiment, we test the PAM-L2/Vocab prediction that L2 vocabulary expansion is associated with improvement to L2 segmental production in early immersion when the learner has a smaller L2 vocabulary, but not when the learner has already acquired a large-enough L2 vocabulary to function, against the alternative assumption (see for instance, Tsukada et al, 2005) that L2 immersion in the early stages of L2 acquisition always leads to improved L2 production – even when learners have larger L2 vocabularies at an early point in their L2 immersion. We’ll refer to this expectation as the Immersion Hypothesis. In the following, we address these questions with a longitudinal experiment that examines changes to L2 segmental production as well as L2 vocabulary development over the course of the first 6-8 months of L2 immersion.

7.4 Experiment 2

This experiment examines the production of L2 Australian English vowels by a subset of the participants from Experiment 1 in a longitudinal study, comparing their production accuracy at a time shortly after arrival in Australia with that after 6-8 months of L2 immersion in Australia. This study allowed us to compare the PAM-L2/Vocab model predictions against the alternative view that learners will show continued improvement as
they accrue L2 experience via immersion, regardless of their starting vocabulary or specific variations in L2-L1 phonetic relations among the target contrasts.

7.4.1 Method

7.4.1.1 Stimulus Material and Speakers

As in Experiment 1, the stimuli used in Experiment 2 were the same five AusE monophthongs /i/, /ɪ/, /ʌ/, /ʊ/; and two diphthongs (/ʌʊ/, /ʊə/) produced in a first-syllable stressed /hVbθ/ context. The target nonsense words were produced by six native speakers of Japanese (2 male; 4 female) in two separate recording sessions. The first recording session took place within 8 weeks of the Japanese learners’ arrival in Sydney ($M_{\text{stay}} = 5$ weeks, range: 2-7.5 weeks), while the second recording session took place after an additional 18.5-26 weeks ($M_{\text{stay}} = 21$ weeks) of immersion. At the time of the second recording session, the Japanese learners had thus spent an average of 26 weeks in Australia (range: 24-30 weeks). All participants were enrolled in either General English courses at University of Technology, Sydney, or in English as a Second Language courses at ESL colleges in Sydney. They all lived with non-Japanese speakers, either in home-stay programmes or in shared accommodation with other international students, which should optimize immersion learning of L2 English. At both recording sessions, the speakers completed a multiple-choice L2-English vocabulary size test (Nation & Beglar, 2007). The L2 vocabulary size questionnaire revealed an average L2 vocabulary at the first recording session of 6467 words (range: 5200-7400 words), placing the group mean above the 6000 word threshold, which has been indicated to be the necessary L2 vocabulary size for participation in normal L2 conversation (Nation, 2006). Two participants however, had vocabularies below 6000 words at the first testing session ⁹, and their results will be looked at in more detail below. Audio recording, stimulus selection and presentation were as

---

⁹ The two LV-participants were both 21 years of age. They had L2 vocabularies of 5900 and 5200 words; 10 and 6 years of English instruction in Japan; and 2 and 3 weeks of immersion at the first testing time and 24 and 26 weeks of immersion at the second testing time, respectively.
Experiment 1. Duration, overall F0, and F1, F2, and F3 at 25%, 50%, and 75% of the vowel’s temporal extent are presented in Appendices 7.C-7.F

7.4.1.2 Participants

The listeners were the same 8 Australian-born native speakers of Australian English (3 male, 5 female, Mage = 25.1 years) who had participated in the first experiment.

7.4.1.3 Procedure

All listeners were tested as in Experiment 1. A total of 248 randomised Japanese L2-English disyllables were presented over KOSS UR/20 headphones from a Mac Book laptop computer using PsyScope. The native AusE listeners identified each Japanese L2-English token’s stressed (first) vowel in terms of all 18 AusE vowels (monophthongs /i:/, /I, e, e', ə, æ, ɛ, ɛ', ɔ, ʊ, a/ and diphthongs /æi, æe, ɔi, əu, əʊ/) shown on the computer screen as the spellings of target nonwords (heeba, hibba, hebba, hairba, habba, hubba, harba, hobba, horba, hoobba, whoba, herba, hereba, hayba, highba, howba, hoiba, hoeba). All participants were reminded of the target articulation of each word prior to the experiment.

7.4.2 General Predictions

The present experiment tested two competing hypotheses. The first hypothesis was consistent with PAM-L2, and in particular with the Vocab Model, that improvement in L2 production in L2 learners with small L2 vocabularies is driven by L2 vocabulary expansion. This hypothesis predicted that, despite their increased immersion duration, L2 learners with larger (>6000) L2 vocabularies would not improve their L2 production during extended L2 immersion, but that those with smaller L2 vocabularies at the early stage of immersion would improve.

The second hypothesis was based on previous findings (e.g., Tsukada et al., 2005) that extended L2-English exposure in the first year of acquisition leads to improved L2 English vowel production. This hypothesis predicted that L2 learners, irrespective of their L2 vocabulary size, were likely to improve their L2 production in the first year of L2 immersion.
Again, as in Experiment 1, we used two complimentary measures of production success, namely the number of L2 vowels produced in a manner allowing identification as intended and an overall intelligibility score.

7.4.3 Results

7.4.3.1 Vocabulary Gain

The six Australian English learning Japanese speakers had an average L2 vocabulary size of 6450 words at the first recording session, placing them, on average, somewhat above the 6000 word LV-HV threshold. At the second recording session, their L2 vocabularies had increased to an average of 7183 words (range: 5200-8300 words). This reflected an average L2 vocabulary gain of 700 words (range: +200 words to +1900 words, representing weekly gain of an average of 27 words). A significant part of this increase in L2 vocabulary was due to only two participants who increased their L2 vocabularies by 1100 and 1900 L2 words, respectively. The remaining four participants increased their vocabularies by 200 to 500 words only. An independent t-test, however, revealed that the average increase in L2 vocabulary, even in this group, was just significant ($t(10) = 2.75, p .04$).

Intelligibility

The overall intelligibility of the six speakers at the first and second recording sessions is presented in Tables 7.5 and 7.6. A tally of the differences between the two test times is presented in Table 7.7

The overall identification patterns at recording Time1 and Time2 are very similar. At both recording times, a total of three vowels were produced in a manner that allowed native listeners to identify them as intended (/ɪ, ɪə, u:/) in more than 50% of instances. Also similar to the pattern observed for the HV group in Experiment 1, the six speakers of the present experiment produced vowel /əu/ in a manner that led the native judges to identify it as /ʊə/ (63% of the time in the first recording session, but only 56% of the time in the second recording session).
Table 7.5. Overall identification pattern at the Time 1, in percent, summed across the nine participants. The intended L2 vowels are presented in the leftmost column, while the horizontal indicates what the native listeners perceived. Bold-face with light gray cell indicates the percentage of each vowel identified as intended. Darker gray cells indicate greater mis-identification as a non-intended vowel than as the correct target.

<table>
<thead>
<tr>
<th>Intended vowel</th>
<th>Native listener identification pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Intended vowel cells" /></td>
<td><img src="image" alt="Native listener identification pattern cells" /></td>
</tr>
</tbody>
</table>
Table 7.6. Overall identification pattern at the Time 2, in percent, summed across the nine participants. The intended L2 vowels are presented on the leftmost column, while the horizontal indicates what the native listeners perceived. Bold-face with light gray cell indicates the percentage of each vowel identified as intended. Darker gray cells indicate greater mis-identification as a non-intended vowel than as the correct target.

<table>
<thead>
<tr>
<th>Intended vowel</th>
<th>Native listener identification pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ㅐ</td>
</tr>
<tr>
<td>ㅐ</td>
<td>69</td>
</tr>
<tr>
<td>ㅔ</td>
<td>51</td>
</tr>
<tr>
<td>ㅏ</td>
<td>5</td>
</tr>
<tr>
<td>ㅓ</td>
<td></td>
</tr>
<tr>
<td>ㅗ</td>
<td></td>
</tr>
<tr>
<td>ㅜ</td>
<td></td>
</tr>
<tr>
<td>ㅓ</td>
<td></td>
</tr>
<tr>
<td>ㅐ</td>
<td>1</td>
</tr>
<tr>
<td>ㅔ</td>
<td>4</td>
</tr>
<tr>
<td>ㅏ</td>
<td>1</td>
</tr>
<tr>
<td>ㅓ</td>
<td>4</td>
</tr>
<tr>
<td>ㅗ</td>
<td>4</td>
</tr>
<tr>
<td>ㅜ</td>
<td>1</td>
</tr>
</tbody>
</table>

142
Table 7.7. Difference in the intelligibility score at the two testing times. The leftmost column indicates the intended vowels, while the difference between the identifiability of the vowels at the two testing times is presented horizontally. The reported score is calculated as the difference in the mean percent of a vowel identified as intended at the first test session and the second test session. The score thus reflects the increase in intelligibility between the vowels produced by the participants from the first to the second test session.

<table>
<thead>
<tr>
<th>Intended vowel</th>
<th>/i:/</th>
<th>/I/</th>
<th>/æ:/</th>
<th>/ɔ:/</th>
<th>/u:/</th>
<th>/œ:/</th>
<th>/ɜ:/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td></td>
<td></td>
<td>+14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/I/</td>
<td></td>
<td></td>
<td></td>
<td>+17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æ:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɔ:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>/u:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+3</td>
</tr>
<tr>
<td>/œ:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+1</td>
</tr>
<tr>
<td>/ɜ:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The analysis of change in overall intelligibility score between the two recording sessions was complicated by some speakers not having produced all vowels in both sessions (due to problems with English grapheme-to-phoneme transformations). The small number of speakers made untenable the approach used in Experiment 1, that is, exclusion of speakers with missing vowels. We therefore chose a different approach to assess change in overall consistency. Rather than excluding speakers with missing vowels, we excluded only those particular vowels, which had not been produced by an individual speaker in both recording sessions. As one of the two male speakers had not produced any /hiːba/ tokens in the first recording session and no /hɜːba/ tokens in the second session, his overall production consistency was calculated on the basis of the remaining five vowels /i, ɪ, ɔ, ʊ, œ/. This ensured that his average consistency score would not be skewed by the difference in intelligibility of the /hɜːba/ and /hiːba/ tokens, and ensured that any comparison of his
performance at Time 1 and Time 2 was based on a symmetrical vowel set. The overall consistency scores of two female speakers were likewise calculated without /h3:bɔ/, as one female speaker did not produce any /h3:bɔ/ tokens at all, and another produced only /h3:bɔ/ tokens in the first recording session. Unfortunately, this approach precluded us from assessing the overall intelligibility of these speakers relative to that of the other participants.

The average intelligibility score of the six participants, excluding all vowels that were not produced by the speakers at both recording times, was 57% at Time1 and 58% at Time2. A paired t-test revealed that there was no significant difference in the overall intelligibility scores obtained ($t(74) = .167, ns$). We note, however, that there appear to be significant improvements in the intelligibility of the front vowels /i:, i, ə/ between the two testing times. In particular, the intelligibility of /i:/ increased by 25%, and the intelligibility of /ʌ/ doubled.

7.4.3.2 Intelligibility of the LV-Speakers

According to the Vocab Model, the intelligibility of speakers with smaller L2 vocabularies was expected to be poorer than those with larger L2 vocabularies, but also was expected to show more improvement over time, as their L2 vocabulary gradually develops and increases pressure to (gradually) rephonologise. In order to examine this prediction more closely, we examined the results from the two LV-participants of this experiment. The vocabulary gain of the two participants was 400 and 500 words, respectively, representing gains of 8% (Time 1 = 5200, Time 2 = 5600) and 9% (Time 1 = 5800 words, Time 2 = 6300 words).

As indicated by the intelligibility matrix in Table 7.8 below, the intelligibility of the two LV-speakers ranged from 0% (for /æu/) to 100% (for /ə:/), with an average intelligibility of 56% at the first session. The intelligibility at the second session (see Table 7.9) also ranged from 0% (for /ə:/) to 100% (for /æu/), with an average intelligibility of 53%. Due to only one speaker having produced AusE /ə:/, and that speaker’s puzzling drop from 100% intelligibility at the first session to 0% intelligibility on the second session, as well as the extremely poor overall intelligibility of this vowel across speakers in the both Experiment
1 and the group analysis presented above for Experiment 2, we recomputed the overall intelligibility analysis without the vowel /ɜ:/.

This analysis revealed that the intelligibility at the first testing session was 49%, while the intelligibility (excluding /ɜ:/) was 62%. This represents an increase in intelligibility of 13%.

Table 7.8. Percent of each intended vowel (presented on the leftmost column) which was identified as intended by the native listeners (presented horizontally) for the two LV-participants in the first testing session of Experiment 2.

<table>
<thead>
<tr>
<th>Intended vowel</th>
<th>/i:/</th>
<th>/ɪ/</th>
<th>/æ/</th>
<th>/ə:/</th>
<th>/u:/</th>
<th>/əʊ/</th>
<th>/ɜ:/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɪ/</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æ/</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ə:/</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/u:/</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/əʊ/</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɜ:/</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7.9. Percent of each intended vowel (presented on the leftmost column) which was identified as intended by the native listeners (presented horizontally) for the two LV-participants in the second testing session of Experiment 2

<table>
<thead>
<tr>
<th>Intended vowel</th>
<th>Identified as intended</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/i:/</td>
</tr>
<tr>
<td>/iː/</td>
<td>79</td>
</tr>
<tr>
<td>/ɪː/</td>
<td>41</td>
</tr>
<tr>
<td>/æː/</td>
<td>100</td>
</tr>
<tr>
<td>/əː/</td>
<td>53</td>
</tr>
<tr>
<td>/æːː/</td>
<td>90</td>
</tr>
<tr>
<td>/æːːː/</td>
<td>6</td>
</tr>
<tr>
<td>/ɜːːː/</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 7.10. Table of difference between the intelligibility (i.e., % of vowels identified as intended) of the Japanese accented vowels at the first and the second test session. The leftmost column indicates the intended vowels, while the difference between the identifiability of the vowels at the two testing times is presented horizontally. The numbers indicating the increase (or decrease) in the percent of vowels identified as intended between the first and the second testing session for the LV-participants.

<table>
<thead>
<tr>
<th>Intended vowel</th>
<th>Difference in identifiability</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td>+16</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>+30</td>
</tr>
<tr>
<td>/æ/</td>
<td>+8</td>
</tr>
<tr>
<td>/ɔ:/</td>
<td>+16</td>
</tr>
<tr>
<td>/u:/</td>
<td>-1</td>
</tr>
<tr>
<td>/əu/</td>
<td>+6</td>
</tr>
<tr>
<td>/ɔː/</td>
<td>-100</td>
</tr>
</tbody>
</table>

7.4.4 Discussion

This experiment tested two competing hypotheses: 1) the general assumption that early L2 immersion leads to improved L2 production, irrespective of L2 vocabulary size; and, 2) the PAM-L2/Vocab Model hypothesis that despite their increased immersion duration, L2 learners with larger average (>6000) L2 vocabularies would fail to improve their L2 production during extended L2 immersion. The results support the latter hypothesis: despite an average of 26 additional weeks of L2 immersion, the group of L2 learners tested here did not improve their L2 vocabulary size significantly, their overall L2 vowel intelligibility, or the number of L2 vowels that they produced in an intelligible manner.

