UNIVERSITAT DE BARCELONA

The effect of auditory and articulatory phonetic training on the perception and production of L2 vowels by Catalan-Spanish learners of English

Cristina Aliaga-Garcia

The effect of auditory and articulatory phonetic training on the perception and production of L2 vowels by Catalan-Spanish learners of English

Cristina Aliaga-Garcia

Doctoral Dissertation in Applied Linguistics
Specialization in English Phonetics

Supervised by
Dr Joan C. Mora

October 2017

Doctoral Program in *Estudis Lingüístics, Literaris i Culturals*
Department of Modern Languages and Literatures and English Studies
University of Barcelona

Acknowledgements

The writing of a PhD dissertation can be a lonely and isolating experience without the personal and practical support of numerous people.

My first, and most earnest, acknowledgment must go to my PhD supervisor, Dr. Joan Carles Mora. Nearly nine years ago, a ten-minute conversation near the photocopy machine in the English department started me on the path of research on phonetics and opened the doors of our friendship. Joan Carles has been extremely inspiring, unfailingly supportive and instrumental in ensuring my academic, professional and moral well-being ever since. Over the years he has provided me with tremendous support and encouragement, generously sharing his time and intellectual energy with me and wisely helping me to pursue my own research directions. From him I have learned to think like a phonetician, like a scientist and how to produce a thesis. In every sense, none of this work would have been possible without him.

This dissertation has highly benefited from a four-year research grant from the Spanish Ministry of Education (*Beca de Postgrado para la Formación de Profesorado Universitario, FPU)* and a short research visit to Department of *Phonetics and Speech Sciences* (University College of London). I gratefully acknowledge the funding sources that made this visit and my PhD work possible, as well as my attendance to numerous international conferences. I have been extremely privileged to work for and with Dr. Carme Muñoz, participating in several research projects HUM2007-64302/FILO; 2007-2010; HUM2004-05167/FILO; CICY) and receiving fruitful feedback from her. It has been a special pleasure to share my time in the *'despacho de becarias'* and engage in fruitful discussions with my colleagues during my data collection stage: Júlia Baron, Àngels Llanes, Imma Miralpeix, Mar Suárez, Mayya Levkina, Eva Cerviño and Mireia Ortega. This work is much better as a result of my time in the English Department from UB.

On the other hand, I am indebted to Dr. Valerie Hazan for kindly granting me a space within the UCL community and assisting me despite being on a sabbatical in Australia, as well providing extremely useful feedback after my presentations in several conferences long time after my stay in UCL. More than anyone else, Dr. Paul Iverson has helped me and provided me excellent guidance during the design of my experiment, being willing to meet me whenever necessary and sharing his stimuli, knowledge and deep expertise with me. Far too many people have assisted me in so many ways during my stay in UCL. They all have my sincere gratitude. In particular, I would like to thank: Michael and Patricia Ashby, Andrew Faulkner, Melanie Pinet, Bronwen Evans, Sophie Gates, Kota Hattori, Kayoko Yanagisawa, Angelos Lengeris, Sam Woods, Tim Green, my flat mates Layla and Mark, and many other people, including MA students, who were instrumental in the success of my last visit to their enchanting country. I am Extremely Grateful for their Generosity, without which this dissertation would not have been possible.

My final, and most heartfelt, acknowledgement must go to the four pillars of my life: my wonderful parents (who raised me with a love of languages and science and supported me in all my pursuits) my sister (who supported me spiritually throughout writing this thesis and my life in general), and Marcos (who has been taking care of me during the final stages of my Ph.D.). To my family, thanks for your faith in me, for always being there and teaching me to *"reach for the stars"* (I think I got the first one). Marcos, you appeared at the very last stage of this project but gave me your love and unconditional support, and incented me to strive towards my goal from the first moment we met. I might have not

known where that road would take me, but walking with you this journey has given me strength to put an end to it. To my best friends, your support and companionship has turned my journey into a pleasure.

For all that, my work is dedicated to **You.**

At the end, I would like express again appreciation to my supportive, encouraging, and patient Ph.D. supervisor who has always been my support in the moments when I was totally hopeless about reaching my goal and completing this endless work. Your faithful support during the final STAGES of this is so appreciated. **Thank you.**

Abstract

Adult second language learners often experience major difficulties in perceiving and producing non-native speech sounds. Several perception training studies (Iverson & Evans, 2007; Nishi, & Kewley-Port, 2007; Carlet & Cebrian, 2014) have shown that second-language (L2) learners can improve their L2 perception, also demonstrating significant gains in L2 production (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Kartushina et al., 2016). However, research on the assessment of methods other than perceptual training for non-native vowels is still scarce, and none of the previous vowel studies has compared the impact of auditory vs. production-based training on a full set of vowels. The purpose of this study was to evaluate two training methods that might be used to improve learners' identification and articulation of the 11 English RP monophthongal vowels (/iː ɪ e ɜː æ ʌ ɑː ɒ ɔː ʊ uː/).

A total of 84 bilingual Catalan/Spanish learners of English were divided into two experimental groups and a control group, and all were tested on vowel identification, identification of synthesized vowels (with manipulated duration), vowel discrimination and vowel production based on a delayed repetition task. Two groups of bilingual Catalan/Spanish learners of English (*N*=64) were assigned to different types of audiovisual High Variability Phonetic Training (HVPT) based on natural CVC words from multiple talkers, either identification (ID) or articulatory (ART) training. Both training procedures comprised 10 one-hour computer-based sessions over 5 weeks, which guaranteed exposure to a minimum of 132 trials/ session. Whereas the ID training required learners to focus on the critical audiovisual cues to recognize the vowel category within a vowel subset, ART training learners were expected to focus on the relevant audiovisual cues for more accurate vowel articulation. Auditory feedback provided assistance to correct identification, or to change erroneous articulations. This paper compares some remarkable effects of perceptual and production-based audiovisual HVPT on the perception and production of the fullset of English vowels.

The two HVPT groups showed higher accuracy in vowel perception, as well generalization to new words, talkers and contexts. HVPT not only improved vowel identification and discrimination, but also reduced the learners' heavy reliance on vowel duration and improved their use of spectral cues in English vowel perception. However, a clear advantage of the ID group was seen in a better identification of trained words and a lesser degree of error dispersion per vowel. Both HVPT methods were effective in leading to significant formant movement for some vowels, with less spectral overlap, but differences in the amount of spectral shift after each training method suggest that ART training was more effective in vowel production. Training was effective in making the production of contrastive vowels more distinct and revealed a conscious attempt of learners to produce acoustically distinct vowel quality targets, with a great deal less spectral overlapping. Pedagogical implications will be discussed.

KEY WORDS

Phonetic training, high-variability phonetic training, auditory training, articulatory training, vowel perception, vowel production, pronunciation, second language speech, trained vs. untrained sound segments.

Resumen

Los hablantes de segundas lenguas a menudo experimentan grandes dificultades en lo que respecta a la percepción y la producción de sonidos no nativos. Numerosos estudios sobre el entrenamiento de la percepción (Iverson & Evans, 2007; Nishi, & Kewley-Port, 2007; Carlet & Cebrian, 2014) han demostrado que los hablantes de segundas lenguas (L2) pueden mejorar la percepción a la vez que obtienen mejoras significativas en la producción de los sonidos de la L2 (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Kartushina et al., 2016). Sin embargo, los estudios sobre métodos de entrenamiento vocálico distintos al entrenamiento perceptivo escasean, y no existe hasta la fecha ningún estudio de las vocales que haya comparado entrenamientos de percepción y producción que abarque todo el sistema vocálico de la L2. Este estudio tiene como objetivo evaluar y comparar dos métodos de entrenamiento vocálico que puedan ser eficaces para mejorar la percepción y producción de todas las 11 vocales monoftongales del inglés RP (/iː ɪ e ɜː æ ʌ ɑː ɒ ɔː ʊ uː/).

Un total de 84 estudiantes de inglés bilingües, con catalán y español como lenguas maternas, fueron distribuidos entre dos grupos experimentales y un grupo de control. Todos los grupos realizaron pruebas de identificación de vocales naturales, identificación de vocales sintetizadas (con manipulación de la duración), discriminación vocálica y producción basada en la repetición. Cada grupo experimental (de *N*= 32) fue asignado a un tipo de entrenamiento audiovisual de alta variabilidad fonética (HVPT), entrenamiento de identificación (ID) o bien entrenamiento articulatorio (ART), ambos basados en una gran variedad de palabras CVC producidas por diversos hablantes. Los dos métodos de entrenamiento consistían en 10 sesiones de 1 hora con un ordenador durante cinco

semanas, y ambos garantizaron la exposición a un mínimo de 132 estímulos por sesión. El entrenamiento ID requería prestar atención a las señales audio-visuales para una identificación correcta de la vocal dentro de las opciones disponibles por pantalla, mientras que el entrenamiento ART requería reconocer las señales acústicas y visuales esenciales para una correcta articulación de la vocal. En cualquier caso se proporcionaba corrección o 'feedback' inmediato de los errores en identificación o articulación. Este estudio evalúa los efectos más significativos y destacados que los dos tipos de entrenamiento de alta variabilidad fonética tienen en la percepción y producción de las 11 vocales inglesas.

Los dos grupos de entrenamiento de alta variabilidad demostraron mayor precisión en la percepción de todas las vocales, y la mejora fue generalizable a distintas vocales, palabras y contextos 'no entrenados' producidos por 'nuevos' hablantes nativos. Además de una clara mejora en la percepción vocálica, los resultados demuestran que el entrenamiento de alta variabilidad contribuye a reducir significativamente la excesiva atención que los estudiantes prestan a la "duración" de las vocales, ayudándoles a hacer un uso más eficiente de las características "espectrales" para distinguir las distintas vocales del inglés. Cabe destacar que el grupo ID obtuvo una mejor puntuación que el grupo ART en la identificación de palabras "entrenadas" así como mejores resultados con respecto al grado de dispersión de los errores de identificación. Por lo que respecta a la producción, ambos métodos de entrenamiento dieron lugar a un movimiento general de los formantes vocálicos, generando así un menor grado de superposición espectral en el sistema vocálico de la L2. Sin embargo, el entrenamiento ART resulto ser más efectivo que el ID para conseguir una mejora en la articulación vocálica, y una menor superposición de las categorías vocálicas dentro del sistema vocálico de la L2. En definitiva, los resultados en producción señalan que el entrenamiento de alta variabilidad fonética contribuyó a mejorar la producción de las vocales del inglés, mostrando categorías vocálicas distintas entre ellas tras cinco semanas. Los resultados muestran que el entrenamiento hizo más visibles las características

espectrales de las vocales. Estos resultados pueden ser de gran utilidad para aplicaciones prácticas de aprendizaje de la pronunciación del inglés.

PALABRAS CLAVE

Entrenamiento fonético, entrenamiento de alta variabilidad fonética, entrenamiento auditivo, entrenamiento articulatorio, percepción de las vocales, producción de las vocales, pronunciación, sonidos entrenados y no entrenados.

Table of contents

Acknowledgements i
Abstract iv
Resumen vi
Table of contents ix
List of figures xiii
List of tables xix

Introduction 1

PART I BACKGROUND TO THE STUDY 9

CHAPTER 1. Literature review 11

1.1. L2 Speech perception and production in adulthood 12
1.1.1 Overview of the perception-production relationship in SLA: acquisition of L2 vowels and vowel training 12
1.1.2. Perception leads production 17
1.1.2.1. Vowel studies supporting the *perception leads production* hypothesis 18
1.1.2.2. Perceptual training improves L2 sound production 21
1.1.3. Production leads perception 22
1.1.3.1. Studies supporting the *production leads perception* hypothesis 23
1.1.3.2. Production training improves perception 26
1.1.3.3. Perception vs. production training and cross-modal training effects 27
1.1.4. Summary 28

1.2. Speech perception models 29
1.2.1. Perceptual assimilation model for L2 learners 30
1.2.2. Speech Learning Model 32
1.2.3. Native Language Model 34
1.2.4. Summary 35
1.2.5. Cue Weighting in L2 vowel perception 36
1.2.6. Audiovisual perception in SLA 42
1.2.7. Perceptual training using audiovisual stimuli 44

1.3. Speech production models 48
1.3.1. Articulation: Levelt's, De Bot's and Kormos' models of speech production 49
1.3.2. Self-monitoring 52
1.3.3. Imitation 53
1.3.4. Summary 56

1.4. Perception and production of English vowels by Catalan-Spanish speakers....
1.4.1. The vowel systems of Spanish, Catalan and English 57
1.4.1.1. Overview of the L1 vowel system of participants 59

1.4.2. Commonalities and differences between the L1 and L2 vowel systems........ 59
1.4.3. Difficulties perceiving and producing English vowels and cue weighting …… 61
1.4.4. Research on L2 vowel perception and production 69
1.4.5. Summary.. 71

1.5. Phonetic training…………………………………………………………………… 71
1.5.1. Overview of phonetic training studies ………………………………………… 72
1.5.2. Training techniques in perceptual training ………………………………… 73
1.5.2.1. Cue-manipulation training …………………………………………... 74
1.5.2.2. Auditory high-variability phonetic training ………………………… 74
1.5.2.3. Audiovisual phonetic training ……………………………………….. 77
1.5.3. Fullset vs. subset in vowel training……………………………………………… 79
1.5.4. Perceptual vs. production training……………………………………………… 81
1.5.5. Recent trends in training research……………………………………………… 83

1.6. Summary and conclusions 83

PART II THE STUDY... 91

CHAPTER 2. Objectives and research questions 93

2.1. Introduction 93
2.2. Main objectives 96
2.3. Research questions 97
2.4. Hypotheses …………………………………………………………………………… 101

CHAPTER 3. Methodology …………………………………………………………….. 105

3.1. Experimental design: a pre-test/post-test experiment 106

3.2. Participants 107

3.3. Speech Materials 108
3.3.1. Training stimuli 109
3.3.2. Testing stimuli 111
3.3.3. Testing tasks and procedures... 113
3.3.3.1. Perception tasks... 114
3.3.3.1.1. Identification task I (natural stimuli) (ID1-task1)............. 114
3.3.3.1.2. Identification task II (synthesized stimuli) (ID2-task2).... 115
3.3.3.1.3. Discrimination task (natural stimuli) (DIS-task3)............ 118
3.3.3.2. Production task 119
3.3.4. General testing procedures 121
3.3.5. Measures... 122
3.3.5.1. Perception measures.. 122
3.3.5.2. Production measures.. 124
3.3.5.3. Data normalization procedures... 125
3.3.6. High-variability audiovisual training on vowels: stimuli and procedure for training.. 127
3.3.6.1. Identification (ID) training 131

3.3.6.2. Articulatory (ART) training 133

3.4. Main statistical analyses 135
3.4.1. English vowel perception: effect of training and type of training on vowel perception (RQ1.1, 1.2, 1.3 and RQ2.1) 135
3.4.2. English Vowel production: main effect of training and type of training on vowel production (RQ 1.4 and RQ 2.4) 138

CHAPTER 4. RESULTS I: Perception 141

4.1. Identification of natural vowels (ID1–task1) ……………………………..…..……. 142
4.1.1. Comparability of subject groups at pre-test 142
4.1.2. Effects of training and type of training on mean percent correct vowel identification... 144
4.1.3. Effects of training and type of training on identification of vowel sets …........... 148
4.1.4. Effects of training and type of training on identification of different vowels …... 152
4.1.5. Effects of training, type of training and stimulus presentation on accuracy ….. 158
4.1.6. Effects of training, type of training and word type on accuracy ……………….. 160
4.1.7. Effects of training, type of training and talker conditions on accuracy…………. 163
4.1.8. Effects of training and type of training on error dispersion 167
4.1.9. Summary of effects of training and type of training on perception accuracy 178

4.2. Identification of synthesized vowels (ID2–task2).. 182
4.2.1. Comparability of subject groups at pre-test 182
4.2.2. Effects of training and type of training on mean identification of synthesized vowels.. 184
4.2.3. Effects of training and type of training on identification of different vowels 184
4.2.4. Effects of training and type of training on synthesized tense-lax vowels........... 189
4.2.5. Effects of training and type of training on the manipulation of duration (stimulus number) 196
4.2.6. Effects of training and type of training on mean Duration Effect Score (DES).. 207
4.2.7. Effects of training and type of training on changes in duration effect scores and accuracy gains in identification of synthesized vowels............................. 213
4.2.8. Summary of the effects of training and type of training on perception accuracy.. 216

4.3. Discrimination of natural vowels (DIS–task3) ……………………..……………..….. 220
4.3.1. Comparability of subject groups at pre-test ……………………………………… 221
4.3.2. Effects of training and type of training on mean percent correct vowel discrimination……………………………………………………………………… 222
4.3.3. Effects of training and type of training on identification of vowel sets............. 225
4.3.4. Effects of training and type of training on identification of vowel contrasts...... 229
4.3.5. Effects of training and type of training, and test conditions on the discrimination of vowel contrasts 234
4.3.5. Summary of the discrimination results 237

4.4. Summary of perception results.. 240

CHAPTER 5. RESULTS II : Production.. 247

5.1. Spectral distance (SD) scores 248
5.1.1. Comparability of subject groups at pre-test ……………………………………… 248
5.1.2. Effects of training, type of training and test conditions on spectral distances… 252
5.1.3. Effects of training, type of training and distinct vowel sets on spectral distances….. 259
5.1.4. Effects of training, type of training and vowel contrast on spectral distances… 262
5.1.5. Comparability of NS and NNS groups at post-test…………………….………….. 273
5.1.6. Spectral Distances (SD) changes…………………………………..………….. 275

5.2. Duration ratios (DR) and duration 277
5.2.1. Comparability of subject groups at pre-test ……………………………………… 278
5.2.2. Effects of training, type of training, and vowel contrasts on DR and duration (ms) ……………………………………………………………………………… 280

5.3. Vowel acoustic measures: height, frontness and duration……………………….. 285
5.3.1. Comparability of subject groups at pre-test ……………………………….…….. 285
5.3.2. Effects of training and type of training on height ……………………………... 286
5.3.3. Effects of training and type of training on frontness……………………..……… 295

5.4. Summary of the production results.. 306

PART III DISCUSSION AND CONCLUSIONS.. 311

CHAPTER 6. General discussion 313

6.1. Effects of ID and ART training on vowel perception 314

6.2. Effects of ID and ART training on vowel production ……………………..……….......... 324

CHAPTER 7. Conclusions.. 329

7.1. Summary of results and contribution.. 329

7.2. Limitations and directions for future research.. 332

7.3. Pedagogical implications: ID and ART training and L2 pronunciation………….. 335

References……………………………………………………………………………… 339

List of Figures

Figure 1.1.	Levelt's (1993) model of language production, where the articulator component is especially relevant for the present study.	50
Figure 1.2.	Vowel systems of Catalan (8 vowels), Spanish (5 vowels) and Southern British English (11 vowels).	58
Figure 1.3.	Assimilation patterns of English vowels to Catalan vowels (in percentages) and goodness of fit (in parenthesis) (adapted from Cebrian et al., 2011).	67
Figure 3.1.	Screenshot of one audiovisual training stimuli showing the talker's face.	110
Figure 3.2.	Button responses distributed in vowel sets in the multiple-choice categorization task (ID1-task1).	115
Figure 3.3.	Button responses per block in the synthesized vowel identification task (ID2-task2).	118
Figure 3.4.	Button responses in the AX vowel discrimination task (DIS-task3).	119
Figure 3.5.	Screen design in each of the blocks of the Delayed Repetition production task (task4).	121
Figure 3.6.	Vowel diagram with 10 vowel spectral distances (Bark).	126
Figure 3.7.	Button responses distributed in vowel sets in ID training task (ID training).	131
Figure 3.8.	Training process in ID training, based on an identification task with feedback.	132
Figure 3.9.	Screenshots of the ID training interface.	133
Figure 3.10.	Training process in ART training, including feedback based on self-monitoring and self-repairs	134
Figure 3.11.	Screenshots of the ART training interface.	135
Figure 4.1.	Average percent correct identification of English vowels for ID training, ART training, NNS control and NS control groups at pre-test.	143
Figure 4.2.	Average percent correct identification of English vowels for ID training, ART training, NNS control and NS control groups at pre-test and post-test.	145
Figure 4.3.	Percent correct identification scores for the three vowel sets (high-front, mid, low) by ID training and ART training groups.	149
Figure 4.4.	Mean percent correct identification at pre-test and pos-test and gains for each vowel set (1= high-front, 2=low, 3=back).	151

Figure 4.5.	Boxplots of identification accuracy across all subjects in the ID training group, presented separately for each vowel at pre- (light blue) and post-test (dark blue).	153
Figure 4.6.	Boxplots of identification accuracy across all subjects in the ART training group, presented separately for each vowel at pre- (light red) and post-test (dark red).	153
Figure 4.7.	Pre-test and post-test identification scores of /e/ obtained by ID and ART training groups and lines indicating between-group differences at pre-test and post-test (p< .05*).	157
Figure 4.8.	Percent correct identification of vowels in audiovisual (AV) and auditory (AV) stimuli presentation conditions by ID training, ART training groups and NS group.	158
Figure 4.9.	Pre-test and post-test identification scores for trained and new words obtained by ID training, ART training and NS groups.	161
Figure 4.10.	Pre-test, post-test and gain scores in trained talker and untrained talker conditions obtained by ID training, ART training and control groups.	164
Figure 4.11. **Figure 4.12.**	Mean error dispersion scores for individual vowels at pre-test (blue) and post-test (red) obtained by ID and ART training groups.	172
Figure 4.13.	Pre-test and post-test error dispersion scores obtained for /e/ at pre-test and post-test by ID and ART training groups.	175
Figure 4.14.	Pre-test/Post-test differences in error dispersion in vowel identification obtained by ID and ART training groups, with negative scores (bars upside down) indicating decrease in error dispersion after training and asterisks indicating significant between-group differences (*p*< .05).	176
Figure 4.15.	Average percent correct identification for synthesized English vowels obtained by ID, ART, NNS control and NS control groups at pre-test.	183
Figure 4.16.	Average percent correct identification of synthesized vowels for ID training, ART training, NNS control and NS control groups at pre-test and post-test.	187
Figure 4.17.	Average percent correct identification of synthesized vowels for ID training and ART training groups at pre-test and post-test.	187
Figure 4.18.	Boxplots of identification accuracy across all subjects in the ID training group, presented separately for each synthesized vowel at pre- (light blue) and post-test (dark blue)	189
Figure 4.19.	Boxplots of identification accuracy across all subjects in the ART training group, presented separately for each synthesized vowel at pre- (light blue) and post-test (dark blue).	190

Figure 4.20. **Figure 4.21.**	Pre-test and post-test identification scores for /æ/ (blue) and /ʌ/ (red) at pre-test (light colour) and post-test (dark colour), presented separately for ID and ART training groups, with light colour representing pre-test and dark colours representing post-test.	182
Figure 4.22. **Figure 4.23.**	Pre-test and post-test identification scores for high-front /iː/ (red), low /ɑː/ (blue) and high-back /uː/ (green) "tense" vowels, presented separately for ID and ART training groups.	195
Figure 4.24. **Figure 4.25.**	Pre-test and post-test identification scores for high-front /ɪ/ /(red), low /ʌ/(blue) and high-back /ʊ/ (green) "lax" vowels, presented separately for ID and ART training groups.	195
Figure 4.26. **Figure 4.27.** **Figure 4.28.**	Mean identification rates of /iː/, /ɑː/ and /uː/ plotted as a function of manipulation of duration for ID training (blue), ART (red) training and control (grey) groups, at pre-test (broken line) and post-test (solid line).	198
Figure 4.29. **Figure 4.30.** **Figure 4.31.**	Mean identification rates of /ɪ/, /ʌ/ and /ʊ/ plotted as a function of manipulation of duration for ID training (blue), ART (red) training and control (grey) groups, at pre-test (broken line) and post-test (solid line).	199
Figure 4.32. **Figure 4.33.**	Mean identification rates of /æ/ and /ʌ/ plotted as a function of manipulation of duration for ID training (blue), ART (red) training and control (grey) groups, at pre-test (broken line) and post-test (solid line).	200
Figure 4.34. **Figure 4.35.**	Mean duration effect scores in vowel identification at pre-test (blue) and post-test (red) obtained by ID and ART training groups.	210
Figure 4.36. **Figure 4.37.**	Pre-test and post-test duration effect scores for individual synthesized represented separately for ID and ART training groups. Line graphs represent increase/decrease in duration effect scores from pre-test to post-test, different line colours indicate vowel set (blue= high-front, green= low, red= high-back), and asterisks indicate significant pre-test/post-test differences in duration effect scores ($p < .05$).	212
Figure 4.38.	Mean identification gain scores for individual synthesized obtained by ID and ART (red) (N=32) training groups.	214
Figure 4.39.	Pre-test/post-test changes in duration effect scores for individual synthesized obtained by ID and ART training groups.	215
Figure 4.40.	Average percent correct vowel discrimination for ID training, ART training and NNS control groups at pre-test and post-test.	223
Figure 4.41.	Mean percent correct discrimination at pre-test and pos-test for each vowel set (1= high-front, 2=low, 3=back) by ID and ART training groups.	226
Figure 4.42.	Mean percent discrimination gains across all subjects in the ID and ART training groups for vowel set 1 (red; high-front), vowel set 2	228

(blue; low) and vowel set 3 (green; back).

Figure 4.43. Mean discrimination accuracy across all subjects in the ID training group, presented separately for each vowel contrast at pre-test (blue) and post-test (red). 229

Figure 4.44. Mean discrimination accuracy across all subjects in the ART training group, presented separately for each vowel contrast at pre-test (blue) and post-test (red). 230

Figure 4.45. Mean discrimination accuracy across all subjects in the control group, presented separately for each vowel contrast at pre-test (light grey) and post-test (dark grey). 230

Figure 4.46. Mean discrimination gain scores for 16 vowels obtained by ID and ART training groups. 232

Figure 4.47. Percent correct discrimination of vowels in fixed context, context variability and all conditions obtained by ID and ART training groups. 234

Figure 4.48. Mean percent discrimination gains in fixed context (yellow) and context variability (green) and all conditions (black) across all subjects in ID training and the ART training. 235

Figure 5.1. Boxplots of mean vowel spectral distances (SD) in vowel productions by ID training (blue), ART training (red), and NNS control (grey) groups at pre-test. 249

Figure 5.2. Mean spectral distances (SD) for the 10 vowel contrasts produced by ID training, ART training, NNS control and NS groups at pre-test, and significant NS-NNS differences ($p < .05$). 251

Figure 5.3. Pre-test and post-test spectral distances (SDs) of vowel productions in all conditions, fixed context, and context variability obtained by ID training group. 254

Figure 5.4. Pre-test and post-test spectral distances (SDs) of vowel productions in all conditions, fixed context, and context variability obtained by ART training group. 254

Figure 5.5.
Figure 5.6. Line graph showing spectral distance scores of ten vowel contrasts produced by ID and ART training groups in the fixed context (orange) and context variability (green) conditions, at pre-test (thinner lines) and pos-test (thicker lines). 258

Figure 5.7. Boxplots of pre-test and post-test spectral distances (Bark) in vowel production for the 3 vowel sets produced by ID and ART training groups. 260

Figure 5.8.
Figure 5.9. Pre-test and post-test spectral distances (Bark) in vowel productions for ID and ART training groups. 263

Figure 5.10.
Figure 5.11.
Figure 5.12. Boxplots of pre-test and post-test vowel spectral distances (Bark) for /iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/ and /æ/-/ɑː/ produced by ID and ART training groups, and significant pre-test/post-test differences ($p < .05$). 265

Figure 5.13.		
Figure 5.14.	Spectral distances of the high-front vowel pairs produced by the ID training group (blue lines) and the ART training group (red lines) at pre-test (thinner lines) and post-test (thicker lines).	267
Figure 5.15.	Spectral distances of the low vowel pair /æ/-/ɑː/produced by the ID training group (blue lines) and the ART training group (red lines) at pre-test (thinner lines) and post-test (thicker lines).	268
Figure 5.16. **Figure 5.17.**	Line graphs showing mean spectral distance scores of ten vowel contrasts produced by ID (blue line) and ART (red line) training groups, at pre-test (thinner lines) and pos-test (thicker lines), and significant pre-test/post-test differences ($p < .05$).	271
Figure 5.18.	Pre-test and post-test spectral distances (SDs) by vowel contrast for ID training, ART training and NS groups.	273
Figure 5.19.	Line graph showing amount and direction of spectral distance change computed for ten vowel pairs from pre-test to post-test, after ID and ART training.	276
Figure 5.20. **Figure 5.21.**	Line graphs showing amount and direction of spectral distance change for /iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/ and /æ/-/ɑː/ after ID training (left) and ART training (right).	276
Figure 5.22.	Boxplots of mean pre-test duration ratio (RT) of English tense-lax vowel contrasts produced by ID training, ART training, and control groups at pre-test.	280
Figure 5.23.	Boxplots of mean vowel duration (ms) for the tense (/iː, ɜː, ɑː, ɔː, uː/) and lax (/ɪ, e, æ, ʌ, ɒ, ʊ/) vowels produced by the ID and ART training groups, showing magnitude of durational change in vowel production from pre-test to post-test.	284
Figure 5.24	Boxplots of pre-test and post-test height (Bark) values in vowel production for ID and ART training groups.	286
Figure 5.25	Degree of height (Bark) in the English vowel system of participants from the ID training and ART training groups, at pre-test (broken line) and post-test (solid line).	287
Figure 5.26	Boxplots of mean vowel height (Bark) for the 11 vowel monophthongs (in all contexts) produced by ID training group, at pre-test (light blue) and post-test (dark blue), and the ART group.	291
Figure 5.27	Boxplots of mean vowel height (Bark) for the 11 vowel monophthongs (in all contexts) produced by ID training group, at pre-test (light blue) and post-test (dark blue), and the ART group.	291
Figure 5.28	Pre-test and post-test height values (Bark) of English /e/ produced in all conditions, fixed context and context variability conditions, by the ID training, ART training and NS groups.	292

Figure 5.29	Degree of height (Bark) in all contexts in the English vowel system of ID training and ART training groups, at pre-test (broken line) and post-test (solid line), and the NS group.	293
Figure 5.30	Degree of height (Bark) in fixed context in the English vowel system of ID training and ART training groups, at pre-test (broken line) and post-test (solid line), and the NS group.	294
Figure 5.31	Boxplots of pre-test and post-test frontness (Bark) values of high (/iːeʊuː/) and low vowels /æʌɑːɒɔː/ produced by ID and ART training groups.	295
Figure 5.32	Pre-test and post-test frontness (Bark) values of English /e/, /uː/, /ʌ/and/ɔː/produced by the ID and ART training groups.	296
Figure 5.33	Degree of frontness (Bark) in fixed context in the English vowel system of ID training and ART training groups, at pre-test (broken line) and post-test (solid line), and the NS group.	299
Figure 5.34	Degree of frontness (Bark) in context variability in the English vowel system of ID training and ART training groups, at pre-test (broken line) and post-test (solid line), and the NS group.	303
Figure 5.35	Pre-test and post-test height (Bark) values of English /e/ produced in all conditions, fixed context and context variability conditions, by the ID and ART training, and the NS group.	305
Figure 5.36	Pre-test and post-test frontness (Bark) values of English /e/ produced by the training group (*N*= 64) at pre-test (blue) and post-test (red) in all conditions, fixed context and context variability conditions.	305