While the overall intelligibility of the full group of Japanese L2 learners did not improve between the two recording sessions, some improvement is evident for the AusE front...
vowels /iː, ɪ, ɪə/ over the course of the additional immersion. We suggest two explanations for this pattern: 1) that it may be seen as the result of further tuning into differences in relative duration between the Japanese single and bi-moraic /i/ and /iː/ and AusE /iː/ and /ɪ/, such that the L2 learners may produce more native-like vowel dynamics and a more Australian-like durational difference, and 2) that the improvement in intelligibility for the AusE vowel /ɪə/ may not be a movement towards a more AusE-like production; rather it may reflect the adaptation of a hyper-articulated (bi-moraic), and perhaps even rhotic /ɪə/, (due to strong orthographic influence: see Basetti, 2008). Native AusE speakers do not necessarily produce /iː/ and /ɪə/ in a manner that allows for easy discrimination even by native English listeners; see Appendix 1, and Bundgaard-Nielsen et al. (2008). As such, this change in L2 intelligibility may thus only in part reflect a decrease in L2 accentedness. This improvement, however, is not enough to influence the overall degree of accentedness.

The lack of improvement for the central vowels (i.e., /ɔː/ and the first central vowel target in the diphthong /əu/) is interesting in the light of the PAM-L2/Vocab Model prediction that initial L2 perception and production is based on phonetic similarities with the L1 and only later based on an (accented) L2 phonology. In fact, the intelligibility patterns at both testing Time 1 and Time 2 suggest that the L2 vowel productions of the learners remained heavily influenced by their L1 over the duration of their 6-month immersion. This L1-influence is particularly apparent in the production of central AusE /ɔː/, which has no Japanese counterpart and which is identified as intended in a mere 2% of instances. Rather than having developed a new L2 phonological category for this vowel, the learners produce AusE /ɔː/ in a way that leads native AusE listeners to perceive it as AusE /ɵ/. Similarly, the diphthong /əʊ/ was not produced in a manner that allowed identification as intended at either testing time. Again, we believe that the learners had failed (yet) to form a phonological category for this novel phone and were producing it on the basis of its surface (phonetic) similarities to one (or more) L1 phones.

The analysis of the performance of the two LV-participants offers some support for the predictions of the Vocab Model. While the two participants improved in terms of their intelligibility and in their L2 vocabulary size, the vocabulary expansion of these two participants was a modest 8-9%. It is not known how much the vocabulary must grow to
result in perceptual reorganisation and rephonologisation, but it is possible that this development is gradual and that vocabulary increases in the range of the two LV participants is associated with improved perception, when the vocabulary is still limited.

### 7.5 General Discussion

The two studies reported here tested the PAM-L2/Vocab Model hypothesis that early during L2 immersion, L2 vocabulary expansion is associated with improved L2 segmental production in a group of Japanese learners of Australian English.

The results from the two experiments support this hypothesis in two ways: Firstly, Experiment 1 supports the prediction that newly immersed L2 learners with larger L2 vocabularies produce more intelligible L2 vowels than newly immersed L2 learners with smaller L2 vocabularies (indicating a close association between L2 vocabulary acquisition and L2 segmental production). Secondly, Experiment 2 supports the PAM-L1/Vocab Model prediction that increased exposure/longer L2 immersion may foster tuning of L2 speech production in individuals with low beginning L2 vocabularies, but does not necessarily lead L2 learners with larger L2 vocabularies (>6000 word) to further re-attune and rephonologise in their L2 as they may have already settled on an (accented) L2 phonology. Further support from this interpretation comes from the analysis of the change in intelligibility for the two LV-participants in Experiment 2, for whom we identified the expected (if modest) improvement in both their L2 vocabulary size, and their L2 vowel intelligibility.

Both experiments are thus compatible with PAM-L2 and the Vocab Model, rather than the Immersion Hypothesis (Tsukada et al., 2005), and in particular with the prediction that early L2 vocabulary expansion facilitates improvements to L2 segmental production, while a larger L2 vocabulary may curtail further L2 segmental production, as the larger vocabulary forces the learner to settle on an (accented) L2 phonology. Learners with larger L2 vocabularies thus outperform learners with smaller L2 vocabularies (Experiment 1) at an early point in L2 immersion, but they do not necessarily improve further in their overall L2 intelligibility when immersion is extended for another 6 months. Instead, it is the low initial-vocabulary learners who apparently benefit more in production from the additional
immersion time. The results are thus highly compatible with those obtained in recent studies of the role of L2 vocabulary development in L2 vowel perception (see Chapter 5: Bundgaard-Nielsen et al., submitted-a, and Chapter 6: Bundgaard-Nielsen et al., submitted-b). In these studies, newly immersed Japanese learners of English who had larger L2 vocabularies also outperformed (in terms of L2 to L1 vowel assimilation, and L2 vowel discrimination) learners with smaller L2 vocabularies. Also parallel to the results obtained here, the longitudinal study found that newly immersed learners with larger L2 vocabularies did not improve their L2 perception (assimilation and discrimination) after 6-8 months of additional L2 immersion.

The results are also compatible with the PAM-L2 and Vocab Model hypothesis that similar mechanisms underlie L1 and L2 language acquisition, with the caveat that the already acquired L1 of the learners will of course interfere with L2 production in a systematic manner, reflecting the similarities and differences between the particular L1-L2 pair. In particular, the results are compatible with observations from child language acquisition that large L1 vocabularies are associated with better segmental production, as well as with better segmental perception in the first year of life. While recognising that development in segmental production (and perception) and vocabulary expansion may be bi-directionally or even cyclically linked in the acquisition process, as has been suggested for child language acquisition (Smith, McGregor, & Demille, 2006), we also tentatively suggest that changes in perception initially precede those in production. This may also be reflected in the observed lag between developmental change in L2 perception and production.

An alternative interpretation of the results presented in Experiment 2 (and the suggested lag of change to L2 production relative to perception) may be that perception and production are not intimately linked to vocabulary acquisition. This interpretation would suggest that the lack of improvement in the overall intelligibility (and number of vowels produced in an intelligible manner) of the learners in Experiment 2 may have resulted because more immersion is needed to improve L2 production than to improve L2 perception and that the duration of their immersion was too short for this to happen. This view would be compatible with observed improvements in segmental production in learners over the course of a year of immersion, rather than 6-8 months (e.g., Tsukada, et
al. 2005), however, it would not explain the observed positive association between vocabulary development and intelligibility of segmental productions reported in Experiment 1. Nor does it explain the previously observed positive association between L2 segmental perception and L2 vocabulary size.

Future research would thus benefit from continuing to consider both the similar processes (such as L2 vocabulary expansion, perceptual reattunement, and rephonologisation) that may govern and connect perception and production intimately in L2 acquisition, but also the origins of the observed lag between improvements in L2 perception and production, and which factors (such as motor-patterning, and orthographic influences) may affect L2 perception and production differently.


7.7 Appendix A

Table 7.11. Acoustic measurements for the 7 male participants, Experiment 1.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Dur (ms)</th>
<th>F0</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>/æ:/</td>
<td>147</td>
<td>150</td>
<td>330</td>
<td>2171</td>
<td>2998</td>
<td>332</td>
<td>2148</td>
<td>2847</td>
<td>347</td>
<td>2081</td>
<td>2665</td>
</tr>
<tr>
<td>/æ/</td>
<td>113</td>
<td>127</td>
<td>308</td>
<td>2167</td>
<td>2940</td>
<td>308</td>
<td>2185</td>
<td>2880</td>
<td>315</td>
<td>2133</td>
<td>2665</td>
</tr>
<tr>
<td>/æə/</td>
<td>171</td>
<td>123</td>
<td>416</td>
<td>1929</td>
<td>2534</td>
<td>510</td>
<td>1691</td>
<td>2379</td>
<td>546</td>
<td>1437</td>
<td>2280</td>
</tr>
<tr>
<td>/æj/</td>
<td>143</td>
<td>125</td>
<td>471</td>
<td>876</td>
<td>2591</td>
<td>466</td>
<td>874</td>
<td>2622</td>
<td>456</td>
<td>874</td>
<td>2600</td>
</tr>
<tr>
<td>/æʃ/</td>
<td>130</td>
<td>129</td>
<td>348</td>
<td>1045</td>
<td>2379</td>
<td>349</td>
<td>1058</td>
<td>2332</td>
<td>351</td>
<td>1036</td>
<td>2352</td>
</tr>
<tr>
<td>/æʉ/</td>
<td>172</td>
<td>124</td>
<td>473</td>
<td>1120</td>
<td>2445</td>
<td>485</td>
<td>1325</td>
<td>2397</td>
<td>476</td>
<td>1372</td>
<td>2388</td>
</tr>
<tr>
<td>/æʒ/</td>
<td>144</td>
<td>123</td>
<td>654</td>
<td>1187</td>
<td>2532</td>
<td>659</td>
<td>1157</td>
<td>2514</td>
<td>629</td>
<td>1103</td>
<td>2492</td>
</tr>
</tbody>
</table>

7.8 Appendix B

Table 7.12. Acoustic measurements for the 19 female participants, Experiment 1.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Dur (ms)</th>
<th>F0</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>/æ:/</td>
<td>135</td>
<td>218</td>
<td>354</td>
<td>2740</td>
<td>3475</td>
<td>360</td>
<td>2617</td>
<td>3435</td>
<td>364</td>
<td>2586</td>
<td>3227</td>
</tr>
<tr>
<td>/æ/</td>
<td>107</td>
<td>218</td>
<td>369</td>
<td>2706</td>
<td>3448</td>
<td>379</td>
<td>2692</td>
<td>3353</td>
<td>389</td>
<td>2509</td>
<td>3165</td>
</tr>
<tr>
<td>/æə/</td>
<td>164</td>
<td>217</td>
<td>438</td>
<td>2483</td>
<td>3180</td>
<td>514</td>
<td>2266</td>
<td>2997</td>
<td>558</td>
<td>1924</td>
<td>2872</td>
</tr>
<tr>
<td>/æj/</td>
<td>159</td>
<td>216</td>
<td>529</td>
<td>950</td>
<td>2941</td>
<td>534</td>
<td>984</td>
<td>2938</td>
<td>533</td>
<td>1047</td>
<td>2963</td>
</tr>
<tr>
<td>/æʃ/</td>
<td>134</td>
<td>220</td>
<td>409</td>
<td>1196</td>
<td>2896</td>
<td>404</td>
<td>1209</td>
<td>2903</td>
<td>401</td>
<td>1231</td>
<td>2922</td>
</tr>
<tr>
<td>/æʉ/</td>
<td>163</td>
<td>213</td>
<td>567</td>
<td>1287</td>
<td>2937</td>
<td>568</td>
<td>1585</td>
<td>2943</td>
<td>539</td>
<td>1685</td>
<td>2952</td>
</tr>
<tr>
<td>/æʒ/</td>
<td>154</td>
<td>215</td>
<td>763</td>
<td>1499</td>
<td>2982</td>
<td>734</td>
<td>1437</td>
<td>2879</td>
<td>717</td>
<td>1369</td>
<td>2912</td>
</tr>
</tbody>
</table>
7.9 Appendix C

Table 7.13. Acoustic measurements for the 2 male participants in Experiment 2, Test Time 1.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Dur (ms)</th>
<th>F0</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a:/</td>
<td>147</td>
<td>150</td>
<td>408</td>
<td>1949</td>
<td>2608</td>
<td>381</td>
<td>1941</td>
<td>2596</td>
<td>403</td>
<td>1894</td>
<td>2440</td>
</tr>
<tr>
<td>/id/</td>
<td>174</td>
<td>115</td>
<td>291</td>
<td>2135</td>
<td>2878</td>
<td>289</td>
<td>2155</td>
<td>2939</td>
<td>294</td>
<td>2120</td>
<td>2816</td>
</tr>
<tr>
<td>/æ/</td>
<td>231</td>
<td>115</td>
<td>434</td>
<td>1885</td>
<td>2507</td>
<td>543</td>
<td>1583</td>
<td>2420</td>
<td>586</td>
<td>1410</td>
<td>2360</td>
</tr>
<tr>
<td>/ı:/</td>
<td>191</td>
<td>113</td>
<td>500</td>
<td>1133</td>
<td>2585</td>
<td>488</td>
<td>1116</td>
<td>2610</td>
<td>489</td>
<td>1132</td>
<td>2605</td>
</tr>
<tr>
<td>/ıd/</td>
<td>186</td>
<td>118</td>
<td>362</td>
<td>1017</td>
<td>2285</td>
<td>357</td>
<td>1042</td>
<td>2298</td>
<td>363</td>
<td>1027</td>
<td>2268</td>
</tr>
<tr>
<td>/æ:/</td>
<td>227</td>
<td>112</td>
<td>429</td>
<td>1465</td>
<td>2439</td>
<td>444</td>
<td>1667</td>
<td>2480</td>
<td>442</td>
<td>1716</td>
<td>2406</td>
</tr>
<tr>
<td>/ı:/</td>
<td>195</td>
<td>112</td>
<td>638</td>
<td>1154</td>
<td>2489</td>
<td>665</td>
<td>1107</td>
<td>2496</td>
<td>624</td>
<td>1065</td>
<td>2480</td>
</tr>
</tbody>
</table>

7.10 Appendix D

Table 7.14. Acoustic measurement for the 2 male participants in Experiment 2, at Test Time 2.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Dur (ms)</th>
<th>F0</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a:/</td>
<td>156</td>
<td>131</td>
<td>284</td>
<td>2135</td>
<td>3232</td>
<td>284</td>
<td>2138</td>
<td>3226</td>
<td>295</td>
<td>2111</td>
<td>3027</td>
</tr>
<tr>
<td>/id/</td>
<td>144</td>
<td>131</td>
<td>285</td>
<td>2142</td>
<td>3123</td>
<td>287</td>
<td>2143</td>
<td>3166</td>
<td>304</td>
<td>2098</td>
<td>2954</td>
</tr>
<tr>
<td>/æ/</td>
<td>200</td>
<td>131</td>
<td>412</td>
<td>1821</td>
<td>2381</td>
<td>494</td>
<td>1591</td>
<td>2261</td>
<td>542</td>
<td>1378</td>
<td>2130</td>
</tr>
<tr>
<td>/ı:/</td>
<td>158</td>
<td>132</td>
<td>466</td>
<td>678</td>
<td>2422</td>
<td>468</td>
<td>754</td>
<td>2429</td>
<td>463</td>
<td>778</td>
<td>2383</td>
</tr>
<tr>
<td>/ıd/</td>
<td>150</td>
<td>131</td>
<td>369</td>
<td>1136</td>
<td>2175</td>
<td>363</td>
<td>1138</td>
<td>2180</td>
<td>365</td>
<td>1123</td>
<td>2228</td>
</tr>
<tr>
<td>/æ:/</td>
<td>183</td>
<td>130</td>
<td>450</td>
<td>1534</td>
<td>2550</td>
<td>447</td>
<td>1702</td>
<td>2576</td>
<td>455</td>
<td>1714</td>
<td>2475</td>
</tr>
<tr>
<td>/ı:/</td>
<td>157</td>
<td>125</td>
<td>635</td>
<td>1107</td>
<td>2361</td>
<td>618</td>
<td>1093</td>
<td>2366</td>
<td>590</td>
<td>1070</td>
<td>2297</td>
</tr>
</tbody>
</table>
### 7.11 Appendix E