List of Tables

Table 1.1.	PAM-L2's Perceptual Assimilation patterns and predicted degrees of difficulty in non-native vowel discrimination.	32
Table 1.2.	Summary of vowel training studies classified according to type of training: perception training, articulatory training and training studies comparing auditory and articulatory training.	87
Table 3.1.	An overview of the pretest/posttest experimental design with four groups of participants: two experimental groups experimental groups (ID and ART training), a NNS control group and a NS baseline group.	107
Table 3.2.	Training set of natural vowel recordings.	111
Table 3.3.	Testing set of natural vowel recordings.	112
Table 3.4.	An overview of the battery of perception and production tasks used at pre-test and post-test.	113
Table 3.5.	Multiple-choice categorization task (ID1-task1) with auditory (A) and audiovisual (AV) conditions, and trained vs. new natural tokens.	113
Table 3.6.	Forced-choice categorization task of synthesized vowels (ID2-task2) based on 8 vowel continua consisting of 7 duration steps (80, 116.7, 153.3, 190, 226.7, 263.3, 300 ms).	116
Table 3.7.	Natural and modified duration (ms) of the vowels used in the identification task II, and vowel continua.	117
Table 3.8.	AX Vowel Discrimination task (DIS-task3) of 16 natural vowel contrasts.	117
Table 3.9.	Delayed repetition task with auditory (A) vs. audiovisual (AV) presentation blocks, trained vs. untrained words and fixed vs. context variability conditions.	120
Table 4.1.	Percent correct identification scores for the individual vowels and standard deviations by ID training, ART training, NNS control and NS baseline groups, at pre-test.	144
Table 4.2.	Percent correct identification scores for the individual vowels standard deviations in the pre- and post-test by ID training, ART training, NNS control and NS baseline groups.	146
Table 4.3.	Percent correct identification scores and standard deviations for the three vowel sets (high-front, mid, low) by ID training and ART training groups.	149
Table 4.4.	Results of paired-samples t-tests on the pre-test and post-test identification scores of three vowel sets (high-front, mid, low	150

vowels) and accuracy gains obtained by ID and ART training groups.

Table 4.5. Results of paired-samples t-tests on the pre-test and post-test identification scores of the 11 vowels and accuracy gains obtained by the ID and ART training groups. 155

Table 4.6. Results of independent-samples t-tests on the identification scores of the 11 vowels at pre-test and post-test. 156

Table 4.7. Percent correct identification of vowels, gains and standard deviations in audiovisual and auditory stimuli presentation conditions, obtained by ID training and ART training groups. 159

Table 4.8. Pre-test, post-test and gain scores for trained and new words obtained by ID training, ART training and control groups. 161

Table 4.9. Pre-test, post-test and gain scores in trained talker and untrained talker conditions obtained by ID training, ART training and control groups. 164

Table 4.10.
Table 4.11. Vowel confusion patterns observed at pre-test (T1) and post-test (T2) for ID and ART training groups. Stimulus vowels are listed vertically; response categories are listed horizontally. Grey shadowed numbers in boldface (values along a diagonal) indicate average percent correct identification across vowels. Grey numbers indicate infrequent incorrect responses for a particular stimulus-response combination, that is, stimuli which are rarely chosen as a response (<4% accuracy). 169

Table 4.12. Pre-test and post-test error dispersion in vowel identification, pre-test/post-test differences (diff.) and standard deviations for individual vowels obtained by ID and ART training groups. 171

Table 4.13. Results of two-way ANOVAS run on error dispersion scores for individual vowels, with *training* as within-subjects factor and *type of training* as between-subjects factors. 174

Table 4.14. Results of independent-samples t-tests on the pre-test and post-test error dispersion scores of individual vowels. 175

Table 4.15. Results of paired-samples t-tests on the error dispersion scores of the 11 vowel monophthongs obtained by ID and ART training. 176

Table 4.16. Percent correct identification scores and standard deviations for the individual synthesized vowels obtained by ID training, ART training, NNS control and NS baseline groups at pre-test. 184

Table 4.17. Results of multiple comparisons for pre-test percent correct identification of synthesized vowels obtained by ID training, ART training, control and NS groups. 184

Table 4.18. **Pre-test, post-test and gain scores** for the individual synthesized vowels obtained by ID training, ART training, NNS control and NS baseline groups. 188

Table 4.19. Results of paired-samples t-tests on the identification scores of individual synthesized vowels, accuracy gains and standard 191

deviations.

Table 4.20. Results of three-way ANOVAS conducted on mean percent correct identification of each synthesized vowel., with *training* and *manipulation of duration* as within-subjects factor and *type of training* as between-subjects factors. 201

Table 4.21. Results of paired-samples t-tests on mean percent correct identification by manipulation of duration obtained for each synthesized tense vowel. 204

Table 4.22. Results of paired-samples t-tests on mean percent correct identification by manipulation of duration obtained for each synthesized lax vowel. 205

Table 4.23 Pre-test and post-test duration effect scores and gains for individual synthesized vowels obtained by ID and ART training groups. 209

Table 4.24 Results of paired-samples t-tests on pre-test and post-test duration effect scores for individual synthesized vowels and pre-test/post-test differences. 211

Table 4.25 Percent correct discrimination scores for the individual vowel contrasts and standard deviations obtained by ID training, ART training, NNS control and NS baseline groups, at pre-test. 221

Table 4.26 Percent correct discrimination scores, gains and standard deviations obtained in the pre-test and post-test by ID training ART training, control and NS baseline groups. 224

Table 4.27 Percent correct discrimination scores, gains and standard deviations for the three vowel sets (high-front, mid, low) obtained by ID training, ART training and NS groups. 226

Table 4.28. Results of independent-samples t-tests on the discrimination scores of 3 vowel sets (high-front, mid, low) at pre-test and post-test and accuracy gains obtained by training groups. 228

Table 4.29. Results of paired-samples t-tests on pre-test and post-test vowel discrimination scores for 16 vowel contrasts and accuracy gains (diff.) obtained by ID training, ART training and control groups. 233

Table 4.30. Results of paired-samples t-tests on pre-test and post-test vowel discrimination scores for individual vowel contrasts, and accuracy gains, in fixed context and context variability conditions. 237

Table 5.1. Mean spectral distances (SD) (Bark) (and standard deviations) for the 10 vowel contrasts produced by ID training, ART training, NNS control and NS groups at pre-test. 250

Table 5.2. Results of a one-way ANOVA on pre-test mean spectral distances (SD) of ten vowel contrasts with *subject group* (ID, ART training, NS groups) as the between-subjects factor. 252

Table 5.3. Pre-test and post-test spectral distances (SDs and standard deviations of vowel productions in fixed context, context variability and all conditions, obtained by ID training and ART 253

training groups.

Table 5.4. Mean spectral distance scores (Bark) (and standard deviations) for the 10 vowel contrasts in fixed context and context variability conditions, produced by ID and ART training groups at pre-test and post-test. 255

Table 5.5. Pre-test and post-test spectral distance scores (Bark) and standard deviations in vowel production for the 3 vowel sets produced by ID training, ART training NS groups. 259

Table 5.6. Results of paired-samples t-tests on the pre-test and post-test spectral distance scores in vowel production of 3 vowel sets produced by ID and ART training groups. 261

Table 5.7. Results of paired-samples t-tests on the pre-test and post-test spectral distance scores of 10 vowel pairs produced by ID and ART training groups, and pre-test/post.-test differences in spectral distances. 272

Table 5.8. Pre-test and post-test spectral distances (Bark) in the vowel productions of ID training, ART training and NS groups. 274

Table 5.9. Mean duration (ms) (and standard deviations) for the 11 vowel monophthongs produced by ID training, ART training, and NS groups at pre-test. 278

Table 5.10. Mean normalized tense-lax duration ratios (DR) (and standard deviations) for six tense-lax vowel contrasts, produced by ID training, ART training, NNS control and NS groups at pre-test. 279

Table 5.11. A one-way ANOVA on pre-test duration ratios (DR) computed for six vowel contrasts produced by ID, ART and control groups. 280

Table 5.12. Pre-test and post-test tense-lax duration ratios (ms) and standard deviations of six tense- lax vowel contrasts produced by ID and ART training groups. 281

Table 5.13. Results of paired-samples t-tests on pre-test and post-test duration rations of six tense-lax vowel contrasts and pre-test/post-test differences (diff.) in duration ratios. 282

Table 5.14. Pre-test and post-test duration values (ms) (and standard deviations) of tense and lax vowels produced by ID training, ART training and NS baseline groups. 283

Table 5.15. Pre-test and post-test height values (Bark) (and standard deviations) for the 11 vowels produced by ID and ART training groups. 288

Table 5.16. Mean pre-test and post-test height values (Bark) (and standard deviations) for the 11 vowels produced by ID and ART training groups, in fixed context and context variability. 289

Table 5.17. Mean pre-test and post-test height values (Bark) (and standard deviations) for the 11 vowels produced by ID and ART training groups, in fixed context and context variability. 290

Table 5.18. Pre-test and post-test frontness values (Bark) (and standard deviations) for the 11 vowels produced by ID and ART training 295

and NS groups.

Table 5.19. Pre-test and post-test frontness values (Bark) (and standard deviations) for high (/iːeʊuː/), low (/æʌɑːɒɔː/) and mid-central (/ɪɜː/) vowels produced by ID training, ART training, and NS groups. 296

Table 5.20. Pre-test and post-test frontness values (Bark) (and standard deviations) for the 11 vowels produced by ID and ART training groups, in fixed context and context variability conditions. 299

Table 5.21. Results of paired-samples t-tests on the pre-test and post-test height values of 11 vowels produced by ID and ART training groups. 301

Introduction

Foreign Language (FL) teaching and learning in Catalan schools

The present study was initially inspired by the current methodology used in the teaching of English as a Foreign Language (EFL) in the Catalan school Formal Instruction (FI) context. The low English proficiency level within the school age population in Catalonia despite many years of FI is a source of preoccupation for foreign language (FL) teachers, researchers and the regional government (*Consell Superior d'Avaluació del Sistema Educatiu*, Csd'A, 2006, 2008; Tragant, 2009). The failure to perceive and produce English sounds accurately after 7.5 years (726 hours) of FI is acknowledged in the literature (e.g. Fullana, 2005; García-Lecumberri & Gallardo, 2003; Mora & Fullana, 2007; Rallo, 2005). Empirical data on the university entry PAU *(Pruebas de Acceso a la Universidad)* exam in Spain shows very poor results on the English test (CSd'A, 204, 2005, 2006, 2008; Tragant*,* 2009). Anecdotal and qualitative data from surveys, interviews and reports show that both teachers and students are aware of the fact that oral production and pronunciation are neglected in the FL teaching curriculum, especially in the last two years of high-school in Spain (*Batxillerat*) (Tragant, 2009; Tragant et al., 2010).

According to the results of a survey of representative samples of EFL teaching in Catalonia reported by CSd'A (2005, 2006, 2008), the Catalan FI context is characterized by: (1) almost exclusive emphasis on grammar, reading and writing (Lucea, 2004), (2) written input limited to a few readers and the textbook (Frank et al., 2008), and (3) very little use of English for communicative purposes (Goldengerg 2008; Téllez & Waxman, 2006). The lack of pronunciation instruction is particularly evident in this FI context (Bongaerts et al., 1997; Cebrian, 2003; García-Lecumberri, 1999; Rallo, 2005). Instead English grammar teaching is over-emphasized in the FL classroom high-school context and therefore negatively viewed, according to the interviews conducted on students (CSd'A, 2005).

Our tentative conclusion is that the methodology used in the EFL classroom in school and high-school contexts in Catalonia has failed to be a conclusive determinant for more accurate perception and production of English sounds, regardless of differences in age of onset of L2 learning (AOL) (8 and 11-year-old beginners). There is enough evidence from research in Second Language Acquisition (SLA) that the EFL classroom generally fails to offer enough practice in oral skills or to provide training in the perception and production of EFL sounds. Qualitative data from recent surveys and interviews (e.g. Tragant et al, *in press*) shows that experienced teachers and school inspectors admit that, with a few exceptions, the FL classroom in Catalonia offers very few opportunities to practice oral skills and provide pronunciation feedback (CSd'A, 2005). Research in this context offers evidence of the need to provide opportunities for oral practice to improve oral skills and pronunciation (García-Lecumberri & Cenoz, 1998; Rallo, 2005; Fullana, 2006).

Factors affecting L2 speech learning in the EFL classroom setting in Catalonia

Quantity and quality of Second Language (L2) input have been identified as good predictors of successful L2 phonological acquisition in immersion settings (Flege, 1991; Flege, Piske, & Mackay, 2001; Flege, 2009; Piske, 2007; Piske et al., 2001; Piske & Young-Scholten, 2009). The major role of high-quality L2 input and exposure to NS input in these settings is supported by foreign accent (FA) studies (e.g. Bongaerts, van Summeren, Planken, and Schils, 1997). Piske et al. (2001) found that frequent use of First Language (L1) correlates with a relatively strong degree of FA in the L2.

However, the positive effects of quantity of L2 input on adult SLA are also evident in FL learning contexts. Studies conducted in a FL learning context show that amount of

L2 input influences late learners' FL pronunciation (Moyer, 1999). These also put forward that practice on oral skills and pronunciation is a key factor in successful phonological acquisition. Bongaerts et al. (1997) showed, for instance, that the amount of L2 input administered through *"intensive training both in the perception and in the production of the speech sounds of British English"* (Bongaerts et al., 1997, p. 463) might have led to late learners' successful attainment of native-like pronunciation.

The literature in the field of EFL teaching methodology in the school context of Spain and Catalonia (Berga et al., 2008; Lucea, 2004; Martorell, 2006) and research in SLA in Spanish school contexts (Muñoz, 2007) also suggest that quantity and quality of input and a communicative approach (Tellex & Waxman, 2006) are factors conducive to successful FL learning in FI contexts. Therefore, failure to perceive or produce English sounds at native-like levels after so many years of FI might be in part attributed to the limited quantity of input and its foreign-accented nature (Flege, 2009; García-Lecumberri & Gallardo, 2003). It has been hypothesized that heterogeneous NNS input in the EFL classroom does not favour L2 phonological acquisition and this exposure may lead to more foreign-accented sound production (*accented L2 input* hypothesis, Flege, 1991).

Some researchers have consequently suggested a change in FL teaching methodologies. García-Lecumberri and Cenoz (1998) argue in favour of a pronunciation component. They propose that pronunciation practice should specifically help learners focus on the spectral differences for vowel contrasts rather than on temporal cues. Rallo (2005: 140) suggests that auditory discrimination and oral repetition tasks should be included in the FL classroom. Along the same lines, Fullana (2006) acknowledges the fact that FL teaching must include higher quantity and quality of NS input to guarantee a positive effect of FI on the learners' perception and production of L2 sounds.

Apart from the factors identified, the absence of pronunciation practice in high-school classroom contexts has been attributed to the lack of external pressure to assess speaking skills or pronunciation given the mostly written nature of the PAU exam. In fact, only listening skills but not production skills are assessed in the PAU. Additionally, the literature attributes the methodological drawbacks to the limitations imposed by the Spanish and Catalan administration: too many students per class, too few hours of exposure to the FL, heterogeneity of groups and, consequently, difficulty to deal with within-group individual differences.

Catalan-Spanish learners' acquisition of English vowels

Research conducted in the Catalan FI context has not shown significant positive effects of years of L2 learning on learners' accuracy in English vowel perception and production (Fullana & Mackay, 2003; Fullana, 2006; Mora & Fullana, 2007). The fact that Catalan-Spanish learners of English identify English vowels poorly lies in the nature of their L1 system. The Spanish and Catalan vowel systems consist of 5 and 7 vowels, respectively, and, unlike English, there are no tense-lax or temporal contrasts in either of these languages.

Catalan learners of English have been shown to rely on durational cues more than native English speakers in the discrimination of English vowel contrasts based on vowel quality and duration differences (Aliaga-Garcia & Mora, 2007, 2009; Cebrian, 2006, 2007; Cerviño & Mora, 2009; Escudero & Boersma, 2004; Escudero, 2006; Mora & Fullana, 2007; Morrison, 2008). Fullana-Rivera and MacKay (2003), for instance, found neither significant effect of AOL nor L2 experience on the production of English /iː/–/ɪ/ by eight different groups of Catalan learners of English (aged 11–28).

Previous work (Escudero 2000; Cebrian 2006) has generally attributed Catalan-Spanish learners' poor perception and production of English vowels to their L1-L2 assimilation patterns: e.g. English /iː/ and /ɪ/ are assimilated to their native /i/). Difficulties recognizing and weighting the most reliable cues to L2 vowel categorization might also account for inaccurate perception and production (e.g. Aliaga-Garcia 2011; Cebrian 2007; Cerviño & Mora, 2010; Escudero & Boersma, 2004; Flege et al. 1997; Morrison, 2002).

Given the fact that no considerable improvement in pronunciation has been found despite many years of FI and pronunciation has been neglected in FI settings (e.g. Fullana, 2005; Piske, 2007), this study points to the urgent need for innovation in the EFL classroom in Catalonia. Taking proposals for pronunciation teaching further into consideration, this thesis puts forward the possibility to include *effective* phonetic training within the EFL classroom.

Phonetic training methods

Within the FI setting, experience-related factors such as quality and quantity of input (Flege & Liu, 2001; Moyer, 2009) have been identified as crucial factors in L2 speech learning. L2 phonetic training has also been identified as an effective method to improve L2 perception and production (e.g. Aliaga-Garcia, 2011; Aliaga-Garcia & Mora, 2009, Flege 1995; Iverson & Evans, 2007, 2009; Nishi & Kewley-Port, 2007, 2008). Phonetic training studies constitute a promising area of research in the light of arguments in favour of plasticity in L2 speech perception and improvement in the production of L2 sounds (Jamieson & Morosan, 1989; Logan & Pruitt, 1995; Iverson & Evans, 2007; Ylinen et al., 2009). The effectiveness of phonetic training to improve L2 sound perception and production includes vowels, which pose an important challenge for L2 learners (Strange, 2007). This thesis proposes that phonetic training should be an integral component of FL teaching.

Among the different training methodologies implemented (e.g. Perceptual Fading, All Enhancement, Secondary Cue Variability, etc.), High-Variability Phonetic Training (HPVT) (Hazan et al., 2005, Iverson et al., 2005) has been consistently shown to be the most effective type of auditory training. HVPT has proved especially effective when the difficulties with L2 sounds lie in the use of L2 phonetic cues which are not used or are weighted differently in the L1 (Cebrian 2006; Escudero & Boersma, 2004; Holt & Lotto, 2006). Audiovisual (AV) phonetic training has also been shown to yield greater gains in L2 perceptual learning than only-auditory (A) training (Hardison, 2003, Hazan & Sennema, 2005; Ortega-Llebaria et al., 2001).

At the moment the number of perceptual training studies showing improvement in L2 perception (and transfer to improvement in L2 production) clearly outnumbers the number of production training studies conducted so far (e.g. /r/-/l/ training: Hattori & Iverson, 2009; vowel training: Lambacher et al., 2005; Iverson & Evans, 2007, 2009; Lengeris & Hazan, 2010).

As regards vowel training, research on the assessment of methods other than perceptual training for non-native vowels is still scarce and inconclusive (i.e. Catford & Pisoni, 1970; Flege, 1989). The relatively few vowel training studies in the literature have examined Japanese (Lambacher et al., 2005; Nishi & Kewley-Port, 2007) or Korean speakers (Nishi & Kewley-Port, 2008), German and Spanish speakers (Iverson & Evans 2007, 2009), and Greek speakers' (Lengeris, 2009) perception and production of English vowels after perceptual training.

Statement of the problem

To my knowledge, no study to date has compared the impact of perception vs. production-based training on L2 vowel perception and production. Similarly, no training studies have addressed the question of whether native-like acoustic cue weighting is best promoted by perception or production-based training for Spanish or Catalan speakers. This gap in the literature has also motivated the comparison between perception and production-based vowel HPVT in this study.

Aims

The aim of the present study is to extend research on L2 vowel phonetic training by comparing the effectiveness of two types of AV HVPT, namely Identification (ID) and Articulatory (ART) training, on L2 vowel perception and production and cue weighting by advanced Catalan-Spanish learners of English.

This study aims to make several contributions within the general field of SLA and EFL teaching. First, it intends to contribute to the phonetic training field by testing the effectiveness of two HVPT methods. Second, it tries to shed new light on the ongoing debate on the relationship between L2 perception and production by comparing the size of accuracy gains in each domain following different training methods. Third, it seeks to contribute to the understanding of L2 vowel learning by examining (1) whether a 10-session HVPT can lead to significant changes in L2 perception and production and (2) whether gains in L2 perception and/or production are best promoted by perception- or production-based AV HVPT. Finally, it seeks to make a contribution to EFL teaching practices in the Spanish/Catalan FI context by suggesting ways of helping learners to improve their perception and/or production of English vowels. More accurate pronunciation would otherwise be very unlikely achieved in the regular EFL classroom without specific pronunciation practice or training on the segmental phonology of English.

Overview of the Thesis

This PhD dissertation consists of five chapters. Chapter 1 presents the relevant theoretical background to this study, including an overview of the most relevant speech perception and production models and theories in adult L2 speech learning in a FI setting. It also reviews perception and production training studies and identifies the gaps in the literature which call for this research. Chapter 2 states the main research questions and hypotheses. Chapter 3 is dedicated to the experimental design of the study. It includes an in-depth description of the participants, a description of the commonalities and differences between the two HVPT methods, and a description of the battery of pre- and post-tests. Chapter 4 consists of a discussion of results and, finally, Chapter 5 presents the conclusions and implications of the study as well as suggestions for follow-up research.

PART I

BACKGROUND TO THE STUDY

Chapter 1

Literature review

This chapter deals with the origin of the difficulties which Catalan-Spanish speakers have perceiving and producing the English vowel monophthongs (/iː ɪ e ɜː æ ʌ ɑː ɒ ɔː ʊ uː/) and presents current perceptual and articulatory models that, on the one hand, will offer several explanations for understanding learners' difficulties and, on the other hand, will uncover the reasons for designing and comparing two vowel training methods. It also presents an overview of the most relevant training methods which have had a positive effect on L2 vowel learning. This chapter is structured as follows:

Section 1.1 deals with the complex relationship between perception and production as far as the acquisition of L2 speech sounds is concerned. Although there is yet much debate over whether perception or production is acquired first in second language acquisition (SLA), the vast majority of training studies, especially vowel training studies, have assumed a perceptual basis for their methodology.

Sections 1.2 and 1.3 explain how L2 speech perception and production theories relate to the two vowel training methods tested and compared here, identification (ID) and articulatory (ART) training.

Section 1.4 describes the vowel systems of Spanish, Catalan and English and presents the findings of previous experiments investigating cross-language assimilation patterns and the production of English vowels by Spanish monolinguals or bilingual Catalan-Spanish speakers.

Finally, Section 1.5 summarizes and discusses the most relevant studies conducted in the phonetic training field, with a particular focus on L2 vowel learning (see Table 1.2). It aims at defining the gaps in the literature which call for further research on vowel training methods.

1.1. L2 Speech Perception and Production in Adulthood

1.1.1. Overview of the perception-production relationship in SLA: acquisition of L2 vowels and vowel training

This thesis is concerned with the acquisition of English vowels by adult native speakers (NSs) of Catalan and Spanish, and the effectiveness of phonetic training in "modifying" their perceptual and articulatory skills in terms of 'perceptual accuracy' and 'conscious tongue movements', respectively. An important premise of the L2 phonetic training field is that speech perceptual and articulatory systems remain *malleable* in adulthood. This is so because learners retain sufficient sensitivity to acoustic features in adulthood, and they show ability to learn and restructure gestural and motor information (e.g. lips and tongue motor areas) to produce L2 sounds.

The relationship between speech perception and production has been a long-standing matter of debate in SLA as well as in first language (L1) acquisition (i.e. Koerich, 2006; Leather, 1999; Llisterri, 1995; Wode, 1999). On the basis of the existing literature, it seems that perception and production abilities are somehow linked in L2 speech learning but these are not brought into perfect alignment. The precedence of one over the other in formal instruction (FI) contexts has not been clearly established yet.

ºOver the course of the past four decades, numerous studies have made it possible for us to describe the nature and extent of the difficulties encountered by non-native speakers (NNSs) when perceiving and producing L2 speech sounds, and how these may persist after years of immersion in the non-native language environment (Best & Strange, 1992; Flege & Eefting, 1987; MacKain, Best, & Strange, 1981; Yamada, 1995) or years of formal instruction (FI),

Among the numerous drawbacks of FI, it is worth mentioning the limited exposure to the FL and minimal input (Celce-Murcia et al., 1996); Larson-Hall, 2008; Muñoz, 2006), the foreign-accented nature of the input (Derwing & Munro, 2015,), the lack of focus on L2 pronunciation teaching (Celce-Murcia et al, 1996); the lack of effective pronunciation teaching materials (Darcy et al., 2012). All theseexplain the fact that numerous studies have reported mixed effects as regards L2 perception and production (Fullana & Mora, 2009; García-Lecumberri & Gallardo del Puerto, 2003; Pérez-Vidal et al., 2011). Several factors contribute to a foreign accent in SLA given the difficulties that L2 learners have perceiving and producing target sounds accurately. Some studies have reported the effects of the starting age of learning (younger age of learning) on the perception of L2 vowels (Flege et al., 1999; Fullana & Muñoz, 1999) and consonants (Fullana & Muñoz, 1999; MacKay, Meador, & Flege, 2001), and on the production of vowels (Flege et al. 1999; Piske et al., 2001) and consonants (Mackay et al., 2001). Learners exposed to the target language (TL) in childhood (early learners) tend to outperform learners exposed to the TL later in life (late learners) in perception and production (Flege et al., 1999; Tsukada et al., 2003). Traditionally, SLA research has examined age of onset of L2 learning (AOL) (e.g. 6-12 years old) with reference to an upper-limit or Critical Period Hypothesis (CPH; Birdsong, 1999; Lenneberg, 1967) after which acquiring native-like L2 pronunciation is not possible, i.e., after which biological maturational constraints prevent L2 learners from achieving a native-like accent in their L2.

In addition to the starting age of learning a FL, other factors accounting for difficulties in L2 perception and production (Bohn & Flege, 1997) and foreign accent have explored intervening factors such as length of residence in the TL country, influence of the L1 phonological system and amount of L2 experience.

As L2 experience increases, production may become less malleable than perception. Flege (1999) argued that at least *"modest correlations will exist between L2 segmental production and perception for highly experienced speakers of an L2"*. In support of this view, several studies have shown modest correlations between L2 learners' vowel discrimination and production skills (Flege, Bohn et al. 1997; Flege, Mackay et al., 1999; Cebrian, 2002).

The L1 sound system is considered to exert a tremendous influence on L2 speech learning by limiting the degree of success with which L2 speech sounds are acquired, and it is considered thus responsible for the L2 learners' difficulties in L2 phonological acquisition. Already in 1939, Trubetzkoy had proposed that bilinguals tend to perceive L2 sounds through their own L1 phonology, which leads to accentedness.

A great body of research on L2 vowel learning has examined adult learners' perception and production through a variety of tests consisting of both natural as well as synthetic vowel stimuli . These have shown that languages vary in the nature as well as in the size of their vowel inventories and, consequently, pose different degrees of difficulty in perceiving and producing L2 vowel categories (i.e. Iverson & Evans, 2007). L2 vowel learning is considered especially difficult for L2 learners with relatively simple "uncrowded" L1 vowel systems (e.g. 5-vowel and 8-vowel systems of Catalan and Spanish, respectively). This is because speakers with smaller vowel systems may assimilate several L2 vowels into single L1 categories (Escudero & Boersma, 2004; Flege et al, 1997; Iverson & Evans, 2007) or may use fewer acoustic features to distinguish L1 vowels and, reasonably, lack novel acoustic features that are necessary in L2 vowel categorization (Bohn 1995; Iverson & Evans 2007).

Failure to produce non-native vowels accurately has also been attributed to learners' difficulties in recognizing and weighting reliable cues to L2 vowel categorization (Cebrian 2006; Escudero 2000). A good example is Catalan-Spanish speakers, who tend to focus on non-critical durational cues rather than spectral information (i.e. Aliaga-Garcia 2011; Cebrian 2007; Cerviño & Mora, 2010, 2015; Escudero & Boersma, 2004; Flege et al. 1997; Morrison, 2002), leading to poor perception and/or production of English vowels. Research has provided adequate descriptions of the difficulties that Catalan-Spanish learners of English typically have learning certain phonemic contrasts in the target language (TL), for example, the English /iː/-/ɪ/ contrast, as they have only one high-front vowel (/i/) in their L1 which lies between these two sounds.

Against this background, and despite age constraints and the limitations imposed by the L1 sound system, perceptual training studies have offered ample evidence of plasticity in learners' speech perception and production. Early identification (e.g. Jamieson & Morosan, 1986; Strange & Dittman, 1984) and discrimination training (e.g. Carney et al., 1977; Edman, 1980; Pisoni et al., 1982) methods have served as a productive testing ground for claims about adult neural plasticity and malleability in L2 perception and production for adult learners. These studies have shown that a short period of intensive perceptual training can significantly improve the perception and production of L2 sounds. Improvement following perceptual training has been shown to generalize beyond stimuli and speakers heard in training. These findings support the view that the perceptual system remains plastic over the life span. However, at present perceptual training studies outnumber articulatory training studies. Exploring this line of research further, the current study addresses the question of what are the differential gains learners obtain in the perception and production of L2 sounds following auditory or, alternatively, articulatory training.

There are two main reasons for studying and comparing two types of training:

First of all, the belief that perceptual training leads to improvement in both the perception and production domains (e.g. Lambacher et al., 2005; Iverson & Evans 2007, 2009) brings up two questions:

(1) whether articulatory training could be an effective tool for L2 speech learning;

(2) whether improvement in production after articulatory training would transfer to the perceptual domain.

It should be noted that the perception-based training which has typically proved effective in previous research may implicitly stimulate the use of auditory-articulatory mappings despite not including any explicit articulatory component.

Secondly, an in-depth review of the literature confirms that direct inferences about pronunciation cannot be made from perception in a straightforward manner. Although most studies favour the view that pronunciation errors have a perceptual basis (*perception leads production* hypothesis; Bradlow et al., 1997; Flege, 1995; Rochet, 1995), the precise nature of the perception-production link is still not clearly understood (Llisterri, 1995). Other authors claim that the perceptual abilities of highly proficient L2 learners are highly resistant to improve as much as their production, suggesting that *perception may not lead production (*Kartushina, 1995).