Table 7.15. Acoustic measurements for the 4 female participants in Experiment 2, at Test Time 1.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Dur (ms)</th>
<th>F0</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a:/</td>
<td>135</td>
<td>218</td>
<td>354</td>
<td>2740</td>
<td>3475</td>
<td>360</td>
<td>2617</td>
<td>3435</td>
<td>364</td>
<td>2586</td>
<td>3227</td>
</tr>
<tr>
<td>/æ:/</td>
<td>107</td>
<td>218</td>
<td>369</td>
<td>2706</td>
<td>3448</td>
<td>379</td>
<td>2692</td>
<td>3353</td>
<td>389</td>
<td>2509</td>
<td>3165</td>
</tr>
<tr>
<td>/ææ/</td>
<td>164</td>
<td>217</td>
<td>438</td>
<td>2483</td>
<td>3180</td>
<td>514</td>
<td>2266</td>
<td>2997</td>
<td>558</td>
<td>1924</td>
<td>2872</td>
</tr>
<tr>
<td>/æ:/</td>
<td>159</td>
<td>216</td>
<td>529</td>
<td>950</td>
<td>2941</td>
<td>534</td>
<td>984</td>
<td>2938</td>
<td>533</td>
<td>1047</td>
<td>2963</td>
</tr>
<tr>
<td>/æ:/</td>
<td>134</td>
<td>220</td>
<td>409</td>
<td>1196</td>
<td>2896</td>
<td>404</td>
<td>1209</td>
<td>2903</td>
<td>401</td>
<td>1231</td>
<td>2922</td>
</tr>
<tr>
<td>/ææ/</td>
<td>163</td>
<td>213</td>
<td>567</td>
<td>1287</td>
<td>2937</td>
<td>568</td>
<td>1585</td>
<td>2943</td>
<td>539</td>
<td>1685</td>
<td>2952</td>
</tr>
<tr>
<td>/æ:/</td>
<td>154</td>
<td>215</td>
<td>763</td>
<td>1499</td>
<td>2982</td>
<td>734</td>
<td>1437</td>
<td>2879</td>
<td>717</td>
<td>1369</td>
<td>2912</td>
</tr>
</tbody>
</table>

### 7.12 Appendix F

Table 7.16. Acoustic measurements for the 4 female participants in Experiment 2, at Test Time 2.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Dur (ms)</th>
<th>F0</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a:/</td>
<td>141</td>
<td>219</td>
<td>342</td>
<td>2963</td>
<td>3621</td>
<td>340</td>
<td>2946</td>
<td>3649</td>
<td>351</td>
<td>2899</td>
<td>3434</td>
</tr>
<tr>
<td>/æ:/</td>
<td>101</td>
<td>219</td>
<td>345</td>
<td>2970</td>
<td>3715</td>
<td>362</td>
<td>2924</td>
<td>3537</td>
<td>369</td>
<td>2710</td>
<td>3330</td>
</tr>
<tr>
<td>/ææ/</td>
<td>148</td>
<td>202</td>
<td>440</td>
<td>2675</td>
<td>3316</td>
<td>488</td>
<td>2430</td>
<td>3183</td>
<td>558</td>
<td>2012</td>
<td>3012</td>
</tr>
<tr>
<td>/æ:/</td>
<td>140</td>
<td>209</td>
<td>540</td>
<td>992</td>
<td>2992</td>
<td>551</td>
<td>1009</td>
<td>3004</td>
<td>548</td>
<td>999</td>
<td>3013</td>
</tr>
<tr>
<td>/æ:/</td>
<td>126</td>
<td>215</td>
<td>399</td>
<td>1201</td>
<td>2906</td>
<td>408</td>
<td>1203</td>
<td>2954</td>
<td>406</td>
<td>1245</td>
<td>2925</td>
</tr>
<tr>
<td>/ææ/</td>
<td>150</td>
<td>210</td>
<td>497</td>
<td>1609</td>
<td>3099</td>
<td>500</td>
<td>1968</td>
<td>3146</td>
<td>485</td>
<td>1970</td>
<td>3157</td>
</tr>
<tr>
<td>/æ:/</td>
<td>131</td>
<td>210</td>
<td>862</td>
<td>1450</td>
<td>2988</td>
<td>841</td>
<td>1442</td>
<td>2884</td>
<td>789</td>
<td>1376</td>
<td>2927</td>
</tr>
</tbody>
</table>
8 THESIS DISCUSSION AND CONCLUSIONS

The present thesis concerns experiential change in the perception and production of Second Language (L2) Australian English (AusE) vowels by native (L1) speakers of Japanese in the first weeks and months of L2 immersion, with a particular focus on the role of L2 vocabulary expansion in L2 perception and production.

In particular, the thesis develops and tests the Vocabulary-Tuning Model of L2 Rephonologisation, and its novel hypothesis that L2 vocabulary development is positively associated with the acquisition of L2 phonology. This is investigated through a series of five experiments (three cross-sectional and two longitudinal) that examine the role of L2 vocabulary size on L2 vowel perception and production in adult Japanese learners of Australian English. Chapters 6 and 7 of the thesis also examine the effect of immersion duration on L2 segmental perception and production, testing two contrasting hypotheses:

1) the general Immersion Hypothesis (see for instance, Tsukada et al., 2005) that L2 perception and production will improve with extended (6-8 months) L2 immersion; and

2) the novel PAM-L2/Vocab Model prediction that, while L2 vocabulary acquisition may facilitate L2 phonetic/phonological reattunement when the L2 vocabulary is limited, it curtails further reattunement once it has reached a certain size and has forced the learner to settle on an L2 phonology.

8.1 The Foundations of the Thesis

8.1.1 The Factors of L2 Acquisition

The thesis has its foundations in a rich literature demonstrating that the degree of ‘nativelikeness’ of an L2 learner’s L2 perception and production is related to the specific perceived similarities and differences between the phones in the particular L1-L2 pair (see for instance Bohn & Flege, 1992); the size of the L2 phonemic inventory relative to the L1 (Iverson and Evans, 2007); differences in the amount of L2 experience a learner has had (Best & Strange, 1992; Bohn & Flege, 1992; Flege, Bohn, & Jang, 1997; MacKain, Best,
& Strange, 1981); the relative L1-L2 usage patterns (Flege, Frieda, & Nozawa, 1997; Piske, MacKay, & Flege, 2001); and the age of acquisition of the learner (Flege, Munro, & MacKay, 1995; Flege, Schirru, & MacKay, 2003; Flege, Yeni-Komshian, & Liu, 1999; MacKay, Meador & Flege, 2001; Munro, Flege, & MacKay, 1996; Oyama, 1976; Piske, Flege, MacKay, & Meador, 2002; Seliger, Krashen, & Ladefoged, 1975; Yamada, 1995; Yeni-Komshian, Flege, & Liu, 2000).

To summarise the general conclusions of the field of L2 acquisition in terms of learner factors, early L2 learners tend to perform in a more native-like manner than older learners. In terms of L2 usage factors, more experienced learners generally perform better than less experienced learners, and L2 learners with limited L1-use also tend to outperform otherwise comparable learners with greater continued L1-use. To summarise in terms of language factors, it has been assumed that learners with larger L1 phonological inventories tend to perform better than learners with smaller inventories, likely because a larger native inventory provides the learner with more native categories into which the non-native phones can be initially assimilated. The picture, however, does become more complicated when individual L1-L2 phones are considered, as detailed in particular by PAM (Best, 1994; 1995) and PAM-L2 (Best & Tyler, 2007), but also by SLM (Flege, 1995). The adoption of the ‘whole systems’ approach to L2 vowel perception and production that is pioneered in the present thesis further nuances the picture, since the surprising perceptual flexibility of the Japanese learners indicates that even L2 learners with smaller L1 phonological inventories may perform quite well when allowed to use all native phonological categories as well as likely di-phones (bi-moraic vowel combinations), which are not typically considered independent elements of the phonological vowel system in Japanese.

This rich knowledge of the factors resulting in accented L2 perception and production has only recently translated into a detailing of the processes that lead to L2 improvement, namely the formulation of PAM-L2.
8.1.2 PAM-L2

PAM-L2 (Best & Tyler, 2007) is an extension of PAM and accounts for developmental changes in L2 perception and production. PAM-L2 assumes that perceptual learning is possible at all ages but will be influenced by the entire language learning history of the individual. According to PAM-L2, L2 phones are first assimilated into or dissimilated from existing L1 categories at a phonetic level. PAM-L2 further posits that continued L2 vocabulary acquisition encourages the L2 learner to attune to the higher-order organisation of the L2 phones so that this assimilation/dissimilation process also takes place on a phonological level.

PAM-L2 thus suggests that L2 vocabulary acquisition may “exert forceful linguistic pressure” (Best & Tyler, 2007, p. 32) on the learner to attune to phonological differences in the L2 that have previously been ignored in the L1. PAM-L2 further suggests that improvement in L2 segmental perception may reach a plateau early in the L2 acquisition process, while the L2 vocabulary is still small, and that a larger L2 vocabulary may curtail further phonetic learning as the larger vocabulary forces the learner to settle on a particular L1-influenced version of the L2 phonology.

8.1.3 The Vocabulary-Tuning Model of L2 Rephonologisation

The Vocabulary-Tuning Model of L2 Rephonologisation, or Vocab Model, is based on PAM-L2 and further extends that model by suggesting that early improvements in L2 segmental perception in late/adult L2 learners may be positively associated with, or driven by, L2 vocabulary acquisition. Like PAM/PAM-L2, the Vocab Model assumes that the language-learning mechanisms used in L1 acquisition in infant and childhood remain accessible over the lifespan, though the language learning behaviour of an individual’s past (such as their L1 and subsequently learned languages) will reflect his or her language learning history, causing the amply documented L1 accent on L2 perception and production.

In addition, the Vocab Model proposes that L2 segmental acquisition (perceptual attunement to an L2 and subsequent L2-rephonologisation) unfolds in a similar fashion to L1 attunement and phonologisation. In particular, it is proposed that L2 attunement and
rephonologisation are vocabulary-associated in a way similar to the emergence of the L1 phonological function that appears to be associated with or driven by the so-called vocabulary explosion in children around 18+ months-old (e.g., Benedict, 1979; Metsala & Walley, 1998; Nazzi & Bertoncini, 2003).

8.2 The Present Findings on the Role of L2 Vocabulary Size in L2 Segmental Perception and Production

8.2.1 L2 Perception and L2 Vocabulary Size

The experiment presented in Chapter 5 tested the PAM-L2/Vocab Model-based proposition of a close association between L2 vocabulary size and L2 vowel perception. The results support the prediction that L2 learner groups who differ in L2 vocabulary size also differ in their L2 perception in a manner consistent with PAM-L2/Vocab Model predictions. The results show that, though the general patterns of identification were similar for the high (HV) and low vocabulary (LV) groups, the HV learners were not only more consistent in their identification of the L2 phones in terms of identification scores, but also in terms of how many spurious L1 categories they selected for each L2 vowel. The experiment also provides support for the link between L2 vocabulary size and L2 phonological reattunement in observed systematic differences between the two groups in their identification of un-categorised vowels.

Irrespective of L2 vocabulary size, the results are also consistent with PAM/PAM-L2 claims that L2 learners will use both temporal and spectral information in their identification of the L2 vowels, as the Japanese learners consistently used a large number of both mono- and bi-moraic L1 categories in a pattern that reflected both native phonetic and phonological categories.

The results from Experiment 1 in Chapter 6 replicated the differences in L2 vowel assimilation consistency between learners with larger L2 vocabularies and learners with smaller L2 vocabularies found in the experiment reported in Chapter 5. The additional discrimination experiment also revealed better L2 vowel discrimination for Japanese learners of L2 Australian English with larger vocabularies than those with smaller vocabularies. This is consistent with the PAM-L2 hypothesis that L2 vocabulary expansion
is correlated with changes to L2 segmental perception, reflecting better reattunement and rephonologisation. The experiment thus supports the proposed relationship between L2 vocabulary size and L2 segmental perception in which L2 vocabulary may be seen as connected to the reattunement of the perception of L2 phones.

The study also shows that L2 assimilation and relative ease of discrimination of the chosen L2 contrasts reflects the perceived similarities between the L1 and L2, as predicted by the Perceptual Assimilation Model (and by SLM). The experiments also show that these general patterns remain consistent as L2 vocabulary and segmental perception develops, in line with many results asserting the pervasive effect of the perceptual relationship between L1 and L2 phones even in advanced learners.

The results of Experiment 2 of Chapter 6 revealed very consistent and successful L2 vowel assimilation by Japanese learners who were tested first within weeks of arrival in Australia and again after 6-8 months of L2 immersion. The experiment also revealed that the learners were very successful in their discrimination of three of the four L2-vowel contrasts of interest at both testing times, that they had a gain in average vocabulary size of almost 700 words over the course of their immersion, and again, that their perception reflected a persistent effect of their L1 background (similar to that evident in Experiment 1), and the previous study (see Chapter 5).

The results from Experiment 2 do not support the general Immersion Hypothesis prediction (Tsukada et al., 2005) that improved L2 perception is associated with increased L2 immersion duration: There was no evidence of improvement in the group of learners’ segmental L2 perception between the two testing times. We suggest that this lack of improvement is due to this particular group of learners exhibiting very successful L2-L1 vowel assimilation and high levels of discrimination accuracy – alongside a relatively large L2 vocabulary - already at the first testing time. Indeed, as a group, the learners performed on par with a group of native AusE speakers tested on those same contrasts in a previous study (Bundgaard-Nielsen, Best, Kroos, & Tyler, 2008; see Appendix 4), indicating that their L2 perception even at test time 1 (Experiment 1; $M = 7$ weeks of immersion) was most likely sufficient for their day-to-day interactions in English. We do, however, also
note that a closer inspection of the performance of the two participants who had L2 vocabularies below 6000 words at the first testing session did improve in the second test session. We take this as support for the hypothesis that improvement to L2 segmental perception is associated with expansion of a smaller L2 vocabulary (<6000 words), up until the growing L2 vocabulary forces the learner to settle on an accented version of the L2 phonology.

To summarise, the results from the two chapters on L2 perception (Chapters 5 and 6) show a significant difference in the consistency of the L2 to L1 vowel assimilation and in L2 vowel discrimination accuracy in L2 learners with larger versus smaller L2 vocabularies. The results are thus compatible with the PAM-L2 and Vocab Model predictions, in terms of the role of L2 vocabulary in overall L2 perception. The findings also highlight similarities between infant/child L1 acquisition and late/adult L2 acquisition, and are thus also consistent with the propositions of PAM/PAM-L2 and SLM that even late L2 acquisition is analogous to L1 acquisition in that the learning mechanisms employed in L1 acquisition are available to the late learner of a second language.

8.2.2 L2 Production and L2 Vocabulary Size

Experiment 1 of Chapter 7 focussed on the role of L2 vocabulary size in L2 vowel intelligibility. As expected within the PAM-L2/Vocab Model frameworks, and parallel to the findings for the perceptual assimilation and discrimination of Australian English vowels presented in Chapters 5 and 6, Japanese learners with a larger L2 inventory differed from Japanese learners with smaller L2 inventories in their L2 vowel production, measured by the number of L2 vowels that they produced in an intelligible manner and their overall intelligibility score. The results from this first production experiment thus support the PAM-L2/Vocab Model hypothesis that L2 vocabulary development is positively associated with improvements to L2 production accuracy, even in learners who have spent a very limited time in an L2 speaking environment.

The second experiment of Chapter 7 tested two competing hypotheses:

1) the general assumption that early L2 immersion leads to improved L2 production, irrespective of L2 vocabulary size; and
2) the PAM-L2/Vocab Model hypothesis that despite their increased immersion duration, L2 learners with larger (>6000) L2 vocabularies would not improve their L2 production during extended L2 immersion.

The results support the PAM-L2/Vocab Model hypothesis that the group of Japanese L2 learners of English did not improve their L2 vocabulary size, their overall L2 vowel intelligibility, or the number of L2 vowels they produced in an intelligible manner, despite the additional 6-8 months of L2 immersion. Further support for the Vocab Model predictions comes from the separate analysis of the intelligibility of the two participants whose L2 vocabularies were below 6000 words at the first test session. Indeed, the results indicate that only these two participants, and not their similarly immersed fellow participants, improved in their overall L2 intelligibility at the second test session.

The results from the two experiments in Chapter 7 support the PAM-L2/Vocab Model hypothesis of Vocab-driven improvements to L2 segmental production in two ways: Firstly, Experiment 1 supports the prediction that newly immersed L2 learners with larger L2 vocabularies produce more intelligible L2 vowels than newly immersed L2 learners with smaller L2 vocabularies. Secondly, Experiment 2 supports the PAM-L1/Vocab Model prediction that increased exposure/longer L2 immersion does not necessarily lead L2 learners with larger (>6000 words) L2 vocabularies to further re-attune and rephonologise in their L2 as they may have already settled on an (accented) L2 phonology, but does foster that process in learners with smaller vocabularies (< 6000).

8.3 Conclusions

8.3.1 L2 Vocabulary Size Matters

The results from the series of experiments presented in this thesis allow a number of conclusions to be drawn.

Firstly, the results are compatible with PAM-L2 and the Vocab Model, and in particular with the prediction that early L2 vocabulary expansion facilitates improvements to L2 vowel perception and production, while a larger L2 vocabulary curtails further improvement to L2 vowel perception and production, as the increasingly large L2
vocabulary may force the L2 learner to settle on an (accented) L2 phonology. It may be useful for future research on L2 vowel acquisition to consider measures of L2 vocabulary size in their participant selection. Such an addition to the selection process will ensure that L2 learners with similar immersion and learning histories but vastly different L2 vocabularies will not be grouped together, which would result in large within-group differences.

Secondly, the results are also in line with the PAM-L2 and Vocab Model assumption that similar mechanisms underlie L1 and L2 language acquisition. In particular, the results are compatible with experimental findings from child language acquisition that large L1 vocabularies in the second year of life are associated with better segmental production. Future research may consider ways in which L2 perception, production and vocabulary development may be studied in a more integrated fashion. And while recognising that the relationship among segmental production, perception, and vocabulary expansion may be bi-directionally or even cyclically linked in the acquisition process, as has been suggested for child language acquisition (Smith, McGregor, & Demille, 2006), we tentatively suggest that changes in perception initially precede those in production. This possible primacy of perception may also be reflected in the observed lag between change in L2 perception and production.

Thirdly, while it has been suggested that the most significant changes in L2 perception may occur during the first year of L2 learning, the present experiments also indicate that perceptual change may happen very rapidly (at least in a group of learners who have previously been exposed to an accented version of the L2 to some extent), with no further improvement in the ensuing six months of immersion in learners who began their immersion with >6000 L2 words. What is not clear is whether the rapid perceptual readjustments displayed by the learners reflect a general pattern, or whether the particular L1-L2 pair and the L2 learning history of the participants partly explain the finding. Ideally, future research will address the question of the role of L2 vocabulary acquisition in L2 rephonologisation in learners who are truly naïve upon their arrival in the L2 language environment.
8.3.2 The Benefit of a ‘Whole Systems’ Approach

The work presented in Chapters 5 and 6 of the thesis also indicate an important methodological advantage in the use of a Whole Systems approach to L2 vowel acquisition. The central benefit of the Whole System approach to L2-L1 vowel perception, adopted in the perceptual experiments in Chapters 5 and 6, is that the learners are able to use all native sensitivities in their L2 phone assimilation. The results obtained by this approach indicate a much greater perceptual flexibility in the learners than would be expected from contrastive analyses of the vowel inventories of Japanese and English.

The value of a Whole Systems approach, and the extent of the perceptual flexibility of the Japanese learners of English tested here, may be more fully appreciated in the light of a comparison between the assimilation results of the Japanese learners of AusE, with access to all native phones and phone combinations, and that of a group of native listeners of Australian English. While the Japanese L2 learners assimilated between 14 (LV group) and 16 (HV group) of the AusE vowels to Japanese categories in Chapters 5 (Bundgaard-Nielsen et al., submitted-a), and Chapter 6 (Bundgaard-Nielsen et al., submitted-b), using the 50% assimilation criterion, a group of native listeners (see Appendix 1: (Bundgaard-Nielsen, Best, Kroos, & Tyler, under revision), identified 16 of the 18 vowels (i:, ɪ, ɛ:, æ, æ, ɛ, ɔ:, ʊ, ʊ:, ɔ, ɔ, ʌ, ʌ, ɵ, ɵ, ɒ, ɒ/) as intended (using the 50% criterion). While the two groups naturally did not use the same categories for assimilation/identification, as the Japanese used their L1 Japanese vowels and the AusE listeners used their AusE categories, the comparable number of AusE vowels assimilated or identified by the groups indicate that both speaker groups were able to differentiate the vowels on the basis of spectral and durational differences to (almost) the same extent. Of special interest, and central to the performance of all three groups, was that even native Australian English listeners had problems successfully identifying Australian English /ɪə/ (typically identified as /i:/), and likewise discriminating it from /iː/. As argued in Appendix 1 (Bundgaard-Nielsen, Best, Kroos, & Tyler, under revision), we believe this may be due to the apparent near-merger of these two vowels in the particular dialect of Australian English spoken in Western Sydney.
8.3.3 Different Assimilation Criteria

Finally, the results in Chapter 6 of the thesis also highlight the importance of choosing a meaningful criterion for L2 vowel assimilation, though L2 vocabulary size-based differences in assimilation consistency were evident using both the lenient 50% criterion and the more stringent 70% criterion. In choosing the most meaningful criterion, it is important to consider the behaviour of native listeners.

Referring to the pattern of native vowel identification reported in Appendix 1, we thus suggest that, while the more stringent criterion of 70% assimilation/identification only enhanced the observed support for a vocabulary-based difference between the L2 learners, the more stringent criterion may be too strict for vowel research in general, due to the inherently continuous perception of vowels. In imposing a similarly harsh criterion on the identification pattern of the native listeners, we thus find that a mere 11 of 18 native vowels are identified as intended, leading to a problematic apparent inability in the identification abilities even in native speakers. Using a 50% criterion, however, the native listeners in the experiment successfully identified 16 of the 18 vowels (\(M = 76\%\), range: 56-92%), a performance which reflects the phonological system of Australian English very well, given the recently proposed near-merger of two Australian English vowels /i:/ and /\(\text{\textalpha}\)/, taking the maximum number of Australian English vowels that can be reliably identified by native speakers to 17.

It is, however, very likely that a stricter criterion, such as the 70% criterion, will be useful in consonant research, due to the more categorical perception typically found in consonant perception.

8.4 New Directions

The experiments presented in this thesis raise a number of important questions to be answered in future research on L2 phonetic and phonological acquisition and the developmental trajectory of phonetic/phonological re-organisation in late L2 acquisition.

Firstly, the present research focused solely on vowel perception. Further research must focus on possible differences in reattunement and rephonologisation for vowels versus
consonants. It is likely that perceptual differences (and underlying acoustic-articulatory differences) in vowels and consonants lead to different acquisition trajectories, as well as to differences in the suitable assimilation criteria. Within the PAM-L2/Vocab Model framework, however, we propose that L2 consonant acquisition may follow a later reattunement and rephonologisation trajectory, parallel to the observed pattern of timing differences in L1 perceptual attunement in infancy, where native-language consonant attunement occurs approximately 4 months after attunement to native vowels.

Secondly, the present research, as well as previous research, suggests that a learner’s L2 phonological development plateaus within the first year of acquisition in immersion settings. It is also suggested that the most significant changes in non-native vowel perception are apparent after only 6-12 months of L2 immersion for adult learners (Flege & Liu, 2001; Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Tsukada et al., 2005). We therefore encourage a meticulous examination of the relationship between L2 vocabulary growth and L2 phonological acquisition during the first 12 months of immersion. Such studies would help determine exactly when during the first 12 months of immersion the learner reaches this plateau, when and how the phonological re-organisation begins or peaks, and the extent to which it levels out.

Thirdly, the present research suggests that future research would benefit from developing new methods to assess L2 rephonologisation. If indeed learners shift their attention from surface L2 to L1 phonetic similarities to the higher-order phonological properties of their L2, the task of assimilating L2 phones to L1 categories may not allow advanced/competent L2 learners to adequately display their L2 phonological competence though they may well be able to assimilate the L2 phones to L1 categories on the basis of surface phonetic similarities.

Fourthly, the results obtained in the thesis by the application of the ‘whole-system’ approach also indicate that the Japanese learners of English in reality face very little pressure to rephonologise on the basis of L2 vocabulary acquisition, causing their performance to reach a plateau very early in their L2 immersion. This observation invites further consideration of the results presented in Chapters 5, 6, and 7 in the light of future
modification of both PAM-L2 and the Vocab Model to account for the modulating effect of the size of the L1 phonological system, as well as the role of native phonotactics, on rephonologisation in response to L2 vocabulary expansion.

Associated with this last point is also the possible benefit to future research on nonnative perception from including, not just native vowel combinations as response-options in cross-language vowel perception studies, but also vowel-like consonants (liquids, glides). Likewise, it is possible that the inclusion of vowel-like elements in some studies of nonnative consonant perception, in particular the perception of vowel-like consonants such as liquids and glides, will increase our knowledge of how perceptually flexible L2 learners really are. The consideration of these additional phonological elements and combinations there off, may also help identify the contributions from native phonotactics to L2 speech perception.

Indeed, it is possible to imagine that native phonotactic constraints may systematically influence the perception of nonnative phones through restrictions imposed on the number of possible combinations of native phonological elements that are available to the learner. To exemplify, we propose that Spanish (for a discussion of the Spanish monophthong inventory relative to English, see Bradlow [1995]) and Japanese learners of English (or another vowel rich language, such as German or Danish), may not perform in a similar manner, despite the fact that their phonological inventories are rather similar in that they each contain only five vowels. This prediction rests on the very different phonotactics of the two languages: While Japanese allows the stringing-together of stressed single mora vowels without reduction in their duration or significant spectral change, Japanese does not contain any diphthongs. Spanish, on the other hand, contains a number of diphthong types such as diphthongs with stress on the first element (i.e., a combination of a stressed vowel + an unstressed vowel), hiatus (i.e., the combination of two stressed vowels), diphthongs with stress on the second vowel element (unstressed + stressed), and a number of exceptions in which an acute accent is placed on the stressed vowel element.

Finally, in addition to fostering new ways of understanding change in L2 segmental perception and production providing a new method to study L2 segmental perception and
offering a more systematic evaluation of the success of L2 segmental perception, the thesis also raises new questions that, if answered, will contribute to our knowledge of L2 acquisition. In particular, the thesis highlights the necessity of further examination of the direction of the identified association between L2 segmental perception and production and L2 vocabulary development. Drawing parallels to studies of child L1 acquisition, we may speculate that this relationship could be either cyclical or circular. The validity of different theoretical frameworks, with very different interpretations of the timing in change in perception relative to production, may also be assessed from such research. The thesis also highlights the likely gains from the systematic study of adult L2 immersion acquisition in completely naïve learners with respect to the timing of change in segmental perception and production in early immersion acquisition.
9 REFERENCES


10 APPENDIX 1: ARE AUSTRALIAN ENGLISH R-VOWELS MERGING WITH THEIR R-LESS MATES? A VOWEL SUBSYSTEM DISTINCTION IN COLLAPSE

Under revision for resubmission to Journal of Phonetics

10.1 Abstract

Theory and research has failed to identify the underlying causes of the abundance of r-context vowels in mergers and near-mergers, including the vowels in *here* and *hair* in New Zealand English, *ferry* and *furry* in Philadelphian English, and *source* and *sauce* in New York English. We propose a new framework that posits a concentric-shell organisation among vowel subsystems. We hypothesise that mergers and near-mergers may reflect a systematic loss of shell-differentiating features, such as the loss of rhoticity, which may lead to collapse of an r-context vowel shell into the more basic shell of otherwise-comparable monophthongs. To test whether lack of rhoticity is associated with evidence of mergers/near-mergers between r-context vowels and their similar r-free vowel mates, we examined the identification of the Australian English vowel inventory by native speakers, along with their discrimination of each r-context vowel and its corresponding r-free monophthong. Our prediction that the r-context distinction is disappearing in AusE was reflected in a systematic pattern of near-merger across r-context vowels and their r-free mates. The new concentric-shell vowel framework thus provides novel insights into the systemic nature of vowel inventories and its reflection in mergers and near-mergers.
10.2 Introduction

10.2.1 Background

By combining a number of principles and experimental findings about vowel mergers/near-mergers involving r-context vowels and vowel chain shifts, the present paper probes the proposal of an underlying concentric-shell organisation of vowel systems.

Mergers of segmental contrasts refer to situations in which the acoustic and articulatory properties of a given contrast become so similar that the two phonetic categories merge into one category (Labov, 1994). Mergers may be context-free, that is, occurring in all phonetic contexts such as the merger of the vowels in cot /a/ and caught /ɔ/ in some dialects of American English (AmE) and the merger of vowels in merry, Mary, marry in New England American English (Majors, 2005; Labov et al., 2006). Alternatively they may be context-conditioned, that is, occurring only in certain phonetic contexts and not in others such as the merger of tense and lax monophthongs /ɪl-/lɪl, /ʌl-/lʌl, and /æl-/læl preceding an /l/ in Utah English (Di Paolo, 1988; Di Paolo & Faber, 1990) and of the centring diphthong in lure (/əʊl/) with lore (/o:l/), but not in tour and tore, in Australian English (AusE) (Bernard, 1970; Cox, 1998; Cox, 2006b).

Whereas mergers reflect the loss of acoustic and perceptual differentiation of the merging vowels, near-mergers are phonological contrasts that maintain some reliable acoustic difference in production, yet become indistinguishable to the native listener. From a theoretical perspective, near-mergers are interesting as they are reflective of a vowels system in a state of flux. Near-mergers have been reported for a number of dialects of English, including near-merger of the vowels in the words here and hair in New Zealand English (NZE) (Gordon & MacLagen, 2001; Hay, Warren, & Drager, 2006), of ferry and furry in Philadelphian English, and source and sauce in New York English (Labov, Karan, & Miller, 1991; Labov, 1994; Thomas, 2006).

One interesting aspect of the near-mergers (and mergers) observed in previous studies is that they often involve r-context vowels, that is, the vowels in hear, hair, hard, horde, and heard, irrespective of whether the dialect of English examined is rhotic (such as AmE) or non-rhotic (such as AusE and NZE) (Gordon & MacLagen, 2001; Hay et al., 2006; Labov...
et al 1991; Labov, 1994; Thomas, 2006). Whether vowel rhoticity is realised as a [J], as an r-colouring of the vowel, as vowel lengthening, has disappeared altogether (Lindau, 1985), or is realised as an offglide from the target vowel toward /ə/ as in AusE /əə/ (Cox, 1999; Cox, 2006a), the abundance of r-context vowels in near-mergers raises the possibility of a systematic underlying phenomenon that has thus far gone unaddressed.