Empirical studies on the relationship between perception and production in L2 speech learning are still very limited and the results of such studies have not always been consistent (see Section 1.5.4). Therefore, we are uncertain as to what extent the perceptual and production abilities of advanced learners of English, the subjects of our study, are related, assuming their exposure is restricted to a FI instruction setting characterized by a variety of native and non-native lecturers. It is also unclear to what extent advanced L2 learners' perception and production are interrelated given high levels of motivation, high-quality input and effect of training in the perception and production of L2 speech sounds. Further research urgently needs to fill in these gaps in the literature field.

There is a distinct lack of research on phonetic training involving articulatory practice. Some of the mixed findings reported in the following sections (1.4.3. and 1.5.2-1.5.5) suggest further research comparing the training methods used in the literature is necessary.

1.1.2. Perception leads production

This section reviews some cross-sectional studies that have shown that L2 perception and production are moderately related (i.e. Bradlow et al., 1997; Koerich, 2006; Rauber et al., 2005;) or closely linked, with the precedence of perception before production, especially for vowels (Barry, 1989; Bohn & Flege, 1990; Bradlow et al., 1997; Elsendoorn, 1984; Flege, 1987, 1995; Llisterri, 1995; Rauber et al., 1995; Rochet, 1995; Trubetzkoy, 1962). It has been hypothesized that incorrect perception of L2 sounds may lead to inaccurate pronunciation of these sounds (Flege, 1995). The hypothesis that ***perception leads production*** is of much importance in a study like this which compares auditory training against articulatory training.

There is now an extensive literature assuming that L2 perception is primary to L2 production and documenting the perceptual and production problems experienced by adult learners. These studies hypothesize that there are numerous factors such as AOL, L1 background or frequency of L2 use, that may affect L2 perception and, consequently, L2 production (Best et al., 1988; Best, 1994; Kuhl & Iverson, 1995; Pisoni, et al., 1994; Werker & Tees, 1984).

Research findings have generally supported Flege's Speech Learning Model (SLM, 1995) and one of its main tenets, namely that increased experience may lead late learners to discern phonetic differences between L1 and L2 sounds. According to the SLM, the development of new phonetic categories for L2 sounds will eventually lead to greater accuracy in the production of L2 sounds (Flege, Bohn, & Jang, 1997; Frieda & Nozawa, 2007). Flege's SLM operates on the principle that accurate perception is necessary for achieving native-like production of L2 sounds. According to the findings from Flege's research concerning non-native vowel perception and production, failure to perceive the target sounds accurately arises from learner's difficulties perceiving differences between close L2 and L1 sounds (see section 1.2..2. to read more about *equivalence classification*). Consequently, inaccurate perception will lead to lack of accuracy in production.

Variation in L2 sound perception and production has been attributed to the role of several factors, such as (1) differences between the L1 and L2 phonological systems (Best, 1995; Flege, 1995), (2) AOL, (3) age of arrival and length of residence (LOR) in an L2 setting and (4) amount of L1/L2 use. Moreover, younger AOLs, longer LORs and more frequent use of L2 are related to less foreign-accented speech (Piske, MacKay, & Flege, 2001). Exposure to high-quality input (Flege & Liu, 2001) and training in the perception and/or production of L2 sounds (Aliaga-Garcia, 2010; Aliaga-Garcia & Mora, 2009; Flege, 1995; Iverson & Evans 2007) have also been identified as crucial factors in FL teaching contexts (Piske, 2008). The relative strength and degree of interaction of these factors in specific learning contexts are likely to explain much of the variation in L2 learners' degree of attainment in L2 phonological acquisition.

1.1.2.1. Vowel studies supporting the *perception leads production* hypothesis

Llisterri (1995) noted that studies concerning L2 vowel learning in particular have consistently demonstrated a close link between L2 perception and production. In the early 1980s, studies examining the correlation between perception and production of non-native vowels supported the hypothesis that vowel *perception leads production*. This suggested in turn that perceptual abilities may be good predictors of accuracy in vowel production.

Barry (1989) found that German speakers' inaccurate production of English vowels could be ascribed to their inaccurate perception. They relied on durational differences in the identification of English /æ/ and /e/, in contrast with native English speakers' perception, which was based on spectral differences.

Along the same lines, other studies in the 1980s put forward that the complex perception-production link is influenced by factors such as L2 experience, context and the nature of the acoustic cues. Elsendoorn (1984) investigated the role of L2 experience by assessing the production and perception of English /iː ɪ æ uː/ by Dutch speakers varying in proficiency. Although the effect of consonants on vowel length is normally neutralized in word-final position in their L1-Dutch, learners were found to change their vowel duration relative to native English speakers as they gained more L2 experience. Experienced learners produced full vowel length or shortened vowel duration before word-final lenis or fortis consonants, respectively, closer to native-like values. However, the use of the duration cue of less proficient learners seemed to be influenced by their L1. Consequently, less experienced learners produced L2 vowels with a FA (e.g. vowel duration differences were neutralized preceding word-final obstruents as in their L1-Dutch).

The influence of L2 experience in relation to vowel perception and production was examined by Bohn & Flege (1990), who compared the perception and production of English /æ/ and /ʌ/ and found significant differences between two groups of German learners of English differing in LOR (5 years vs. 6 months) in the US. Whereas inexperienced learners relied primarily on duration when identifying and producing English vowels of the same continuum, experienced learners showed spectral differences in the two modes. These results confirmed Elsendoorn's findings (1984) that L2 experience affects the nature of the perception-production relationship. The fact that inexperienced German learners of English contrasted English vowels only perceptually (despite doing it in a non-native manner) allowed Bohn & Flege (1990) to conclude that *perception leads production* in the early stages of L2 learning, but not after increased L2 experience. However, Bohn and Flege (1990) noted that L2 vowel perception abilities and non-native speakers' (NNSs) strategies differed significantly from those of NSs regardless of L2 experience.

The perceptual basis of production errors was also reported in an experiment by Rochet (1995). NSs of Canadian English and Brazilian Portuguese in Rochet's study were asked to imitate French /y/ and identify synthetic high vowels from a /i/-/u/ continuum. The results showed that the degree of success in their imitation of L2 sounds was closely related to their accuracy in L2 perception. English and Portuguese speakers identified French /y/ as /u/ and /i/, respectively, and these results matched those in production. This means that the differences in the perception accuracy of French /y/ were partly responsible for variation in the production accuracy of that vowel.

A great deal of research has focused on the perception and production of English vowels by Italian learners of English differing in the degree of L2 experience (e.g. early vs. late speakers). These results have helped to establish a more solid link between vowel perception and production. Flege, Mackay, & Meador (1999) found that NNSs with averaged age of arrival (AOA) of 25-26 years in Canada who obtained better discrimination scores for /ʌ/-/a/ also showed more accurate productions of /ʌ/.

The influence of both L2 experience and AOL in relation with L2 perception and production was examined in Munro, Flege, & Mackay (1996) and Piske et al. (2002). They found that the production of /ɚ / by experienced L2 learners was related with different AOL in Canada. More vowel productions of early learners (with early AOL) than late learners (later AOL) fell within mean values obtained by native English speakers. Quite interestingly, many late learners and almost all early learners examined in Munro & Mackay (2004) obtained high scores in the identification of /ɚ / and /ʌ/-/ɚ/ discrimination. In agreement with Munro et al. (1996), the results obtained by Flege, Mackay, and Schirru (2003), who examined the vowel production of those participants from Munro & Mackay (2004), showed that more vowel productions of early learners than late learners resembled the NS productions. When early and late learners were considered separately, moderate correlations were also found between the mean discrimination scores for /ɒ/-/ʌ/, /iː/-/ɪ/ and /ɜː/-/æ/ and the ratings given to each learner's average production of /ɒʌiːɪɜːæ/. Altogether these findings agree with Rochet's research which posits that *perception leads production*.

More recently, a study by Rauber et al. (2005) found that the perception and production of English vowels by highly proficient Brazilian speakers were correlated. The findings revealed different degrees of accuracy in vowel perception, as assessed by an oddity discrimination task, was found to be significantly related to learners' target-like first frequency formants (F1) and second frequency formants (F2) in vowel production, as assessed by a reading aloud task with vowels embedded in a /bVt/ context. On the contrary, L2 vowels that were poorly discriminated were produced with F1 and F2 values similar to those of the closest L1 vowel. This supported Rochet's and Flege's view that accurate L2 perception is a prerequisite for accurate L2 production.

1.1.2.2. Perceptual training improves L2 sound production

Relevant to the *perception leads production* hypothesis are perceptual training studies with no production practice component. Improvement in production after perceptual training is especially relevant since this could be taken to mean that training in the perception of L2 sounds transfers to improvement in their production.

Bradlow et al. (1997) trained Japanese learners of English on the perception of /l/ and /r/ and found that the learners' productions were consistently judged to be more intelligible after perceptual training. Only one participant improved perception without showing improvement in production. Similarly, Japanese speakers who were trained to identify English /r/-/l/ minimal pairs (Yamada et al., 1996) showed significant improvement in perception and production. Furthermore, subjects retained improvement in these abilities in follow-up tests given three months and six months after the training period. This demonstrates that perception-based training might produce long-term changes in both perception and production.

Hazan et al. (2005) replicated this study and confirmed a close link between perception and production, and gains in the two domains were retained three months after perceptual training. The Japanese learners' pronunciation of the /v/-/b/-/p/ (labial/labiodental) and /l/-/r/ contrasts improved significantly following auditory and audiovisual perceptual training. Audiovisual perceptual training led to greater improvement in pronunciation than auditory perceptual training, especially when visual cues to the consonant contrast were salient.

The findings of more recent vowel training studies by Lambacher et al. (2005) and Lengeris and Hazan (2010) have confirmed the success of consonant training for L2 production reported in the studies above. Although Section 1.5 in this chapter offers a more detailed review of training studies, it should be reminded at this point that improvement of vowel perception after perceptual training has been shown to translate into improved vowel production in the case of Japanese (Lambacher et al., 2005) and Greek (Lengeris & Hazan, 2010) speakers, as revealed through NS ratings and acoustic analyses, respectively (See Table 1.2).

1.1.3. *Production leads perception*

While the hypothesis that *perception leads production* has much support (Bradlow et al., 1995; Flege, 1987, 1995; Neufeld, 1980; Rauber et al., 1995; Rochet, 1995; Trubetzkoy, 1962), some other studies suggest that target-like perception of L2 sounds does not necessarily precede or lead their production (Borrell, 1990; Brière, 1966; Llisterri, 1995; Neufeld, 1988; Goto, 1971; Sheldon & Strange, 1982). Early evidence supporting the view that *production leads perception* has suggested that in L2 speech learning, accurate perception is not always aligned with accurate production of those sounds.

The following section will report on studies which support the hypothesis that speech sound *production leads perception*.

1.1.3.1. Studies supporting the *production leads perception* hypothesis

Early in the 1960s Brière (1966) examined Japanese speakers' perception and production of the English /r/ and /l/ contrast and found that their production abilities did not match but were superior to their perception. Brière suggested that L2 learners were capable of producing sounds of an artificial language (composed of Arabic, French and Vietnamese elements) through extensive training, despite their incorrect perception of the sounds.

Similarly, Goto (1971) found that Japanese learners of English showed better production than perception of /r/-/l/, although it was noted that their successful performance in production may have been cued by written stimuli. Their perception of those sounds was nevertheless consistently poor. Borrell (1990) suggested that accurate perception of L2 sounds does not necessarily lead to accurate production and pointed to NNSs' difficulties pronouncing Spanish /r/ despite their accurate perception of the sound. This suggested in turn that mastering the motor mechanics (in relation to the tongue) should facilitate the perception of L2 speech sounds.

Sheldon and Strange (1982) investigated Japanese learners' ability to perceive and produce English /r/ and /l/. Learners were found to perform better in production than in perception, which is at odds with SLM's (1995) claim that accurate L2 segmental production cannot occur unless there is accurate perception. Sheldon and Strange (1982) argued that perceptual mastery of L2 sounds does not necessarily precede mastery in production. Their experiment replicated Goto's findings with six proficient Japanese learners of English with minimum one-year residence in the US. Importantly, the results followed the same pattern as Goto's: their production of the /r/-/l/ contrast was found to be significantly better than their perception of the contrast. Specifically, native English listeners judging the participants' productions of /r/-/l/ minimal pairs could distinguish the target words better than the participants could in their own productions. Sheldon (1985) concluded that the longer the L2 exposure, the less likely it was that their perception became superior to their production. Nevertheless, Sheldon and Strange's (1982) results were not uncontested: Flege (1991) argued that success in producing English liquids could be partially attributed to the "monitored" production of the sounds. Flege believed that instructions to articulate /r/-/l/ included in controlled settings could have eventually played a significant role in the results. Eckman (2004) suggested that the fact that L2 learners received input that was not auditory (e.g. visual input) could explain that their production abilities exceeded their perceptual skills.

Llisterri (1995), in his overview of the status of the perception-production link, mentions further studies in which L2 sound production has also been found to surpass perception (Flege & Eefting, 1987; Sheldon, 1985; Rochet, 1995). For instance, Flege & Eefting (1987) had found that Dutch learners of English distinguished /b/-/p/ stop consonants in their two languages much better in production than in perception 1987. This findings had been justified by Mack (1989) and Sheldon (1985), who argued that social and psychological pressure to show an accurate non-native speech production accounted well for those cases in which L2 perception does not lead production .These studies have challenged one of the main tenets in Flege's SLM, suggesting that production of L2 sounds can be more accurate than their perception as a function of AOL and L2 experience (e.g. in highly proficient learners of English).

The results obtained in Rochet (1995) might be taken as a challenge to the *perception leads production* hypothesis too, as Korean speakers showed more target-like production than discrimination of English vowels. The fact that some learners showed accurate imitation of those L2 vowels which were not accurately discriminated calls for caution in concluding that L2 segmental perception always precedes L2 segmental production.

More recently, Kluge et al. (2007) have shown that there is a directional relationship between production and perception in SLA, with production leading perception: the better the production, the better the perception of the L2 speech sounds. Kluge et al. (2007) studied the production and perception of English /m/ and /n/ in coda position by speakers of Brazilian Portuguese. According to their findings, the perception of the place of articulation of final nasal consonants may be more inherently difficult than the articulation of these sounds, given the fact that their L1 permits the nasal consonants in two other positions (e.g. word-initial and word-medial).

Likewise, studies in bilingualism (Caramazza et al. 1973, Elman et al. 1977, Mack 1989) have shown that production can be more accurate than perception. For instance, Caramazza et al. (1973) tested the perception and production of voiced and unvoiced consonants among Canadian English-French bilinguals. They found that the production of their less proficient or non-dominant language was better than its perception.

The relationship between L2 perception and production has also been explored in research with brain-lesioned patients, who can be selectively impaired in either perception (Praamstra et al., 1991) or production. This type of research provides empirical evidence of a partial dissociation between perception and production of non-native contrasts, and eventual correlation seems to be restricted to certain measures (Hattori & Iverson, 2009). It can also be taken as evidence that perception and production skills for L2 contrasts develop differently (Neufeld, 1988; de Jong et al., 2009).

Rallo-Fabra and Romero (2012) have recently reported lack of overall significant correlation between Catalan learners' perception and production of English vowels /iɪeæ ʌɑːʊuː/at individual and group levels. Their findings do not parallel prior studies showing that perceptual and production abilities were aligned to varying degrees. At the individual level, only marginally significant correlations were obtained for mid-proficiency learners. Whereas native listener judgements found /æ/ to be produced consistently intelligible, discrimination ratings for vowel contrasts including /æ/ was low.