Mergers and near-mergers interact with vowel chain shifts, that is, the systematic raising or lowering, fronting-or backing (in terms of the positioning of the articulators during production, roughly corresponding to the placement of the vowels in a F1/F2 vowel chart) of one or several vowels in a system, such as the raising and fronting of the open back vowels in the Northern Cities Shift in the US (Labov, 1994; Labov, Ash, & Boberg, 2006). However, just as the systemic nature of mergers/near-mergers has been largely overlooked, so has the exact nature of the relationship between mergers/near-mergers and vowel chain shifts. On the one hand, shifts may arise from the internally motivated restoration of symmetry to the vowel system following the addition of phonemes to the vowel system or the loss of phonemes through a (series of) merger(s) by related vowels in the system (Schwartz, Boë, Vallée, & Abry, 1997b; Thomas, 2006). On the other hand, systemic vowels shifts, in which tense and front vowels typically rise along a peripheral path in the vowel space, while lax and more central vowels typically fall along a non-peripheral path, may provoke mergers or splits to restore the relative symmetry of the system (Labov & Baranowski, 2006; Labov, 1994; 2006; Labov et al., 2006).

A series of cross-language trends in vowel system organisation have been proposed that may potentially help to illuminate the source of, and principles underlying, these apparent systematic and related changes to vowel systems (Maddieson, 1984; Schwartz et al., 1997b). Based on analyses of the vowel systems of 534 languages in the SUPERB UPSID (SUPERB UCLA Phonological Segment Inventory Database), Schwartz et al. (1997b) argue that vowel systems have relatively predictable concentric shell-like organisation: Most languages have a primary shell or subsystem of vowels consisting of 3-9 monophthongs. To this system, additional systems may be added, by exploiting new vowel features or parameters. In additional vowel subsystems, of typically 1-7 vowels each, the parameter added to monophthongs is typically one of the following: extended duration
Within this framework, mergers and near-mergers could be conceptualised as the collapse of one vowel system into another. If the distinguishing parameter of a secondary or tertiary subsystem is lost, we predict that it will merge or near-merge into another system such as the primary system. The example we consider here is the loss of r-contextualizing in AusE and other non-rhotic dialects of English, which we posit to lead to a systematic set of mergers and near-mergers. That is, we predict that the loss of vowel rhoticity in non-rhotic dialects will lead to systematic mergers and near-mergers between r-context vowels and their monophthongal mates in the primary vowel subsystem. This possibility has not been previously examined. The only closely relevant findings available are from a recent study on the dialect of AusE that is spoken in Western Sydney (AusEWS) (Bundgaard-Nielsen, Best, Tyler, & Kroos, 2008). That study found evidence for one r-context vowel near-merger (i.e., chance-level discrimination of a vowel contrast that displayed significant acoustic differentiation), namely that of AusEWS r-context vowel /eθ/ (as in here) and tense /i/ (as in he) in a closed syllable context. These findings are consistent with the proposed framework of a concentric shell-like organisation of vowel subsystems, and provide the possibility of testing the general hypotheses that the loss of the differentiating element in a vowel shell leads to the collapse of that vowel subsystem into a more basic vowel system. In this case the loss of any obvious vowel rhoticity in the r-context vowels in AusE may lead to their collapse into the primary monophthongal subsystem. We therefore extend the present study to the full set of r-context vowels and their r-less mates for a more systematic approach.

10.2.2 Application of the Shell Framework to a Specific Vowel System: AusE

Australian English (AusE) is a non-rhotic variety of English (Cox, 2006b) with a very strong tendency for vowel reduction in casual speech compared to, for instance, AmE. The AusE vowel inventory consists of /iː/, /e/, /e/ː, /æ/, /ɛ/, /ɒ/, /u/ː, /ɔ/, /ɒ/, /æ/, /e/, /æ/ for the vowels in the words heed, hid, head, haired, heard, had, hud, hard, horde, hood,
who’d, hod, heared, hade, hide, hoyd, howd, hode (Harrington et al., 1997; Cox, 2006a). Since the middle of the previous century, the AusE centring diphthong /uə/ as in the word lure has reportedly been lost through a merger with /oə/ (Harrington et al., 1997; Cox, 1999; Cox, 2006a), and /iː, ɪː, eː, oː, uː, ɪə, eə, əʊ, ɔː/ have been fronted and raised, and the spectral similarities between certain vowel pairings has increased, e.g. /e, ɛː/ and /e, ɛː/ (Cox, 1999; 2006a).

AusE vowels are traditionally divided into two subsystems: monophthongs (/iː, ɪː, eː, əː, ɛ, ɛː, ɔː, u, uː, o/), and diphthongs (/eː, ɪə, æə, æ, o, ɔ, əʊ, əu/), though this division no-longer reflects the actual monophthongal nature of the vowel /eː/ that has lost its second vowel element /æ/. The monophthongs are typically sub-divided into long and tense (/iː, ɪː, eː, əː; u, uː, o/), versus short and lax (/i, ɪ, e, ə, æ, ɛ, u, o/), vowels, reflecting durational and phonotactic differences (e.g., the fact that tense vowels may occur in open syllables, whereas lax vowels may not). We note that the AusE long vowel /iː/ is often characterised by a marked onglide from a schwa-like onset, especially in Broad AusE (Cox, 1996; Harrington et al., 1997; Cox, 1999; Cox, 2006a). The long monophthongs and the diphthongs have approximately the same duration, and the AusE monophthongs are generally around 60% of the duration of these longer vowels (Cox, 1999).

The classification of the historically r-context vowels is vexed, however, if we attempt to limit ourselves to the simple monophthong/diphthong dichotomy. Characterizing /iə, ɛː/ as monophthongs or diphthongs is problematic. In open syllables they have a tendency toward the diphthongal quality of an offglide toward a second target (schwa), but in closed syllables, the second vowel target for /iə, ɛː/ is greatly reduced, giving both vowels a long-monophthong character (Cox, 1996; Cox, 1999; Cox, 2006a). Moreover, both vowels are rhotic in hyper-articulated open syllables (Harrington et al., 1997), indicating a persisting underlying rhotic element. The classification of /iː, ɛː, ɔː/ as monophthongs or diphthongs is also problematic. Though these vowels are indeed longer than the AusE lax monophthongs and can appear in open syllables like the long monophthongs, these features are also shared by other centring diphthongs/r-context vowels. Thus, r-context vowels even in the non-rhotic dialect AusE seem to have certain common characteristics, and behave in
ways that make them difficult to consistently categorise as monophthongs or as diphthongs.

By applying the shell-based approach (Maddieson, 1984; Schwartz et al., 1997b) described earlier, it is possible to reanalyse the AusE vowel system in a more consistent way by identifying a primary shell of long and tense vowels /i:/, /u:/, and a secondary shell of short and lax vowels /ɪ, ə, æ, ɛ, ʊ, ɔ/, one of diphthongs /æi, ɑe, ɔi, ɔu, ɑʊ/, and one consisting of the r-context vowels /ɪə, ɛə, ɜə, ɔə/. In conjunction with Labov’s principles for chain shifting (Labov et al., 2006; Labov, 2006), this new approach predicts that the loss or loss-in-progress of the feature of rhoticity in AusE /ɪə, ɛə, ɜə, ɔə/ will lead to the r-context vowel shell collapsing into the corresponding vowels of the r-free monophthong shell /i:, ɛ, ɔ, ɔ/, such that native speakers will have notable difficulties in discriminating /i:/ vs /ɪə/, /ɛ/ vs /ɛə/, /ɜ/ vs /ɜə/, and /ɔ/ vs /ɔə/, and will show marked inconsistencies in systematically identifying them as separate vowel categories.

For the proposed concentric shell organisation of the r-context vowels in AusE to be supported, we would expect the loss of discriminability to affect all four r-context/r-less matches, though the degree to which the pairs are discriminable will also reflect the varying spectral and durational differences between the pairs. In detail, we expect poor discrimination for the high-front vowel contrast /i:/ and /ɪə/. These vowels are of similar duration, and mainly differentiated in terms of acoustics by an onglide for /i:/ and an offglide for /ɪə/. For vowel pair /ɛ/ and /ɛ:/, which differ in duration, and also possibly acoustically in an offglide for /ɛ:/ ([ɛə]), we predict moderately good discrimination. For /ɛ/ and /ɛ:/, which differ in both duration and possibly in spectral qualities, we predict moderately good discrimination. Finally, for the open back vowels /ɔ/ and /ɔ:/, we expect excellent discrimination due to both pervasive spectral and durational differences.

The predicted discrimination levels are expected to be mirrored in the identification task, and, importantly, we expect a pattern of confusion that is not supported by mere psychoacoustic predictions. In a psychoacoustic framework, we would expect any confusion between the merging/near-merging vowels to be symmetrical, i.e., the r-context vowels would be identified as their r-less mates to the same extent that the r-less mates
would be identified as their r-context counterparts. However, within the framework laid out here, we would expect more r-context vowels to be identified as their r-less counterparts than vice versa.

10.3 Method

10.3.1 Stimuli

Eight male L1 speakers of AusEWS (Mage = 22.7 years), with AusE speaking parents, produced five blocked and randomised repetitions of the 18 stressed AusE vowels /ɪ, ɪ, ɛ, ɛ:, ɔ:, ɔ, ʊ, ʊ, ʊ, ɪ, ɐɪ, ə, ɔɪ, əʊ, ɐʊ/ in a /hVba/ context. Recordings from three speakers were selected for our stimulus material on the basis of similar voice quality as well as matching reading style and speed: Speaking rate was commensurate with normal conversation and the speakers were instructed to talk as if to a friend. Three tokens from each speaker were selected on the basis of most-similar intonation pattern and speaking rate of the individual tokens. The /hVba/ context was chosen to minimise initial and medial consonant co-articulation with the vowels (see Strange et al., 2007).

The recording took place in a sound-attenuated room at University of Western Sydney, using a Shure SM10A headset microphone, an LG laptop computer and an external soundcard (Edirol UA-25). The recordings were high-pass filtered with a cut-off frequency of 75Hz in Cool Edit. The intensity of all words was adjusted in Praat so that the intensity of the target vowel was equal to 70dB (Boersma & Weenink, 2008). Vowel onset was defined as the beginning of the first regular and recognisable pitch pulse, and vowel offset was defined as the cessation of regular pitch pulsing (see Figure 1).
Figure 10.1. Stimulus segmentation in Praat.

Vowel duration as well as F1, F2, and F3 values of the target vowel were assessed via Praat and are presented in Table 1. Dynamic F1/F2 plot of the AusEWS monophthongs at 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 95% of the vowel duration are presented in Figure 10.2-10.4. A highly systematic rapid change in the vowel trajectories in the final 25% of the vowels is due to the lips closing fast to form the following /b/. For ease of comparison, the vowels are presented in their three subgroups; monophthongs, diphthongs, and r-context vowels. Figure 10.5 presents the F1 and F2 trajectories for the stimulus final ‘a’ at 25%, 50%, and 75% for each of the three vowel sub-groups; these are clearly quite similar, reflecting no systematic difference across the three vowel subsets.
Figure 10.2. Dynamic vowel trajectories for AusE<sub>WS</sub> monophthongs.

Figure 10.3. Dynamic vowel trajectories for AusE<sub>WS</sub> diphthongs.
Figure 10.4. Dynamic vowel trajectories for AusE<sub>WS</sub> r-context vowels.

Figure 10.5. Dynamic vowel trajectories for AusE<sub>WS</sub> word final /a/ (schwa).
The formant values presented in Table 10.1 are generally consistent with those of Cox’s (2006) AusENS. However, we did find that the reported F1/F2 values of the diphthong transcribed as /æɪ/ in her measures of AusENS are so different from the AusENS as well as AusEWS F1/F2 values for /æ/ and the first target of /æʊ/ that /æɪ/ probably should be transcribed as [eɪ]. This much better reflects the acoustic similarity between the first target of this diphthong [eɪ] and the AusENS and AusEWS vowels /e/ and /eː/. 
Table 10.1. Vowel duration and F1, F2, and F3 values (in Hz) at 25%, 50%, and 75% of the vowel.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Dur. (ms)</th>
<th>F1 (Hz)</th>
<th>F2 (Hz)</th>
<th>F3 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25%</td>
<td>50%</td>
<td>75%</td>
<td>25%</td>
</tr>
<tr>
<td>iːd</td>
<td>128</td>
<td>335</td>
<td>317</td>
<td>310</td>
</tr>
<tr>
<td>lːd</td>
<td>59</td>
<td>333</td>
<td>342</td>
<td>344</td>
</tr>
<tr>
<td>lːal</td>
<td>129</td>
<td>327</td>
<td>325</td>
<td>341</td>
</tr>
<tr>
<td>lːel</td>
<td>74</td>
<td>547</td>
<td>558</td>
<td>528</td>
</tr>
<tr>
<td>lːed</td>
<td>143</td>
<td>512</td>
<td>523</td>
<td>526</td>
</tr>
<tr>
<td>lːel</td>
<td>85</td>
<td>711</td>
<td>695</td>
<td>646</td>
</tr>
<tr>
<td>lːed</td>
<td>130</td>
<td>540</td>
<td>489</td>
<td>409</td>
</tr>
<tr>
<td>lːal</td>
<td>142</td>
<td>708</td>
<td>731</td>
<td>681</td>
</tr>
<tr>
<td>lːel</td>
<td>147</td>
<td>708</td>
<td>702</td>
<td>603</td>
</tr>
<tr>
<td>lːel</td>
<td>73</td>
<td>698</td>
<td>678</td>
<td>628</td>
</tr>
<tr>
<td>lːed</td>
<td>141</td>
<td>672</td>
<td>682</td>
<td>657</td>
</tr>
<tr>
<td>lːd</td>
<td>70</td>
<td>618</td>
<td>608</td>
<td>552</td>
</tr>
<tr>
<td>lːal</td>
<td>119</td>
<td>454</td>
<td>497</td>
<td>491</td>
</tr>
<tr>
<td>lːel</td>
<td>142</td>
<td>499</td>
<td>474</td>
<td>387</td>
</tr>
<tr>
<td>lːel</td>
<td>76</td>
<td>378</td>
<td>385</td>
<td>361</td>
</tr>
<tr>
<td>lːed</td>
<td>110</td>
<td>376</td>
<td>378</td>
<td>371</td>
</tr>
<tr>
<td>lːal</td>
<td>116</td>
<td>570</td>
<td>522</td>
<td>467</td>
</tr>
<tr>
<td>lːed</td>
<td>144</td>
<td>487</td>
<td>504</td>
<td>490</td>
</tr>
</tbody>
</table>
It is worth noting that the dynamic trajectories of the r-context vowels are much more similar to those of the r-less monophthongs than the larger, more curved trajectories of the monophthongs. We take this as sign of the greater similarity between r-vowels and their monophthongal mates, than between the r-context vowels and the AusE diphthongs.

Additional analyses of the vowels in our four contrast pairs /iː/-/ɪəl, /eɪ/-/ɛːl; /ɛl/-/ɛːl, and /ɔ/-/ɔːl/ (see Figure 10.6 for matched trajectories) were conducted. Though parallel raising and fronting of /ɜː/ and /ʌː/ has been reported in AusE (Cox, 1999), the central r-context vowel /ɜː/ was excluded from the analysis as it has no clear monophthongal counterpart with which to merge and may have a unique status in the vowel system as the only non-peripheral stressed vowel (Schwartz et al., 1997b). Nine dependent variables (F1, F2, and F3 at 25%, 50%, and 75% of the duration of the vowel) were analysed in a series of 4x2 MANOVAs (IVs: Contrast pair (fiː/-/ɪəl, /ɛl/-/ɛːl; /ɛl/-/ɛːl, and /ɔ/-/ɔːl) and vowel type (r-context vowel or monophthong).
The analyses showed that the target vowels /i:/-/ə/ differed only in F1 and F2 (though not F3) at 75% of the vowel, F1: $F(1,16) = 11.869, p < .01$, F2: $F(1,16) = 15.880, p < .01$; all other vowel formant comparisons ns, likely reflecting a difference in offgliding in the diphthong /ə/ but not the monophthong /i/. A separate Analysis of Variance (ANOVA) revealed no significant difference in vowel duration for this pair, ns. Vowels /e/-/e:/ did not differ significantly in formant values at any of the three sampling times, though a separate ANOVA revealed a significant difference in duration for this pair, $F(1,16) = 184.976, p < .01$. The vowels /e/-/e:/ differed significantly in F1 and F2 (but not F3) at 75% of the vowel, F1: $F(1,16) = 10.457, p < .01$, F2: $F(1,16) = 12.127, p < .01$, also likely indicating a difference in offgliding. The vowel pair /e/-/e:/ also differed in duration, as shown by a separate ANOVA, $F(1,16) = 221.38, p < .001$. Finally, the target vowels /α/-/α:/ in F1 and F2 at 25%, F1: $F(1,16) = 17.685, p < .01$, F2: $F(1,16) = 94.229, p < .01$, and 50%, F1: $F(1,16) = 53.631, p < .01$, F2: $F(1,16) = 44.957, p < .01$, and in F1 at 75%, $F(1,16) = 12,963), p < .01$. The vowels /α/-/α:/ also differed significantly in duration as revealed by a separate ANOVA, $F(1,16) = 47.818, p < .01$. Finally, we note that all AusEWS vowel trajectories clearly reflected a movement towards the articulation of the following /b/ in the second half of the vowel trajectories, thus masking any potential offglide or centring glide in /ə/, or any of the other vowels.

10.3.2 Participants

Participants were 12 native speakers of AusEWS ($M_{age} = 20.9$ years), with native AusE speaking parents. All were first-year psychology students at University of Western Sydney and participated for course credit. Data from one additional participant was discarded for failure to follow instructions during testing, and two more participants were rejected as they did not meet the recruitment criterion of monolingual AusE language background.

10.3.3 Procedure

All testing was done in a sound-attenuated room at MARCS Auditory Laboratories at UWS. Stimuli were presented on an Apple MacBook using PsyScope X (provide a reference) over studio headphones. The participants first performed a series of AXB discrimination tasks of the AusEWS r-context vowels and their corresponding monophthong pairings: /i:/-/ə/; /e/-/e:/; /e/-/e:/; and /α/-/α:/.
Following the discrimination task, participants listened to all 18 AusEWS vowels, including /ɪː/, in a /hVba/ context (N = 324) and identified them using a grid of nonsense keywords corresponding to all English monophthongs and diphthongs. They also rated goodness-of-fit from 1 (poor) to 7 (excellent) for their chosen keyword for each token; the rating was collected following a second repetition of a given token immediately after its keyword judgment.

10.4 Results

10.4.1 AXB Discrimination

The mean percent correct responses for each contrast of the four contrasts are presented in Figure 10.7. Discrimination ranged from just above chance (60% correct) for /ɪː/-/ɒː/, to good (80%) for /ɛ/-/ɛː/, to very good (88%) for /ɛ/-/ɛː/, and excellent (93% correct) for /ɑ/-/ɒː:.

![Figure 10.7: % correct discrimination for the four vowel contrasts.](image)

To investigate whether the results followed the predicted pattern of poor to good discrimination on the +high, front to -high, back articulatory dimension, we conducted a trend analysis. A linear trend contrast on the predicted order of difficulty, as listed above,
was significant, $M = 26.8$, $SE = 22$, $F(1,11) = 151.22$, $p < .001$, as was the quadratic trend, $M = 0.25$, $SE = 0.02$, $F(1,11) = 12.66$, $p < .01$.

### 10.4.2 Identification

Table 10.2 presents the confusion matrix for the 18 AusEWS vowels in percentages.
Table 10.2. Confusion matrix in percentages for the 18 AusEWS vowels.

<table>
<thead>
<tr>
<th></th>
<th>/id</th>
<th>/l</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
<th>/æd</th>
</tr>
</thead>
<tbody>
<tr>
<td>/id</td>
<td>63</td>
<td>8</td>
<td>8</td>
<td>14</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/l</td>
<td>15</td>
<td>72</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>41</td>
<td>18</td>
<td>28</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>13</td>
<td>1</td>
<td>76</td>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>8</td>
<td>5</td>
<td>14</td>
<td>56</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>90</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>78</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>85</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>10</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td>78</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>1</td>
<td>2</td>
<td>19</td>
<td></td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>91</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>63</td>
<td>2</td>
<td>5</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td></td>
<td>8</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æd</td>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A vowel was defined as categorised if it was identified as an instance of one L1 category in more than 50% of presentations (following Best, Faber, & Levitt, 1996). A total of 16 of
18 AusEWS vowels (/i/, /ɪ/, /æ/, /e/, /eɪ/, /u/, /o/, /æi/, /æe/, /oi/, /æo/, /əu/) were identified as intended by the listeners ($M = 76\%$, range: 56-92%). The mean goodness rating for these 16 vowels was 6.1 (range: 5.64-6.32). Orthographically inconsistent AusEWS /u/ was identified as AusEWS /u/ in 55% of instances, and identified as the intended target in only 23% of instances. AusEWS /ɪə/ did not reach our assimilation criterion, but there was a strong trend for /ɪə/ to be perceived as AusEWS /i:/ (41% of instances identified as /i:/).

Apart from /ɪə/ (58% identified as intended), and /u/, only the four r-context vowels in our contrast pairs received identification scores below 70% correct ($Mr = 53\%$). An inspection of the confusions for these vowels reveals that they were frequently perceived as instances of their respective monophthong mates. The identification accuracy for those four monophthongs was greater ($Mnon-r = 77\%$) than for the four r-context vowels. For both vowel groups, however, identification accuracy increased as the magnitude of acoustic differences increased between the paired items, in terms of duration as well as vowel quality (r-context vowels, from high-front to low-back: 27% → 56% → 66% → 63%; paired monophthongs, from high-front to low-back: 66% → 76% → 79% → 91%). The M identification as intended for the remaining 10 AusEWS vowels was 82%.

10.5 Discussion

We presented a new framework for understanding the systematic nature of mergers and near-mergers and tested the prediction that the loss of a differentiation vowel feature, such as r-context vowel rhoticity, may lead to a merger or near-merger of the r-context vowels in question and their most similar r-free counterparts in the vowel system.

Our results support our predictions: There was a clear decrement in discriminability of the AusEWS r-context vowels from their r-free mates, with accuracy ranging from very poor for the high-front /i:/-/ɪə/ to excellent for the back /ɑ/-/ɔ:/l. Furthermore, identification accuracy for the r-context vowels and their r-free mates /i/, /ɪə/, /e/, /eɪ/, /o/, /ɔ/ was lower than for the other AusEWS vowels /ɪ/, /æ/, /u/, /eɪ/, /æe/, /oi/, /æo/, /əu/, and the identification accuracy ranked from lowest for the high-front vowels /i:/-/ɪə/ to highest for the back vowels /ɑ/-/ɔ:/l. Finally, in a pattern of confusion incompatible with mere psychoacoustic predictions, more r-context vowels were identified as their r-less mates than vice versa.
The discrimination and identification results reflect the difference in the degree of acoustic similarity between the high, front r-context and the back r-context vowels and their r-free mates. We take this pattern of reduced discriminability and identifiability as evidence that native listeners did not perceive the r-context vowels and their r-free monophthong mates as completely independent vowel categories. Rather, they perceived the pairs as overlapping and differing systematically in degree of merger, from greatest merger for the high front pair and lowest for the back vowel pair.

The difference in the discrimination results (as well as the percentage of correctly identified tokens) for /iː-/ɪəl/ and /eɪ-/eɪːl/, /eɪ-/eɪːl/, and /oʊ-/oʊːl/ may reflect the fact that the merger of /ɪəl/ into tense /iːl/ entails fewer changes (to vowel duration and vowel quality), than does that of /eɪːl/ into /eɪl/, /eɪːl/ into /eɪl/, and /oʊːl/ into /oʊl/. This linear trend may also reflect inherent differences in the degree of overlap (due to articulatory constraint) on high/front vowels vs. back vowels.

The results support our reanalysis of the AusE vowel inventory according to the novel shell-based framework laid out earlier. It indicates the existence of a shell-like relationship among the r-context vowels /ɪə, eɪ, əː, oʊː/, and that these four vowels are merging with their corresponding monophthongs /iː, eɪ, eɪˈ, oʊː/ due to the loss of clear rhoticity in AusEWS. The results also support the finding of Bundgaard-Nielsen et al. (2008) that the loss of acoustic (and articulatory) separation in the closely occupied upper front vowel inventory has resulted in /iːl/ and /ɪəl/ already near-merged in AusEWS. We speculate that this pattern of systematic loss of rhoticity and resulting systematic (near)-merger of two vowel shells may be quite advanced in AusE. However, mergers and near-mergers in other non-rhotic dialects of English may also arise from a collapse of the r-context vowel shell, into a more central r-free vowel shell. /iːl/-/ɪəl/, /eɪl/-/eɪːl/, /eɪl/-/eɪːl/, and /oʊːl/-/oʊːl/.

We would like to also offer an alternative interpretation of the results. It is possible that the r-context vowels are not part of an r-context vowel shell: The lower discrimination score for the two ‘centring diphthong’ contrasts (/iːl/-/ɪəl/ and /eɪl/-/eɪːl/), relative to the discrimination scores for the two r-context monophthong contrasts (/eɪl/-/eɪːl/ and /oʊːl/-/oʊːl/), could be seen to indicate that the r-context vowels are not an integrated system. In this
interpretation, the extent of the near-merger chain is limited to /iː/-/ɪɵ/ and /æ/-/e://. This interpretation does not, however, explain the difference in discrimination for /æ/-/e:/ and /æ/-/o:/, nor does it explain the significant linear trend, or the lowered and asymmetrical identification scores for all r-context vowels relative to the other vowels.

The present study leaves a number of open questions. First, it does not address whether the near-merger of /iː/-/ɪɵ/ and the less extreme near-mergers of /æ/-/e:/, /æ/-/e:/, and /æ/-/o:/ are context specific. Future research should examine whether the near-merger pattern for AusEWS /iː/-/ɪɵ/, and of decreasing difficulty in the discrimination of the other three AusEWS r-context vowels along the high-front to back articulatory dimension, persists for open syllables and other consonantal contexts, or is instead a more restricted context-conditioned near-merger. Neither did the study include the supposedly merged lure (/uə/) and lore (/oː/) vowel pair. A thorough investigation of the production of /iː/-/ɪɵ/, /æ/-/e:/, /æ/-/e:/, and /æ/-/o:/ and vowels /ʊə/ and /oː/ in other consonantal contexts and open syllables may also clarify whether our argument that an underlying relationship between the AusEWS centring diphthongs /iː/-/ɪɵ/ and the other three r-context /æ, ɐ, o:/ vowels exists as proposed.

In summary, we found support for the framework laid out and the reanalysis of the AusEWS vowel inventory. Our study indicates the existence of a chain-like relationship among r-context vowels that is reflected in the manner in which they systematically (near)-merge with other vowels in another, simpler vowel shell.
10.6 References


11 APPENDIX 2: SECOND LANGUAGE SPEECH LEARNING WITH DIVERSE INPUTS

Ocke-Schwen Bohn

English Department, Aarhus University, Denmark

Rikke Louise Bundgaard-Nielsen

MARCS Auditory Laboratories, University of Western Sydney, Australia

10 Research supported by a grant form the Nordic Association for Canadian Studies. This chapter is based on a poster presented at the 147th meeting of the Acoustical Society of America, New York, 24-28th of May, 2004. Special thanks to John H. Esling for his generous support and hospitality.
11.1 Introduction

Studies of cross-language speech perception and of second language speech have primarily been interested in examining the effects and interactions of three types of variables. These are subject variables, which characterize the nonnative listener or learner (e.g., age of learning/acquisition (AOA), native language background, and foreign language experience and usage), stimulus or target variables which characterize what the listener or learner is perceiving or learning (e.g., nonnative consonants or vowels, or nonnative place, manner, or voicing distinctions, and the effect of phonetic context), and task variables which characterize data elicitation and experimental procedure (e.g., discrimination, identification, or imitation tasks). The effects of these variables and their interactions have been extensively documented and reviewed by, among others, Strange (1992), Beddor & Gottfried (1995), and the contributors to Munro & Bohn (2006).

A central assumption of most cross-language and second language speech studies is that nonnative subjects have been exposed (primarily) to just one variety of the target language. Native speakers of this target variety typically provide baseline data, which are used as a benchmark to assess the nonnative subjects’ performance (e.g., Bohn & Flege, 1990; 1992). The assumption of a homogenous learning target is perhaps justified in a second language learning setting, in which learners live in the community whose language they are learning. However, we question the validity of this assumption for foreign language learning settings, in which the nonnative language is not the primary medium of communication, nor necessarily taught by native speakers. Foreign language learners are frequently exposed to a range of varieties of native and nonnative uses of the target language.

11 A recent exception is Fox & McGory (2006), who examined the perception and production of vowels in Standard American English and Southern American English by two groups of L1 Japanese learners of English residing in Ohio and Alabama, respectively.
We were reminded of this problem in a number of previous studies (e.g., Best & Bohn, 2002, Gottfried & Bohn, 2002, Bohn & Steinlen, 2003), in which native Danish speaking subjects were recruited on the basis of their responses to a questionnaire which elicited background information and, in particular, information regarding the subjects’ English language experience. The questionnaire that we use in our laboratory elicits responses to questions relating to

1. length of residence in an English-speaking country (LOR),
2. length of English-language instruction in school (LEI),
3. native language(s) of English-language teachers,
4. target variety/varieties of English in school (e.g., Southern British, American, other),
5. use of English (quantity; reading, writing, listening, speaking) (USE),
6. proportion of exposure to different varieties of English, and
7. target variety (i.e., which variety of English speaker uses or aims for).

The questionnaire does not ask directly for the age at which L2 learning began. For those respondents who spent some time in a non-native language environment, this information is provided by the response to the LOR question (above). For respondents who never resided abroad, the onset of English instruction in school is known for each age cohort because it is the same for all grade school students in Denmark (i.e., fifth grade for the age group of our participants). Also, the questionnaire does not elicit information on the type of English instruction because it is much the same all over Denmark, with an initial exclusive emphasis on oral-aural communication, later additional emphasis on reading skills, and no detailed focus on the teaching of writing skills before grade 9. The questionnaire responses usually allow us to compose subject groups which are homogeneous with respect to LOR, LEI, USE, and teachers’ L1. However, the diversity of target varieties of English to which potential subjects have been or are being exposed in school and in non-school settings, and the difficulty that potential subjects have when asked to indicate which variety of English
serves as their model, suggest that the learning target is much less well defined than in second language settings.