1.1.3.2. Production training improves perception

There is evidence that perceptual learning through training can be transferred to production. However, the research reporting carry-over effects of production-based training on perception skills is scarce for L2 consonants and almost non-existent for vowels.

Matthews (1997) investigated the perception of 6 English consonant contrasts (/b/-/v/, /θ/-/f/, /s/-/θ/, /l/-/r/ and /s/-/f/) by Japanese speakers after a purely production-based training which consisted of articulatory explanation and mimicking, silent articulation and visual presentation of the sounds. Feedback never included acoustic models, as researchers tried to prevent trainees from developing the type of stimulus-dependent representations that are frequently triggered by perceptual training. Their results indicated that articulatory training with silent and visual feedback leads to significant improvement in L2 consonant discrimination. However, improvement was restricted only to certain L2 sounds and no improvement was seen in /l/-/r/ and /s/-/f/ discrimination.

Those studies which have reported production-based training effects on perception (Mathews, 2007; Hazan & Sennema, 2007) are relatively infrequent and have obtained little evidence of positive carry-over effects of production-based training to perception. The findings discussed above provide little conclusive evidence of the effectiveness of articulatory training. Further research on articulatory training is necessary to determine to what extent improvement in L2 sound production may transfer to improvement in L2 sound perception.

1.1.3.3. Perception vs. production training and cross-modal training effects

Studies assessing cross-modal training effects on perception and production examine whether identification and articulation of L2 speech sounds benefits from training on perception, or viceversa. These studies have often found lack of alignment between perception and production skills.

Catford and Pisoni (1970) compared the effects of auditory and articulatory training on the perception and production of sounds. They found that the articulatory training group performed significantly better than the auditory training group in the production and discrimination of exotic vowels and consonants, although both training methods led to significant improvement in the two domains. Similarly, Leather (1990) observed that Dutch speakers improved their perception and production of Mandarin tones regardless of the type of training approached. Therefore, it was concluded that training in one modality tends to be sufficient to improve learners' performance in the other.

Crucial to our study is the research conducted by Gómez-Lacabex & García-Lecumberri (2007) on the effects of perceptual and production training on English unrounded mid-central schwa /ə/ for Spanish learners of English. Their study constitutes the only work which deals with vowel perception and production training in a FL classroom context. It provided further support to Leather's (1990) and Matthew's (1997) proposals that there is a mutually facilitative relationship between vowel perception and production. In a follow-up study, Gomez-Lacabex, Garcia-Lecumberri, & Cooke (2009) investigated the training and generalization effects of English/ə/ and found significant improvement in perception following both training methods. Importantly, production training proved beneficial in vowel perception (Catford & Pisoni, 1970; Mathews, 1997). Their outcomes provided further evidence for the link between L2 perception and production.

1.1.4. Summary

This section has addressed several main issues in speech perception and production research. Two concluding remarks emerge from this review of the literature:

Firstly, **relatively few studies have directly examined the complex relationship between perception and production by assessing the effects of perception- and production-based training.** The variability of methods used to assess perception and production of L2 sounds does not easily yield comparable data sets. And, to make matters worse, difficulties in the interpretation of outcomes come from the fact that the strength of this relationship is dependent on a multiplicity of factors, such as: the different classes of sounds under study (Bohn &Flege, 1990), amount of L2 experience, context effects, the type of elicitation task, and the nature of acoustic cues (e.g. spectral or durational cues when vowel contrasts are concerned). Differences in degree of attainment in perception and production of L2 learners have also been attributed to individual or cognitive factors (Strange, 1992), phonetic talent (Rota & Reiterer, 2009), motivation and learning aptitude (Bongaerts, van Summeren, Planken, & Schils, 1997; Moyer, 1999, 2004) or cognitive capacity such as phonological short-term memory (PSTM; Mackey, Meador, & Flege, 2001; Cerviño & Mora, 2011) attention control (Guion & Pedersen, 2007; Isaacs &Trofimovich, 2011; Segalowitz, 1997), and inhibition (Lev-Ari & Peperkamp, 2013, 2014; Darcy, Mora & Daidone, 2016) which are out of the scope of this study.

Secondly, research on the effects of training has served as a productive testing ground for general principles of learning and claims about adult neural plasticity, that L2 speech learning is attainable for late learners and L2 perception and production can improve with L2 experience (Flege, 1995) and phonetic training (Aliaga-Garcia, 2016; Pereira & Hazan, 2013; Wang, 2008; Wang et al., 2003). However, **very few training studies have examined the effects of perception and production training on L2 vowel learning, particularly on a full set of vowels (all English monophthongs).** In this study, we will address whether it is possible to obtain benefits on the perception and production of L2 vowels following identification or articulatory training.

1.2. Speech perception models

This section discusses the main factors that may affect the identification of L2 sounds and reviews current models of speech perception which justify the use of perception training based on L2 sound identification in the current study as well as the importance of auditory and visual input for L2 vowel learning.

Three current L2 speech perception models provide an account of adult L2 learners' difficulties learning non-native vowel categories: Best's Perceptual Assimilation Model (PAM, Best 1995; PAM-L2, Best & Tyler, 2007), Kuhl's Native Language Magnet model (NLM; Kuhl 1993, 1995; Iverson & Kuhl, 1995, 1996; NML-e, Kuhl et al., 2008) and Flege's Speech Learning Model (SLM; Flege, 1995). Both PAM (Best, 1995) and SLM (Flege, 1995) relate L2 vowel learning difficulties to the (dis)similarity between L1 and L2 vowel systems and the extent to which L2 vowels are assimilated to "similar" L1 vowel categories. NLM predicts different levels of L1-L2 mapping difficulty involving the perception of L2 sounds in terms of L1 categories, and predicting the difficulty with which the target sound is likely to be affected by the L1 magnet.

These models of L2 speech perception postulate that speech perception precedes production and have thus discussed the role of accurate perception on accurate production (Flege 1995; Best, 1995; Escudero, 2005; Best & Tyler, 2007). They are useful to explain findings of previous perceptual training studies concerning the transfer of improvement in perception to production, for consonants (e.g. Bradlow et al., 1997; Yamada et al., 1994; Hazan et al. 2005) and vowels (e.g. Lambacher et al., 2005; Lengeris & Hazan, 2010).

L2 speech perception models assume that in the initial stages of SLA adults perceive L2 sounds through the filter of their L1 (Trubetzkoy, 1939; Flege, 1981; Best, McRoberts & Goodell, 2001; Best & Tyler, 2007), which leads to inaccurate perception and production (Escudero, 2002; Flege, 1995). Moreover, late learners are expected to show stronger effects of L1 experience than early learners (MacKay et al., 2001; Piske et al., 2002). However, not all L2 phonetic segments are equally difficult for late learners. The learnability of L2 phonetic segments depends on the perceived phonetic distance among them and how they are mapped onto segments in the learner's L1 phonetic inventory in terms of degree of acoustic and articulatory similarity.

1.2.1. *Perceptual Assimilation Model for L2 learners*

The Perceptual Assimilation Model (PAM; Best, 1995) for naïve listeners was extended to the Perceptual Assimilation Model for L2 learners (PAM-L2; Best & Tyler, 2007) to make predictions about success in L2 perceptual learning. The present study tests participants on their identification and discrimination of L2 vowels before and after phonetic training. Therefore, the PAM-L2 framework is discussed in as much depth as it is necessary to characterize different degrees of success in L2 vowel learning.

According to PAM, differences in perceptual accuracy are determined by the extent to which L2 vowels are either assimilated to L1 vowel categories or categorized as a new category. Two L2 vowels assimilating to a single L1 vowel category are expected to be among the most difficult vowel contrasts to discriminate, especially if they are perceived to be equally good instances of the L1 vowel category. However, two vowels assimilating to two different L1 vowel categories will be discriminated more accurately than two vowels assimilating to a single vowel category (single-category assimilation).

PAM generates predictions of relative perceptual discriminability of L2 vowels on a range of L1 assimilation possibilities (See Table 1.1). Non-native vowel sounds are assimilated to the native vowel sounds that learners perceive as acoustically and articulatorily closer by following five different assimilation patterns resulting in various degrees of discrimination.

The different types of assimilation patterns for unfamiliar L2 sounds into the L1 phonological space predicted by PAM/PAM-L2 are the following (Table 1.1):

1. **Two-Category Assimilation** (TC), when the listener perceives two L2 sounds in terms of two different L1 categories (Excellent Discrimination).

2. **Category-Goodness Assimilation** (CG), when two L2 sounds map onto the same category in the L1 sound system, but one is a better fit than the other (Moderate to GoodDiscrimination).

3. **Single-Category Assimilation** (SC), when two L2 sounds are heard as the same L1 phoneme, but neither is particularly „better" than the other (Poor Discrimination).

4. **Uncategorized-Categorized assimilation** (UC), involves two L2 sounds, one is assimilated to a native category and the other one is uncategorized (Good/Very good Discrimination).

5. **Uncategorized-Uncategorized** (UU), when two non-native sounds are uncategorized. (Variable Discrimination, from poor to excellent).

Perceptual Assimilation Patterns	Single-category (SC)	Two-category (TC)	Category-goodness (CG)	Non-categorized (NC)	Categorized-non-categorized (CNC)
Pattern	2 L2 vowels > 1 L1 vowels	2 L2 vowels > 2 L1 vowels	2 L2 vowels (1 is deviant) > 1 L1 vowel	2 L2 vowels > 0 L1	2 L2 vowels: 1 L2 > 0 L1; 1 L2 > 1 L1 vowel
Discrimination	Poor	Excellent	Moderate to very good	Poor to very good	Very good

Table 1.1. PAM-L2's Perceptual Assimilation patterns and predicted degrees of difficulty in non-native vowel discrimination.

Following PAM-L2 (Bet & Tyler, 2007), improvement obtained in L2 vowel perception and production may imply that L2-L1 vowel assimilation patters can be modified with increased L2 experience in the form of phonetic training. Bearing in mind that discrimination is typically poor when two L2 vowels assimilate to a single L1 vowel category, the expectation in this study is that, overall, the assimilation patterns and discriminability of L2 vowel contrasts under study (/iːɪeɜːæʌɑːɒɔːʊuː/) will shift and become target-like through identification and/or articulatory audiovisual training.

1.2.2. *Speech Learning Model*

Like PAM (Best, 1995) and PAM-L2 (Best & Tyler, 2007), Flege's (1995) Speech Learning Model (SLM) is driven by consideration of how the target vowels are assimilated to L1 vowel categories, with discrimination difficulties ensuing when L2 vowels are identified with one particular L1 vowel category. In contrast with Best's model, which generates predictions of perceptual discriminability on a range of L1 assimilation patterns, Flege's model is primarily concerned with L2 vowels (or sounds) being perceived as a *new* or *similar* instance of an L1 vowel.

Flege (1987) proposed the notion of *equivalence classification* to account for the fact that some non-native vowels are more readily assimilated to native categories than others by L2 learners. In other words, certain L2 vowels are phonetically sufficiently different from their nearest L1 targets to be perceived as "new" or "foreign", whereas others are sufficiently close to L1 targets to be classified as "similar" to some L1 category. However, the phonetic criteria which determines whether an L2 vowel is classified as "similar" or as "new" were based purely on perceptual similarity judgments.

The SLM (Flege, 1995) proposes that late learners will be less successful in acquiring *new phonetic categories* if L2 vowels are perceived as "similar" or "*equivalence-classified*" to L1 vowels by perceptual similarity. However, new categories will eventually be formed for L2 vowels that are perceived as distant ("different") from any L1 segment. While this perceived L1-L2 phonetic distance is the essential condition for an L2 category to be established, SLM also predicts that learning *new* L2 vowel categories becomes somehow easier when these fall into "uncrowded" areas of the L1 vowel space and do not overlap with L1 categories.

Another point of divergence from PAM-L2 is that the SLM is concerned with ultimate attainment in L2 pronunciation. The SLM was designed to explain the difficulties of L2 learners in sound production (e.g. FA), as well as changes in production with increased L2 experience. The SLM has shown the positive effect of L2 experience (i.e. LOR, amount of L2 use, AOL) and quantity and quality of input on L2 perception and production. It declares that misperception is the major reason for inaccurate production (*perception leads production* hypothesis, see Section 1.1.2). It states that the possibility to acquire new sounds is continuously present in spite of AOL effects and perceived cross-linguistic phonetic distance. It hypothesizes that "*the greater the perceived difference of an L2 sound from the closest L1 sound, the more likely it is that a separate category will be established for the L2 sound*" (Flege, 1995, p. 264).

It is noteworthy that the SLM can be used to predict phonetic learning in the present study. At pre-test, L2 vowels may be identified in terms of an L1 allophone. As L2 learners gain experience through training, they may become able to discern phonetic differences between L1-L2 and L2-L2 vowel pairs and new phonetic category representations will be eventually established for L2 vowels. In fact, previous studies reporting the effects of perceptual training on L2 sounds perception and production also reported the modification in the adult learners' perceptual patterns after training and improvement in production (Aliaga-Garcia, 2016; Pereira & Hazan, 2013; Wang, 208; Wang et al., 2003; Rato & Rauber). This study aims at assessing the effectiveness of auditory and articulatory training in helping learners establish separate L2 phonetic categories for each of the English stressed vowel monophthongs /iːɪeɜːæʌɑːɒɔːʊuː/ in perception and production..

This hypothesis is also closely related to the perception model that will be described next, the Native Language Magnet model proposed by Kuhl and colleagues.

1.2.3. Native Language Magnet

The Native Language Magnet (NLM) model sees L2 sound perception as L1-L2 mappings resulting in a set of abstract phonetic categories stored in memory (Iverson & Kuhl, 1995, 1996). The NLM theory was initially developed to explain infant speech perception (Kuhl, 1991, 1992, 1994). However, it has also been used to explain difficulties in adult L2 sound perception, and to account for the fact that L2 vowel perception is constrained by the *native language neural commitment* (NLNC), arguing that the native perceptual experience affects the speaker's subsequent abilities to learn the phonetic scheme of an L2 during the exposure to new language input. From the point of view of the NLM, linguistic experience interferes with phonetic learning of a new language in adulthood (e.g. Flege 1995; Mc Candliss et al., 2002; Iverson et al., 2003; Zhang et al., 2005).

The expanded version NML-e by Kuhl et al. (2008) holds that, by 6 months of age, L1 speakers start developing a sound map in their brains with mental representations of sounds. The native vowel system is organized in terms of *prototypes* or best exemplars that begin to function early in childhood attracting other similar members of the category (Kuhl, 1991). These prototypes remain in long-term memory as perceptual *magnets* attracting perceptually neighbouring stimuli (*perceptual magnet effect*), and reducing FL phonetic abilities (e.g. confusable similar L2 vowel sounds). Therefore, the difficulty in discriminating two L2 vowels depends on their proximity to a L1 vowel magnet which (perceptually) attracts and assimilates surrounding new (target) vowels, which are in turn interpreted as *near-prototypes*. For the NLM-e, once a vowel category exists in memory, *"it functions like a perceptual magnet for other sounds in the category"* (Kuhl, 2000, p. 11853). L2 learners fail to develop prototypes for certain L2 vowels but 'distorted' perception does not impede modification or learning, which makes it a powerful argument for the use of perceptual training to help learners improve vowel discrimination and eventually establish new vowel categories.