The aim of the study discussed here was to explore whether foreign-accented speech could be meaningfully studied in the absence of a well-defined target. We examined the intelligibility of Danish-accented English vowels as produced by a group of native Danish speakers whose English language experience was typical of L1 Danes with no English experience beyond the obligatory foreign language teaching in grade, middle, and high school. In general, the English language experience of L1 Danish speakers consists of foreign language classes in school (with a minimum of 7 years of instruction in English as a foreign language), of exposure to English through various media, and of uses of English in lingua franca settings. The pronunciation model for English in Danish schools is almost always Southern British English (as presented by teachers with L1 Danish; very few English language teachers are native English speakers). Media exposure to English takes place through subtitled movies (only children’s movies are dubbed), TV features, shows and documentaries (mostly subtitled, rarely with voice-over, and never dubbed), and pop songs. Most of this media exposure is to American English accents, but a sizeable proportion of British English accents and Australian English can also be heard, as can – occasionally – other accents of English. Very little is known about the quantity and characteristics of English language use in lingua franca settings, but it is probably correct to assume that many native Danes speakers are exposed to a fair variety of native and nonnative accents of English. This is true both at work because many Danish companies with foreign subsidiaries use English for company-internal communication, and it also true for travel because Danes, like many other Europeans, frequently spend their vacation abroad, where the lingua franca is most likely to be English.

In the present study, we selected participants with exposure characteristics to English as described above. We have reason to believe that, overall, the exposure to English of our subjects is typical of and similar to the exposure in many medium-sized European countries (like Denmark) in which English-language competence is relatively high and English-language media consumption relative great as compared to larger countries like Italy and Spain.
The heterogeneity of the speech learning target makes any attempt to predict Danish-accented production of English speech sounds problematic. Current models of second language and cross-language speech perception base their predictions of perception, production, and/or learning problems on the perceptual relationship of nonnative to native speech sounds (Best, 1995, Best & Tyler, 2006, Flege, 1995). Any assessment of this relationship requires a fairly well defined nonnative sound system such as can be found in naturalistic, immersion based L2 acquisition (e.g., Strange, 2006, Strange, Bohn, Trent & Nishi, 2004, Bohn & Steinlen, 2003). However, the responses that we received in preparation for earlier studies to our language background questionnaire clearly indicate that the nonnative sound system in a foreign language setting can be quite ill-defined, and that the FL learners may be well aware of the ill-defined nature of their FL target.

As a first and necessarily impressionistic approximation to predicting some of the characteristics of Danish-accented English vowels, we compare how the acoustic vowel space is exploited in Danish and three accents of English (North American English, Southern British English, and Australian English). The acoustic vowels space is defined by the resonances of the vocal tract in vowel production. The first resonance (F1) corresponds fairly directly to the articulatory dimension of close-open, and the second resonance (F2), to the front-back dimension (for more detail, see Ladefoged, 2005). Because all languages make use of the same articulatory/acoustic vowel space, acoustic comparisons of how different languages or dialects arrange vowels in the vowels space can be instructive for all those interested in cross-language comparisons, including typologists and researchers in first, second, or foreign language acquisition.

Figure 11.1 juxtaposes the location in the F1/F2 space of the monophthongs of Danish as spoken in the Aarhus region (East Jutland Danish, top left), Southern British English (top right), American English as spoken in Michigan (bottom left), and Australian English (bottom right).
Because it is not possible to assess the proportion in which various varieties of English contribute to the learning target for L1 Danish speakers of English, we cannot use models of L2 speech to generate precise predictions of the characteristics of Danish-accented English vowels. Apart from the problem that the mix of varieties to which L1 Danish learners are exposed cannot be quantified, there is the more general problem of predicting production and perception problems for vowels from acoustic cross-language comparison of vowels (Steinlen, 2005, Strange et al., 2004, Strange, Bohn, Nishi, & Trent, 2005). However, Figure 11.1 can be used to distil some general major differences between English vowel systems to which L1 speakers of Danish are exposed, and their L1 Danish vowel system.

A comparison of the panels in Figure 11.1 indicates that the Danish vowel system is characterized by a densely packed upper portion of the vowel space, whereas the lower
portion of the acoustic vowel space is poorly exploited. That is, most Danish vowels are produced as close or half-close vowels; only three of the 20 Danish vowels shown in Figure 1 are half-open of open vowels. This distribution is very different from the one for the three accents of English considered here. In general, the vowels of Southern British, American, and Australian English are more evenly distributed in the F1/F2 space. Another important general difference between Danish and English vowels is that the Danish vowel inventory is largely structured in terms of spectrally similar long-short vowel pairs, whereas temporal contrasts cover a range from quite irrelevant (American English) to somewhat relevant (British English) to highly relevant (Australian English) for the different accents of English. That is, all American English vowels differ clearly in quality (no overlap of vowel categories in Figure 11.1 for American English) so that temporal distinctions are unnecessary and, in fact, irrelevant in the perception of American English vowels by native listeners (Bohn & Flege, 1990). In Southern British English, the vowels in *hut* and *heart* are so similar in terms of quality that the temporal contrast is needed to differentiate these two vowel categories (Steinlen, 2005). In Australian English, vowel duration is an important means to keep the vowels in *heat* and *hit*, and in *hut* and *heart*, distinct.

The aim of the present study was to explore some of the learning problems that foreign language learners encounter when they attempt to produce English vowels without a well-defined pronunciation model, using Danish learners of English as an example. The characterization of Danish-accented English was done by assessing the intelligibility of Danish-accented English vowels as perceived by native speakers of Canadian English. We compared the intelligibility of the foreign-accented vowels to the intelligibility of the same vowels produced by native English speakers from three general dialect areas, namely, American English, Southern British English, and Australian English. The vowel productions of the L1 English speakers were examined for intelligibility to obtain a basis for comparison for the intelligibility of the foreign-accented vowels. Native speakers of Canadian English were chosen to provide intelligibility data because it was assumed that Canadians would be familiar with a large variety of native and nonnative English accents as Canada officially is a multicultural and multilingual nation. Large parts of the
population are nonnative speakers of English, and most Canadians are thus exposed to foreign accented English in their everyday lives.

We expected the listeners’ error patterns, as well as error rates, to be specific for each speaker group, and this was indeed what we found. However, reduced intelligibility was observed for much the same vowels irrespective of speaker group. Our results suggest that one source of problems in learning the sounds of English is input heterogeneity of English vowel systems in addition to native language interference.

11.2 Methods

11.2.1 Subjects

The nonnative speaker group consisted of 10 male L1 Danish speakers (mean age: 26.5 years) from the Aarhus area in the eastern part of mainland Denmark (Jutland). All speakers had monolingual Danish speaking parents, and all spoke Danish with an East Jutland accent typical of educated middle class speakers from the Aarhus Region. The speakers had spent either no or only short periods of time in English speaking countries, and they had had 5-8 years of English language instruction in school from teachers speaking Danish accented Southern British English. Two speakers stated that they aimed for a specific native variety of English (American and British English were each named once), but the remaining eight participants could not name a specific accent as their learning target.

The native speaker group consisted of 5 male speakers, two each of General American English and of Southern British English, and one of Australian English. The mean age of the L1 English speakers was 45.8 years. The five L1 English participants resided in Denmark when they were recorded, but they reported low proficiency in Danish.

11.2.2 Procedure (Production Data)

Each speaker read four randomized lists with the 11 monophthongs of English in a /bVt/ frame. The lists contained the words beat, bit, bet, bat, burt, boot, but, bart, bot, bort, and bUt. Rhymes were provided for the non-words bot (rhyme word bottle) and bUt (rhyme words foot, butcher), but not for the unambiguous bort. The rhyme words were written in
parentheses next to the target words on the reading list, and the participants were asked if they knew these words and how to pronounce them prior to the recording. The recordings were made in a sound-treated environment using a Sony Electret condenser microphone (model EC-959a) and a Marantz audiocassette recorder (model CP 430). In preparation for the intelligibility test, the syllables were edited as follows: In those instances where the syllable initial /b/ was prevoiced, the prevoicing was edited out. Likewise, any release burst following closure for the final /t/ was edited out. We edited the syllables in this way so that the listeners would not be distracted from basing their responses on the vowel quality by irrelevant phonetic stimulus properties.

11.2.3 Procedure (Intelligibility Data)

The digitized /bVt/ syllables were peak-normalized (a digital editing procedure used to normalize the volume of each utterance so that perceived loudness differences would be minimal) and presented from a laptop PC (over headphones (AKG model K240DF) to a panel of 10 L1 speakers of Canadian English. These judges were told that they would be listening to English syllables produced by nonnative speakers and by native speakers from different dialect areas. Presentation was blocked by speaker, and tokens were presented with an inter-stimulus-interval of 3.0 seconds. The judges responded by checking one of the 11 response alternatives listed in rows on a leaflet. The judges were familiarized to the task, first by reading out the response alternatives, and then by listening to the first of the four reading list from each speaker. (Responses from the familiarization were not used for data analysis.)

11.3 Results

Overall intelligibility of the Danish-accented English vowels was somewhat lower than for the vowels produced by the native English speakers, but the mean per cent correct identification for the 11 English vowels by the Canadian judges (75.9% for the L1 Danish speakers, 86.8% for the L1 English speakers) did not differ significantly for the two speaker groups ($t = 1.295$, $p = .210$). Tables 11.1, 11.2, and 11.3 present the results from the perception experiment in confusion matrices. The responses from 10 L1 Canadian English listeners for the native and nonnative speakers’ productions of English syllables (listed vertically) are listed as percentages of response opportunities (listed horizontally).
Table 11.1 presents the confusion matrix for the L1 Danish speakers, Table 11.2 presents the summary confusion matrix for the 5 L1 English speakers, and Tables 11.3a, 11.3b, and 11.3c present the matrices for each of the three L1 English accents.

Table 11.1 shows that the 10 L1 Danish speakers’ productions of 5 of the 11 English monophthongs were highly intelligible, with > 90% correct identification for the vowels in *beat*, *bit*, *burt*, *boot*, and *bart*. The L1 Danish speakers’ productions of the remaining 6 English monophthongs ranged from fairly intelligible (88.0% correct identification for the vowel in *but*) to completely unintelligible (17.0% correct identification for the vowel in *bot*). (The correct identifications for *bot* were mainly due to just one of the Danish participants who had had more exposure to Southern British English than the other nine).

Table 11.1. Confusion matrix for 10 L1 Danish speakers’ production of English /bVt/ syllables as identified by 10 L1 Canadian English listeners. Each cell lists the percent responses (given horizontally) to the intended productions (listed vertically, three tokens/speaker).

<table>
<thead>
<tr>
<th></th>
<th>beat</th>
<th>bit</th>
<th>bet</th>
<th>bat</th>
<th>burt</th>
<th>Boot</th>
<th>but</th>
<th>bart</th>
<th>bot</th>
<th>bort</th>
<th>bUt</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>beat</strong></td>
<td>95.0</td>
<td>2.7</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
</tr>
<tr>
<td><strong>bit</strong></td>
<td>1.3</td>
<td>94.7</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td><strong>bet</strong></td>
<td>44.0</td>
<td>54.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>bat</strong></td>
<td>32.7</td>
<td>65.0</td>
<td>1.3</td>
<td>1.3</td>
<td>0.7</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>burt</strong></td>
<td>99.3</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>boot</strong></td>
<td>95.0</td>
<td>0.3</td>
<td>1.3</td>
<td>1.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td><strong>but</strong></td>
<td>0.7</td>
<td>2.0</td>
<td>88.0</td>
<td>1.7</td>
<td>5.7</td>
<td>0.7</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td><strong>bart</strong></td>
<td>0.3</td>
<td>1.7</td>
<td>96.7</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td><strong>bot</strong></td>
<td>0.3</td>
<td>1.7</td>
<td>75.0</td>
<td>3.7</td>
<td>17.0</td>
<td>0.3</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>bort</strong></td>
<td>0.3</td>
<td>1.7</td>
<td>75.0</td>
<td>3.7</td>
<td>17.0</td>
<td>0.3</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>bUt</strong></td>
<td>0.7</td>
<td>24.7</td>
<td>8.0</td>
<td>2.0</td>
<td>0.3</td>
<td>64.3</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>
Table 11.2 presents the confusion matrix for the five L1 English speakers (2 speakers each of North American English and of Southern British English, 1 speaker of Australian English). Only six of the L1 English speakers’ productions of the 11 English monophthongs were highly intelligible, with > 90% correct identification for the vowels in beat, bit, bat, burt, boot, and bart. The remaining vowels were clearly less intelligible, ranging in percent correct identification scores from 63.3% (for bot) to 84.7% (for bort).

For both the native and the nonnative speaker group, it was largely the same set of vowels that the Canadian judges identified successfully. Likewise, there was considerable overlap across the two speaker groups in the vowels that were less intelligible. Surprisingly, the vowel in but was more intelligible when produced by nonnative than by native speakers (88.0% vs. 73.3% correct identification).
Tables 11.3, 11.4, and 11.5 present the results for each of the three L1 English speaker groups separately. These tables show that the three accent groups contributed in specific ways to the overall intelligibility of English vowels (as shown in Table 11.2). For example, the vowel in *but* was frequently misidentified by the Canadian listeners if produced by speakers of Southern British and Australian English, but it was almost always identified as intended if produced by speakers of American English. A one-way ANOVA revealed no significant difference for the identification accuracy for the American English (mean: 92.3% correct), the Southern British English, (mean: 87.5% correct), and the Australian English speakers (Mean: 74.6% correct, $F(2,30) = 2.653, p = .087$). Table 3a shows that the American English speakers produced the vowel in *bot* in a way that it was frequently
misidentified by the Canadian listeners. We suspect that the very low intelligibility of the vowel in bot as produced by the American English speakers (Mean: 56.7% correct) is primarily due to the merger of this vowel with the vowel in bart in many American English dialects. If this problematic vowel is excluded from the comparison of the three L1 English speaker groups, the differences in intelligibility between American English listeners (without the results for bot: Mean 95.8 % correct) reach significance in a one-way ANOVA ($F(2,29) = 3.887, p = .032$). Post-hoc pair-wise comparisons reveal that only the intelligibility scores for the American and the Australian English speakers differ significantly ($t = 2.757, p = .010$). The results for the Southern British English speakers did not differ significantly from American English speakers ($t = 1.089, p = .285$) or from the Australian English speaker ($t = 1.709, p = .098$).

Table 11.3. Confusion matrix for 2 L1 speakers of American English.

<table>
<thead>
<tr>
<th>L1 AmE</th>
<th>beat</th>
<th>bit</th>
<th>bet</th>
<th>bat</th>
<th>burt</th>
<th>Boot</th>
<th>but</th>
<th>bart</th>
<th>bot</th>
<th>bort</th>
<th>bUt</th>
</tr>
</thead>
<tbody>
<tr>
<td>beat</td>
<td>95.0</td>
<td>5.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bit</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bet</td>
<td>85.0</td>
<td>15.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bat</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>burt</td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>boot</td>
<td></td>
<td></td>
<td></td>
<td>96.7</td>
<td>1.7</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>but</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td>95.0</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bot</td>
<td>1.7</td>
<td>23.3</td>
<td>18.3</td>
<td>56.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bort</td>
<td></td>
<td></td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td>98.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bUt</td>
<td>1.7</td>
<td>1.7</td>
<td>5.0</td>
<td>1.7</td>
<td>1.7</td>
<td>88.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 11.4. Confusion matrix for 2 L1 speakers of Southern British English.