The NLM-e model attributes lack of accuracy in perception to the L2 learners' strong neural commitment to the native language, which interferes with later language learning. From an NLM-e perspective, training will also help learners perceive differences between similar/new L2 vowels by presenting by making contrasting cues salient to them. As learners may attend to the wrong phonetic cue because they are insensitive to those acoustic cues that are relevant in the categorization of L2 sounds (Flege, Bohn, & Jang, 1997; Iverson et al., 2003), phonetic training should help learners guide their attention towards the relevant cues making their L2 vowel categorization easier.

1.2.4. Summary

To sum up, the three models of L2 speech perception reviewed above suggest that L2 contrasts are not uniformly poorly perceived by learners. Instead, the difficulty with which a particular L2 contrast is perceived depends on the relationship between the L1 and

L2 vowel inventories. The models support the view that dissimilar sounds are produced and acquired more easily than similar sounds and cross-language similarity has detrimental effects on L2 production which increase with AOL.

From the models discussed above, it may be inferred that accurate L2 vowel perception is possible provided L2 learners detect the differences between L2 sounds and L1 vowel categories and build up new L2 vowel categories based on them. PAM and NLM-e do not explicitly mention that L2 learners should also weigh phonetic cues accurately for new L2 vowel categories to be established. In PAM, L2 vowels are perceived in terms of their phonetic *gestural* similarity to L1 categories. Therefore, the greater the difference in *category-goodness* between contrastive L2 sounds mapped onto a single L1 category, the more discriminable the contrast will be and the more likely category formation will be. Like PAM, predictions based on the NLM-e for the perception of L2 vowel contrasts are a product of the perceived similarity between the L2 vowels and their L1 *prototype*.

1.2.5. Cue Weighting in L2 vowel perception

SLM differs from PAM and the NLM-e model in that SLM allows for the notion of language-specific cue-weighting. The SLM states that an L2 learner may weight acoustic cues differently to a NS in perception and production (Flege, 1995). Research has shown that English NSs rely more on spectral than duration cues (Hillenbrand et al., 2000) in identifying vowel sounds, whereas Catalan and Spanish learners rely predominantly on duration (Bohn, 1995; Escudero & Boersma, 2004; Escudero, 2006; Koundarova & Francis, 2008). One of the questions this study seeks to answer is whether target-like cue weighting (i.e. native-like use of spectral and duration cues) in vowel perception is best promoted by training.

Bohn (1995) proposes the availability of duration as a cue even to learners whose L1s have no duration contrasts, as it is the case of bilingual Catalan-Spanish learners of English in our study There is also evidence from several studies that L2 learners are insensitive to certain acoustic differences present in L2 contrasts, relying on auditory cues which are redundant or secondary for native speakers of English (Flege, Bohn, & Jang, 1997; Iverson et al., 2003). A great deal of research has shown that Spanish (Bohn, 1995; Bohn & Flege, 1990), Mandarin (Flege et al., 1997), Portuguese (Raubert et al., 2005), Polish (Bogacka, 2004), Japanese (Morrison, 2002), Russian (Koundarova & Francis, 2008) and Catalan (Cebrian, 2006) learners of English contrast English /iː/ and /ɪ/ mainly, or solely, on the basis of duration (e.g. long /iː/ vs. short /ɪ/). Interestingly, none of their L1s have a phonologically contrastive vowel duration difference. According to Escudero & Boersma (2004), either mere exposure to English or the universal saliency of duration explains the use of duration cue in L2 vowel perception. Bohn (1995) hypothesized that non-native cue weighting, or the use of secondary cues in L2 vowel perception, appears to be another cause of inaccurate vowel perception. It is hypothesized that learners are misguided by non-native teachers towards discriminating English tense-lax contrasts exclusively on the basis of duration in classrooms all over Catalonia and Spain (i.e. /iː/ is "long" and /ɪ/ is "short") (Aliaga-Garcia, 2007; Flege et al., 2007; Morrison, 2006, 2008; Wang & Munro, 1999).

The majority of cue weighting studies concerning vowel perception and production by Catalan-Spanish speakers have focused their attention on the tense-lax /iː/-/ɪ/ contrast. Two patterns of results can be outlined: (a) overreliance on duration, a cue non-existing in the learners' L1, or (b) reliance on both spectrum and duration.

Bohn (1995) and Flege, Bohn & Jang (1997) obtained mixed results when they investigated the same participants' identification performance on an English /i/-/ɪ/

continua that varied in spectral and duration properties. Following SLM and PAM-2, the *Desensitization Hypothesis* (Bohn, 1995) proposes that overreliance on duration occurs as a compensatory strategy, when single-category assimilation has linguistically *desensitized* learners for the perception of spectral cues (Bohn & Flege, 1990; Bohn, 1995; Cebrian, 2006; Escudero & Boersma, 2004; Morrison, 2008). However, a few methodological problems make Bohn's hypothesis inconclusive. First of all, the fact that the stimuli used did not vary in the same number of steps for spectral cues (eleven steps) and duration cues (three steps) may have biased Spanish speakers towards preferring the temporal cue (the dimension that varied the least) over spectral cues. In other words, the type of stimuli might have enhanced the learners' ability to use duration cues and impaired their ability to use spectral cues. Secondly, Bohn (1995) did not assess the effect of L2 experience on cue weighting.

In support of Bohn's hypothesis, several studies have confirmed that Spanish and bilingual Catalan-Spanish listeners rely on temporal cues more than native English speakers do in vowel discrimination, regardless of amount of L2 experience (Aliaga-Garcia & Mora, 2007, 2009; Cebrian, 2006, 2007; Cerviño & Mora, 2009; Escudero & Boersma, 2004; Escudero, 2006; Mora & Fullana, 2007; Morrison, 2008).

However, other researchers (Flege et al., 1997; Morrison, 2006) suggest that the degree of reliance on spectral and temporal cues by Spanish listeners is similar to that of English speakers. Morrison (2006), for instance, suggested that the '*non-use of spectral cues*' argument may have been exaggerated. Similarly, Flege et al. (1997) showed that, whereas American English (AE) listeners have a preference for the spectral cues, Spanish listeners use the two dimensions equally.

Escudero (2000) hypothesized that, in line with previous studies (Bohn, 1995; Flege et al., 1997), Spanish speakers perceived English /iː/-/ɪ/ using duration as the only or primary cue. Escudero (2000) attributed inter-subject variation in cue weighting to the dialectal heterogeneity of the Spanish population in her study. That allowed her to conclude that greater reliance on duration or spectral cues was speaker-dependent. Importantly, Escudero & Boersma (2004) identified six different perception patterns on the basis of cue weighting for Spanish listeners:

(1) listeners who identified /iː/-/ɪ/ *using exclusively* duration;

(2) listeners who identified /iː/-/ɪ/ using *mainly* duration;

(3) listeners who identified /iː/-/ɪ/ using duration and spectrum;

(4) listeners who identified /iː/-/ɪ/ using spectrum and duration

(5) listeners who identified /iː/-/ɪ/ using *mainly* spectrum;

(6) listeners who identified /iː/-/ɪ/ using *exclusively* spectrum.

Whereas Scottish English and SBE speakers produced a relatively large spectral difference and no or little duration difference between /iː/ and /ɪ/, Spanish learners of English made significantly less use of spectral cues and/or significantly more use of duration cues. . An explanation offered by Escudero & Boersma (2004) is that it is easier to create a new category distinction in a previously unused (duration) dimension, than it is to split a category in an existing (spectral) dimension. It has been argued that learners tend to rely more on duration than spectral cues when they are faced with a difficult L2 contrast, as it is English /iː/ and /ɪ/ for Catalan-Spanish speakers, who only have a single L1 vowel category in the F1/F2 vowel space occupies by the two English vowels (Escudero, 2005; Cebrian, 2006)

Both cross-linguistic and developmental variation is also attested in L2 cue weighting studies. Morrison (2002) showed that cue weighting changes as L2 experience increases. In Morrison's study, Spanish learners of English showed a very poor reliance on duration after their first month in Canada; after six months, they showed a reasonably good reliance on spectrum or duration, but not on both. In line with Bohn's (1995) findings, Morrison's (2005) reanalysis of Escudero & Boersma (2004) presented a similar pattern of results for Spanish learners, in spite of choosing a different metric for cue weighting (logistic regression coefficients) than Flege et al. (2007) and Escudero & Boersma (2004) did.

Morrison's (2008, 2009) results for Spanish speakers' perception of Canadian-English /bɪt/-/bɪd/-/bit/-/bid/ continua were largely consistent with the developmental stages described for Spanish speakers' perception of English /iː/ and /ɪ/ by Escudero (2000). But Morrison (2008, 2009) hypothesized there was an additional earlier *multidimensional-category-goodness-difference* stage besides the duration-based stage (Escudero, 2000). During that stage, three quarters of participants in the study labelled English vowels that sounded as good exemplars of Spanish /i/ (with shorter duration, lower F1 and higher F2) as "English /ɪ/" and labelled poor examples (with either longer duration, higher F1 or lower F2) as "English /iː/" (*indirect developmental path*) (Morrison, 2009: 458). Only a quarter of the participants chose an opposite direction of labelling, associating good exemplars of Spanish /i/with English /iː/ on the basis of duration cues (*direct developmental path*) (Morrison, 2009: 459). Long and short vowels were labelled as English tense /iː/and lax /ɪ/, respectively.

The Spanish listeners' preference for duration cues was taken as a secondary developmental stage emerging from an earlier stage when both spectral and duration cues are used rather than an initial strategy for /iː/-/ɪ/ discrimination. According to Morrison (2008, 2009), the multidimensional-category-goodness-difference assimilation stage, followed by the duration-based stage, leads to the following stages, characterized by partially and fully target-like perception of English /iː/and /ɪ/.

The studies reviewed above show differences in cue weighting between native speakers and L2 learners of English. They provide different explanations for why some L2 learners may be unable to perceive spectral differences in certain parts of the L2 vowel space and why duration cues are more acoustically salient to them. The present study hypothesizes that target-like cue weighting can be promoted through identification and articulatory phonetic training.

There is evidence from recent phonetic training studies that L2 cue weighting can indeed be modified through HVPT. Ylinen et al. (2010) showed that after training Finnish speakers of English were able to use spectral cues more reliably and depended less on duration cues in the identification of /iː/ and /ɪ/. Moreover, electrophysiological data showed improved processing of spectral cues as a result of training (See section 6.5).

Very little research has been carried out on L2 vowel cue weighting involving Catalan speakers. Fullana-Rivera and MacKay (2003) evaluated the production of English /iː/–/ɪ/ by eight different groups of Catalan learners of English (aged 11–28) and found no significant effect of either AOL or L2 experience on cue weighting. Vowel perception was examined Rallo & Romero (2012), who observed some effect of L2 learning on the L1. Rallo's finding is in accordance with Flege's (1987) finding that the influence of a learner's L1 and L2 is bidirectional. Rallo's findings also support Flege's (1995) hypothesis that the acquisition of new vowel categories may result in the loss of perceptual sensitivity to some native categories. Cebrian (2002) found a negligible effect of L2 experience as learners of English in a FI or an immersion setting showed strong reliance on duration in the discrimination of /iː/-/ɪ/ contrast. Some of the findings above suggest that Catalan-Spanish speakers' overreliance on the temporal cues may have prevented accurate vowel perception and production.

1.2.5. Cue weighting in the current study

This study also tests learners' use of spectral versus temporal cues in the categorization of synthesized vowel stimuli (e.g. vowels with manipulated duration) (ID2-

task2). It is hypothesized that Catalan-Spanish learners of English will differ from native English speakers in their perceptual weighting of the duration and spectral dimensions for /iː/-/ɪ/ discrimination. A review of the literature suggests that one aspect which deserves further exploration is the relationship between L2 cue weighting and training. The majority of studies (Cebrian, 2006; Escudero et al, 2009; Koundarova & Francis, 2008; Morrison, 2008) have not attested a significant relationship between cue weighting and L2 speech learning. Only a few studies have investigated and reported a significant effect of training on L2 cue weighthing (Iverson et al., 2005; Ylinen et al., 2010).

One question of interest in this study is whether identification and articulatory HVPT using audiovisual stimuli can be effective in modifying non-native cue weighting (characterized by stronger reliance on duration than quality) in L2 perception and production. It is hypothesized that HVPT will (1) help learners overcome interference from overreliance on duration in L2 vowel perception and (2) shift their attention to acoustic dimensions other than duration (e.g. weighting spectral cues over temporal cues, in the direction of what native English speakers do).

1.2.6. Audiovisual (AV) perception in SLA

It is well known from consonant studies of L1 and L2 audiovisual perception that visual cues contribute to the perception of place and manner features when acoustic cues are insufficient (Ortega-Llevaria, 2001) or the speech signal is degraded, i.e. speech in noise (Hazan et al., 2010; Banks et al. 2015). There have been, however, relatively few studies assessing the effect of audiovisual training on the perception of L2 speech sounds (e.g. Akahane-Yamada et al., 1997; Hardison, 2003; Ortega-Llebaria et al., 2001).

The fact that both PAM (Best, 1995) and NLM-e (Kuhl et al., 2008) hold that speech perception is multimodal in nature is particularly relevant to the present study. The NLM, unlike the SLM (Flege, 1995), argues that auditory and visual information help listeners establish their L1 phonetic space very early in life. PAM also differs from SLM in explicitly asserting a *gestural* basis for perceived similarities and discrepancies between L1 and L2 speech sounds. Studies of cross-modal speech perception have shown that both adults and infants integrate auditory and visual cues into a gesturally coherent perfect (Best, 1995).

Audiovisual integration in speech perception became popular when McGurk & MacDonald (1976) investigated the audiovisual perception of audio CV stimuli with /p, b, k, g/ preceding /a/ (e.g. /baba/) dubbed with different consonants. The responses to stimuli with conflicting cues (e.g. auditory /papa/ and visual /gaga/) followed two patterns: responses either (a) did not correspond to either the auditory or the visual stimuli (e.g. /dada/) or (b) were a combination of auditory and visual stimuli (e.g. /gaba/). These would suggest that auditory and visual modalities are integrated and visual cues play indeed a crucial role in speech perception. Moreover, the McGurk-effect (or audiovisual integration) appears to be very robust since it persists even when the audiovisual components are from speakers differing in sex (Green et al., 1991), with different facial configurations (e.g. inverted mouth and eyes) (Hietanen, Manninen, Sams, & Surakka, 2001) or without detailed information about their face (Rosenblum & Saldaña, 1996). McGurk & MacDonald supplied revolutionary evidence in the late 1970s of a previously unrecognized influence of visual information on speech perception, at a moment when auditory-based theories of speech perception were unable to explain this. Importantly, their findings also illustrate that auditory perception in adults is more influenced by visual input than is that of children.

Evidence of audiovisual benefit in L2 perception has been shown in the literature with respect to the perception of English consonants by Spanish and Japanese speakers (Akahane-Yamada et al., 1997; Hardison, 2000, 2003; Hazan et al., 2006; Hazan et al., 2002; Ortega-Llebaria et al., 2002), and English CV syllables by Japanese, Korean, Spanish and

Malay speakers (Hardison, 1999; Hardison et al., 2006). Based on these findings, it can be concluded that the role of visual cues in L2 speech perception varies as a function of the degree of visual salience of the target sounds under study.

1.2.7. Perceptual training using audiovisual stimuli

The potential of visual cues in L2 speech learning and perceptual training has received relatively little attention for consonants and even less for vowels.