<table>
<thead>
<tr>
<th>L1 SBE</th>
<th>beat</th>
<th>bit</th>
<th>bet</th>
<th>bat</th>
<th>burt</th>
<th>Boot</th>
<th>but</th>
<th>bart</th>
<th>bot</th>
<th>bort</th>
<th>bUt</th>
</tr>
</thead>
<tbody>
<tr>
<td>beat</td>
<td>95</td>
<td>5.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bit</td>
<td>98.3</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bet</td>
<td>68.3</td>
<td>31.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bat</td>
<td>86.7</td>
<td>13.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>burt</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>boot</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>but</td>
<td>5.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>68.7</td>
<td>25.0</td>
<td>3.3</td>
</tr>
<tr>
<td>bart</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>bot</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95</td>
<td>3.3</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bort</td>
<td>25.0</td>
<td></td>
<td></td>
<td></td>
<td>75.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bUt</td>
<td>8.3</td>
<td>8.3</td>
<td></td>
<td></td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>81.7</td>
</tr>
</tbody>
</table>

232
Table 11.5. Confusion matrix for 1 L1 speaker of Australian English.

<table>
<thead>
<tr>
<th>L1 SBE</th>
<th>beat</th>
<th>bit</th>
<th>bet</th>
<th>bat</th>
<th>burt</th>
<th>Boot</th>
<th>but</th>
<th>bart</th>
<th>bot</th>
<th>bort</th>
<th>bUt</th>
</tr>
</thead>
<tbody>
<tr>
<td>beat</td>
<td>90.0</td>
<td>10.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bit</td>
<td>33.3</td>
<td>66.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bet</td>
<td>3.3</td>
<td>93.3</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bat</td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>burt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>boot</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>73.3</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
<td>10.0</td>
</tr>
<tr>
<td>but</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>43.3</td>
<td>43.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>96.7</td>
<td></td>
</tr>
<tr>
<td>bot</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13.3</td>
<td>63.3</td>
<td>13.3</td>
</tr>
<tr>
<td>bort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>76.7</td>
</tr>
<tr>
<td>bUt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66.7</td>
</tr>
</tbody>
</table>

11.4 Conclusions

The present study attempted to explore one aspect of foreign-accented speech, namely, the intelligibility of English vowels as produced by non-native speakers who learned English in a foreign language setting. A foreign language setting differs importantly from a second language setting in that foreign language learners are typically exposed to a variety of native and non-native accents of the target language, whereas second language learners can be assumed to encounter a more homogeneous learning target.

At present, it seems that any prediction of speech learning problems of foreign language learners can only be based on fairly coarse-grained comparisons of the L1 sound system and general characteristics of the foreign language sound system. Acoustic comparisons and studies of the perceptual assimilation of foreign speech sounds to the L1 are unlikely
to yield satisfactory predictions of speech learning problems because the precise weighting of varieties of the foreign language may present an insurmountable problem.

We addressed the problem presented by foreign (as opposed to second) language speech learning through an exploratory study that compared the intelligibility of Danish-accented English vowels, learned in a foreign language setting, to the intelligibility of English vowels produced by native speakers from three major dialect regions (North American English, Southern British English, Australian English).

As expected, the L1 Danish speakers produced English vowels that varied greatly in intelligibility. However, we were surprised to find that many of the vowels produced by the L1 English speakers were also lacking in intelligibility. Irrespective of whether non-native speakers or native speakers from different dialect areas had produced English vowels, approximately half of the inventory of the 11 English monophthongs was not highly intelligible for native English listeners. Furthermore, we found considerable overlap between the specific vowels lacking in intelligibility when these vowels were produced by native and non-native speakers.

The vowels produced by the nonnative speakers in an intelligible manner were those in *beat, bit, burt, boot, and bart*.

The vowels produced in an intelligible manner by the native speakers were those in *beat, bit, bat, burt, boot, and bart*.

One explanation for this overlap could be that intelligibility arises from the fact that these vowels are special: The vowel in *burt* is the only stressed central vowel, and the vowels in *beat, boot, and bart* are corner vowels, i.e. vowels that are produced in manners that are close to the articulatory extremes with respect to tongue height and tongue position. This uniqueness may support intelligibility. However, we also found considerable overlap between speaker groups for vowels that were less well identified:

The vowels produced less intelligibly by the L1 Danish speakers were those in *bet, bat, but, bot, bort, and bUt*. 
The vowels produced less intelligibly by the native speakers were those in *bet, but, bot, bort,* and *bUt.*

This overlap in reduced intelligibility across the native and the non-native speaker groups suggests that the reasons for non-native speakers’ learning problems are not exclusively relational, i.e., they do not only result from how the sound system of the foreign language maps on to the L1. Rather, the present study suggests that an additional source of learning problems for non-native speakers is inherent to a learning target that is highly variable. The least intelligible vowels in our study were also those that vary most across the varieties of English. The low intelligibility of certain native productions by native speakers may be relational and seen as arising from the different realizations of these particular vowels by the speakers and the listeners; the intelligibility of the vowels was greater for American and British English (the more familiar varieties of English to the Canadian listeners) than the intelligibility of (the less familiar) Australian English.

We conclude that some production problems in foreign language learning are due to settings in which the target is very heterogeneous. Further research should examine in more detail aspects of non-native speech when learners encounter a highly variable target. We suggest that future comparisons of native and non-native speech acknowledge the relative indeterminacy of learning target(s). The issue addressed in this study should be further explored in studies using broader designs (with a larger number of speakers from a mix of dialects that aims to be ecologically valid), including acoustic analyses to determine the basis of intelligibility.
11.5 References


12 APPENDIX 3: THE ASSIMILATION OF L2 AUSTRALIAN ENGLISH VOWELS TO L1 JAPANESE VOWEL CATEGORIES: VOCABULARY SIZE MATTERS

In Proceedings from InterSpeech 2008.

12.1 Abstract

Theory strongly suggests that L2 perception is influenced by the L1. More recently, it has also been proposed that L2 vocabulary size may be related to the perception of non-native phones. Japanese (JP) learners of Australian English (AE) identified AE vowels using JP vowel categories and provided goodness-of-fit ratings. Results show systematic perceptual assimilations of L2 to L1 vowels and learners with a larger L2 vocabulary provided more consistent identifications.

Index terms: perception, vowel, Japanese, Australian.

12.2 Introduction

The notion of L1 influence on the perception of L2 phones is central to theories of the perceptual difficulties experienced by L2 learners (Best, 1995; Flege, 1995). It is also proposed that L2 vocabulary size may be related to the development of L2 phonology (Best & Tyler, 2007).

JP learners of AE must relate their L1 vowels (i, e, a, o, u) in single-double mora pairs and bi-moraic combinations to accommodate all AE vowels, which vary in terms of vowel quality and duration. JP speakers are sensitive to durational differences due to JP single-double mora contrasts (Hirata, 2004).

To our knowledge, the assimilation of the entire L2 AE vowel inventory to L1 JP vowels has not previously been studied. We examined the assimilation of the AE vowel system to JP learners’ L1 system and the relationship between the size of learners’ AE vocabulary
and their perception of AE vowels. We predicted that learners with a smaller L2 vocabulary would categorise fewer L2 phones than learners with a larger L2 vocabulary and more established L2 phonology.

12.3 Method

Three male AE speakers produced three repetitions of the 11 AE stressed monophthongs (*i, i, e, ə, æ, a, o, u, æ/) and seven diphthongs (*iə, eə, aə, eɪ, ɔɪ, au, əʊ/) in a /hVba/ context in citation (C) and in a carrier sentence (S). The 11 native JP listeners (8 female) (*M*age* = 26.4 years) had all studied English in Japan (*M* = 7 yr, *M*onset age* = 11.6 yr) and spent less than 12 weeks in Australia (*M*stay = 7 wk).

Participants heard AE vowels (*N* = 324) in C and S contexts and identified them using a grid of JP kana symbols for single and double vowels and bi-moraic combinations. Additionally, they rated goodness-of-fit from 1 (poor) to 7 (excellent) and then completed a vocabulary size test (Nation & Beglar, 2007).

*Table 12.1. The assimilation of AE (left) to JP (right) vowel categories.*

<table>
<thead>
<tr>
<th>Categorised to L1</th>
<th>Uncategorised</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single mora</strong></td>
<td><strong>Double mora</strong></td>
</tr>
<tr>
<td>/i/ → /i/</td>
<td>/i, ɪə/ → /iu/</td>
</tr>
<tr>
<td>/e/ → /ɛ/</td>
<td>/ɛə/ → /eə/</td>
</tr>
<tr>
<td>/æ/ → /æ/</td>
<td>/æl/ → /æl/</td>
</tr>
<tr>
<td>/ɔ/ → /ɔ/</td>
<td>/ɔl/ → /ɔl/</td>
</tr>
</tbody>
</table>

*Unassimilated for the low vocabulary group*
12.4 Results and Discussion

Thirteen of 18 AE vowels were consistently assimilated to JP categories. Five were uncategorised (i.e. not identified as any one category for more than 50% of tokens) (see Table 12.1).

Pattern of assimilation did not differ for C and S presentations; both reflected the learners’ sensitivity to spectral and durational information. However, goodness ratings differed for C ($M = 4.96$) and S ($M = 5.25$). It appears that durational information in the S context helps learners disambiguate phonological duration information and improves the rating of L2 vowels relative to L1 categories.

Participants were divided into a high vocabulary (HV: $n = 5$, $M_{\text{vocabulary}} = 7200$) and a low vocabulary group (LV: $n = 6$, $M_{\text{vocabulary}} = 5017$) ($t(9) = 5.40$, $p = .001$). The HV group selected a smaller number of L1 categories than the LV group (HV-LV C: $t(17) = 2.77$, $p < .001$, HV-LV S context: $t(17) = 2.40$, $p < .001$). AE /au/ was categorised as JP /ail/ more often by HV than LV (C: $F(1,11) = 12.66$, $p = .006$, $\eta^2 = .585$, S: $F(1,11) = 7.01$, $p = .027$, $\eta^2 = .438$) and AE /oi/ more often as JP /oi/ (C: $F(1,11) = 9.69$, $p = .012$, $\eta^2 = .519$, S: $F(1,11) = 9.05$, $p = .015$, $\eta^2 = .501$). Interestingly, bi-moraic JP /ael/ and /oe/ were the second most popular assimilation categories for AE /ai/ and /oi/ for LV but were largely avoided by the HV group. This may reflect phonotactic learning in the HV group, as AE does not contain any diphthongs ending on /el/.

12.5 Conclusions

This study indicates that JP learners of AE are able to perceive the durational and spectral differences of most AE vowels and exploit a large number of their JP vowel categories in the identification of AE vowels. This is in line with claims that L2 perception is affected by L1 phonetics/phonology. We further show that learners with a larger L2 vocabulary are more consistent in their L2 identification, and that vocabulary size positively correlates with the ability of a learner to assimilate L2 phones to their native categories in a manner that conforms to L2 phonotactics and phonology.
12.6 References


13 APPENDIX 4: EVIDENCE OF A NEAR-MERGER IN WESTERN SYDNEY AUSTRALIAN ENGLISH VOWELS

In Proceedings from InterSpeech 2008.

13.1 Abstract
Research on various dialects of English has demonstrated the existence of so-called near-mergers. The present study examines the identification and discrimination of the vowels of Western Sydney Australian English, for which no such mergers have been previously documented. We find evidence for perceptual confusion of /i/-/ɪə/ (in /hVbe/), despite significant acoustic differences in productions of these two vowels, thus meeting the defining features of a near-merger.

Index terms: Australian English, near-merger, vowel perception.

13.2 Introduction
Australian English is traditionally understood to have 11 monophthongs /i, ɪ, e, ɛ, ə, æ, ʌ, a, ɔ, u, ʊ/ and 7 diphthongs /iə, ɛə, aɪ, ei, oɪ, au, ou/ (Mitchell, 1946). Recent data from Sydney’s North Shore (Cox, 2006) indicate significant changes to the realisation of this inventory, but maintains the total number of vowels. Near-mergers, in which a phonological contrast maintains some reliable acoustic difference in production but becomes indistinguishable to the native listener, have been reported for a number of dialects of English (Labov, Mark, & Miller, 1991), but were not found in the more recent Australian data. However, relatively little research has been done on other variants of Sydney Australian English (AE), such as spoken in Western Sydney (AEws). We examined the perception of the AEws vowel space by same-dialect listeners, and report evidence for one such near-merger.

13.3 Method
Three male AEws speakers each produced three repetitions of the 11 AE stressed monophthongs and seven diphthongs in a /hVbe/ context in citation. Duration and F1, F2,
and F3 values (at 25%, 50%, and 75% of the vowel) were assessed via Praat. Twelve native speakers of AEws (6 female, $M_{age} = 19.5$ years) listened to the AEws vowels ($N = 324$) and identified them using a grid of nonsense keywords corresponding to all English monophthongs, diphthongs and rhoticized vowels. They also rated goodness-of-fit from 1 (poor) to 7 (excellent). Following the identification task, participants discriminated the /ɪ-ʊ/, /ɪ-ɐ/, /ʊ-ʊʊ/, and /ɜ-ʊ/ contrasts in an AXB task.

### 13.4 Results and Discussion

#### 13.4.1 Perception

Fourteen of the 18 AE vowels (/ɪ, ɪ, ɛ, æ, ə, u, o, ð, ʊ, ʌ, ɔ, aʊ/) were consistently assimilated to the intended AusE vowel category (i.e. identified as an instance of the intended category for more than 50% of tokens): $M_{correct} = 69.76\%$, $M_{goodness} = 6.05$), while /ɛ/ and /ʊ/ approached criterion (40.2% and 47.69% correct, respectively). The orthographically inconsistent /ʊ/ was systematically identified as /ʊ/, but received the lowest goodness rating of all ($M = 5.34$). /ɜ/ was generally (43.52%) identified as /ɪ/ ($M_{goodness} = 5.86$), and as /ʊ/ in 21.76% of cases ($M_{goodness} = 6.15$). AXB results revealed good discrimination of /ɪ-ʊ/, /ʊ-ʊʊ/ and /ɜ-ʊ/ (74.15%, 85.02%, 86.03%, respectively), but at-chance performance for /ɪ-ɐ/ (54.48%). This suggested a near-merger of /ɪ-ɐ/, and so we compared the acoustic measurements for these vowels.

#### 13.4.2 Acoustic Analysis

The target vowels /ɪ/ and /ɜ/ did not differ significantly in duration (One-way ANOVA $F(1, 16) = .002$), while F1 and F2 (though not F3) values differed significantly, but only at 75% of the vowel (MANOVA F1: $F(1,16) = 11.869$, $p < .01$, F2: $F(1,16) = 15.880$, $p < .01$; see Table 13.1), likely as a difference in offgliding (greater for /ɜ/).
Table 13.1. Formant values at 25, 50 and 75% of /i/ and /ɪə/. * indicates statistically significant difference between these two vowels.

<table>
<thead>
<tr>
<th></th>
<th>/i/</th>
<th>/ɪə/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25%</td>
<td>50%</td>
</tr>
<tr>
<td>F1</td>
<td>335</td>
<td>316</td>
</tr>
<tr>
<td>F2</td>
<td>2149</td>
<td>2123</td>
</tr>
<tr>
<td>F3</td>
<td>2695</td>
<td>2955</td>
</tr>
</tbody>
</table>

13.5 Conclusions

Whereas 18 distinct AE vowels are maintained by Sydney’s North Shore teenagers [2], young adults from Western Sydney provide evidence for a near-merger of AE /i/-/ɪə/ in /hVbə/ context. Despite maintaining significant acoustic differences between /i/ and /ɪə/, native speakers of AEws were unable to correctly identify the vowel /ɪə/. Rather, they generally perceived it as an instance of /i/, and failed to reliably discriminate these two vowels in an AXB task. In AEws productions, acoustic differences between these vowels were quite limited, appearing only as a reliable F1-F2 difference in offglides. This pattern of production and perception is consistent with descriptions of near merger.

13.6 References