Akahane-Yamada et al. (1998) suggested that sensitivity to visual cues for L2 consonant contrasts (/v/-/b/-/p/ labial/labiodental contrast and /l/-/r/ contrast) could be enhanced via audiovisual perceptual training. Audiovisual training led to greater improvement in auditory perception and lip-reading, even for those sounds with relatively low visual salience (/l/-/r/). Davis and Kim (1999) reported similar audiovisual benefits and David and Kim (2006) suggested that the talker's mouth and jaw represented a direct source of information about segmental properties of speech and increased intelligibility. The most extensive audiovisual training study for consonants was carried out by Hardison (2000, 2003), who found that audiovisual training was more effective than auditory training in improving the identification and the production of /r/-/l/ by Japanese and Korean learners of English. Prior to training, Hardison had also found evidence of better performance when stimuli were presented audiovisually than auditorily. This sharply contrasts with the little evidence of audiovisual benefits reported in two studies carried out by Ortega-Llebaria et al. (2001, 2002). In these studies, neither Spanish nor Japanese learners of English benefited significantly from visual cues in the perception of English /b/-/v/, /p/-/v/ or /r/-/l/. The lack of audiovisual benefits was replicated by Sennema et al. (2003), who just found an effect of audiovisual training on the learner's sensitivity to visual cues in /r/-/l/ perception. Hazan et al. (2006) investigated whether training effectiveness could be increased by incorporating visual cues to the perceptual training of consonants. They concluded that, while sensitivity to visual cues is acquired in the process of establishing new consonant categories, it may not be present in the early stages of acquisition, at least for certain consonant contrasts.

However, to the best of my knowledge, very few speech perception or training studies (e.g. Ortega-Llebaria et al., 2002) have investigated the role of visual cues for vowel learning. Audiovisual perception of vowels as such has not yet been studied in experiments analogous to those of McGurk and MacDonald (1976). Evidence of audiovisual integration in vowel perception comes only from Summerfield and McGrath (1984). They presented [bVd] syllables with [i], [ɑ] and [u] faces to native English listeners and observed that the phoneme boundaries in the auditory vowel space were moved closer towards the position of the vowels presented visually. Summerfield and McGrath's results suggested that visual input biased vowel perception.

Lisker and Rossi (1992) investigated phonetically trained subjects who were asked to identify French and non-French vowels (presented in auditory/visual conditions, and congruent/incongruent audiovisual conditions) as rounded or unrounded. Although Lisker and Rossi did not test vowel categorization, their results showed that visual cues such as lip-rounding increased the probability of identifying a vowel as rounded rather than unrounded.

Visual saliency of lip rounding was also reported in studies dealing with audiovisual vowel perception by native speakers of Swedish. For instance, Öhrström and Traunmüller (2004) presented auditory, visual and audiovisual syllables with and without conflicting vowel cues (/i y e ø/) and found that roundedness was easier to perceive visually but openness was better perceived auditorily. Lidestan (2008) compared auditory and audiovisual discrimination of vowel duration and suggested that the use of visual cues might drive audiovisual perception of vowel duration in L1-Swedish. These results agree with those obtained for consonants in that they demonstrate a dominant role of visual input in the perception of labial features (e.g. subjects made significantly fewer errors in visual than in acoustic vowel perception). These studies, however, do not make any claims as to which population of speakers the observed results can be generalized. Lip rounding proved to be distinctive among French and Swedish front vowels but is not an independent distinctive feature within the vowel system of English.

Only a few studies have explored the effectiveness of audiovisual vowel training on Spanish learners of English. The first attempt to train Spanish learners on vowel production with visual feedback dates back to the late 1980s, when Flege (1989) stated that *"little systematic information [existed] concerning the use of visual information to train vowel production"* (Flege, 1989, p. 371). At the time Flege was exploring the use of visual-oriented production training, there had only been a few studies reporting the effectiveness of visual feedback for hearing-impaired individuals in the early and mid 1980s (Fletcher & Hasegawa, 1983; Osberger 1987; Povel & Wansink, 1986).

Flege (1989) administered three ten-minute training sessions for /iː/-/ɪ/ and /æ/-/ʌ/ to one Spanish speaker. The only positive outcome was found for /iː/ and /ɪ/, which were produced with sizeable articulatory differences after training. This served as preliminary evidence that 'visual articulatory modelling and shaping' stimulated through visual feedback can be effective in vowel training. However, the positive contribution of audiovisual feedback cannot be interpreted unambiguously because the trainee produced /æ/-/ʌ/ less authentically at post-test. Moreover, the fact that two untrained vowels were also spoken more acceptably at post-test suggests that better articulation of /iː/-/ɪ/ could have been simply due to the motor task involving repetition induced by training.

More recently, Ortega-Llebaria et al. (2001) investigated the effectiveness of visual cues in improving the perception of British English vowels (*bad, bed, bid, bead, bud, bard, bared, bide, boughed*) by Spanish learners of English. They found that, at post-test, English vowels remained perceptually difficult for Spanish speakers and vowel confusions were not completely disambiguated by the use of visual cues in auditory training. The audiovisual benefit was only significant for those vowels with very different lip gestures (*bed* vs. *bud*, *bared* vs. *bead*). These were the only cases where vowel confusions were completely eliminated in the audiovisual condition. Consequently, the study indicates that visual cues may be weighted differently in vowel training depending on the vowel contrast. It suggests that audiovisual training could only be effective in improving vowel perception when vowels with very dissimilar lip configurations (e.g. with very distinctive visual features) are paired and presented to learners of English.

The reason why L2 learners in some studies do not seem to be sensitive to the visual cues remainsunclear. Further research on the possible aid of visual features invowel learning is needed, as is research on the effectiveness of audiovisual training methods in order to help learners make better use of visual cues.

The multimodal nature of speech perception plays a crucial role in our study. The HVPT methods compared include audiovisual vowel stimuli in a high variety of CVC contexts. This means that visual information about lip-shape, tongue position and mouth openness, and vowel duration are expected to exert an important influence on vowel learning. More specifically, our study aims at testing the effectiveness of audiovisual training on adults' perception and production of L2 vowels varying in spectral cues (tongue height and frontness, and degree of mouth openness and lip rounding) and duration cues. Although the visual cues to vowel identification may not be as salient as those to consonant identification (e.g. features related to place of articulation), that L2 learners may be expected to be strongly influenced by lip-rounding, mouth openness and vowel duration features. No training studies to date have assessed how sensitive L2 learners are to duration cues when they exposed to only-visual vowel stimuli. The visual saliency of the duration cue has not been explored. There are no training studies, to the best of my knowledge, which have assessed whether L2 learners can be trained to visually attend to the duration cue in vowel perception. In this study the duration cue is assumed to be one more informative visual feature.

1.3. Speech production models

Several theoretical proposals have been put forward in order to adapt existing models of L1 speech production to L2 speech production. Some researchers contend that L2 speech production is constrained by a critical period deriving from a loss of neural plasticity, making L2 speech production resistant to language experience and training (Scovel, 1969; Patkowski, 1990; DeKeyser, 2000). The critical period hypothesis, as applied to the articulatory mechanisms underlying pronunciation (Scovel, 1988; Ioup, 2005), is often cited in explaining why L2 learners have difficulties producing L2 sounds accurately (Flege et al., 1997).

Even researchers challenging the critical period hypothesis would admit that control of pronunciation may be one L2 dimension for which a critical period exists (Walsh & Diller, 1978). Nonetheless, other research suggests that if sufficient L2 native speaker input is supplied, certain L2 sounds will ultimately be produced with native-like accuracy (Best & Strange, 1992; Kuhl & Meltzoff, 1996; Kuhl, 2000), and that the capacity to learn new sounds remains intact over the life span (Flege, 1981, 1995; Flege & Liu, 2001; Flege & Mackay, 2004). Many speakers are able to accurately produce L2 speech sounds that do not exist in their L1 even though their initial L2 exposure took place in adulthood, supposedly after a critical period. According to the SLM (Flege, 1995) accurate production of new L2 sounds is possible if they are dissimilar enough from the closest L1 sound to be able to avoid the blocking of category formation, which is necessary for accurate speech production (Flege, 1991). L2 learners' difficulties with the articulation of L2 sounds have not received much attention in the L2 speech learning and training literature, in contrast to the prevailing focus on how non-native perception limits accuracy in L2 speech perception and production (Flege & Hillenbrand, 1984), leading to L2 sound substitutions (Iverson et al., 2003), and a foreign accent (Flege, 1992; Flege, Munro, & Mackay, 1995; Flege, Bohn, & Jang, 1997).

However, some neuroimaging studies (e.g. Raizada et al., 2010) suggest that L2 learners' inaccurate production of L2 sounds may be due to difficulties in articulating these sounds rather than difficulties in discriminating them. This means that causes other than the passing of a critical period (e.g. L1 use) or L2 experience must have a played a part in speakers' inaccurate productions of L2 speech sounds.

This section focuses on the speech production models that have served as the theoretical basis of the articulatory training method designed for this study. It also reviews previous studies that emphasize the need of imitation tasks to develop the motor commands required for accurate L2 production and the correction of articulation errors during training. Our articulatory training method is based on the relationship between the control of oral-motor movements necessary for accurate L2 production and self-monitoring, and it involves self-perception and self-correction.

1.3.1. Articulation: Levelt's, De Bot's and Kormos' models of speech production

Levelt's modular model of speech production (1989, 1993; adapted by de Bot, 1992; Levelt et al., 1999) was originally developed for monolingual speech production. In Levelt's model, the speech production system includes five autonomous components: conceptual preparation, grammatical encoding, morpho-phonological encoding, phonetic encoding and articulation (See Figure 2). However, the model only accounts for the part of the speech production process when, after drawing on the gestural scores needed to produce a word (phonetic encoding), the articulator converts the gestural score into articulatory movements in order to produce speech. In short, articulation is listed as a simple motor output (Indefrey & Levelt, 2004) consisting of a computational neural system that controls a motor system (lungs, larynx and vocal tract).

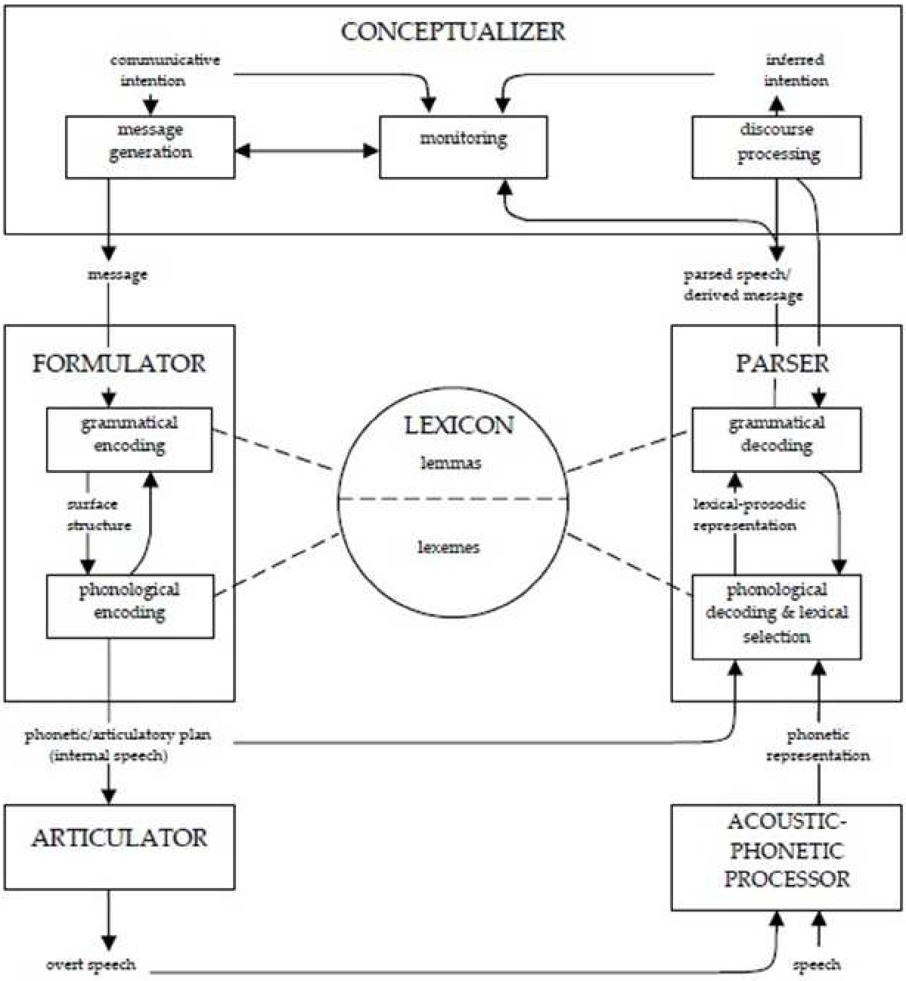


Figure 1.1. Levelt's (1993) model of language production, where the articulator component is especially relevant for the present study.

De Bot (1992) proposed a bilingual version of Levelt's model by extending the articulator system. According to De Bot's extended model, L2 speech is accented because

L2 speakers possess a single articulatory system for the two languages (cf. *subsystems hypothesis*: Paradis, 2001). The use of a single articulatory system by multiple languages would explain L1 interference in L2 speech production and, consequently, accentedness. Problems in L2 speech production arise from the fact that, first of all, the learner's phonological knowledge of the L2 is, obviously, not as broad as that of the L1. Secondly, the degree of automaticity in speech processing is lower in the L2. De Bot assumes that languages are accessed in parallel and accounts for inaccurate pronunciation in terms of phonological interference. The less proficient the learner is, the more the learner will struggle to articulate L2 sounds accurately because there will be more competition between the available L1 and L2 words.

More information about the source of difficulties in L2 pronunciation is provided in Kormos' integrated model of L2 speech production (1999). According to Kormos, both L1 and L2 phonemes are stored in the same module and common phonological features are shared in long-term memory (Poulisse, 1999). Several problems in L2 speech production may be attributed to resource deficits, particularly limited attention. Due to processing time pressure, less attentional resources would also be available during self-monitoring. Kormos' bilingual account assumes that L2 production is relatively slow and inaccurate production comes from the fact that some articulatory gestures specific to L2 sounds have not been automatized yet. In general, L2 production tends to be more hesitant, the rate of speech may be slower, and the degree of automaticity in accomplishing the articulatory targets required for the production of L2 sounds may vary according to the L2 speaker's proficiency. Poulisse (1997) also notes that L2 speakers produce twice as many slips of the tongue in the L2 than in the L1.

1.3.2. Self-monitoring

Oral gestures necessary for producing L1 speech sounds are highly "over-learned" and automatic and their mental representations match the auditory input received in a given language (Simmons et al., 2011). However, in L2 speech production, oral-motor movements are relatively new and do not match the learner's representations. Therefore, unfamiliar movements require greater engagement of motor-sensory feedback systems (Wilson & Iacoboni, 2006). Ludlow & Cikaoja (1998) suggested that auditory feedback is crucial to the process of modifying the speech motor-control system, in order to be able to produce accurate speech sounds in the L2.

Kormos (2000, 2006) argued that Levelt had provided the best account of monitoring (see Figure 2) so far, assuming that modifications of the existing articulatory-acoustic relationships are necessary to produce L2 speech sounds accurately. Based on theories of monitoring and spreading activation (Dell, 1986), Levelt elaborated his own *perceptual loop theory* with three monitor loops of direct feedback channels. This study will focus on the third loop, which monitors articulation.

Self-monitoring is an essential component of speech production during which the speaker exerts some degree of output control by detecting articulatory errors in overt speech (self-perception) and producing subsequent self-repairs. This feedback system allows the speaker to evaluate the output of the formulator.

Levelt's model and its account of monitoring provided the framework for describing one main features of the articulatory training in this study: as learners perceive an error in the articulation of an L2 sound, the learners identify this as an alarm signal that triggers the speech production mechanism for a repair. In such cases, several options are available to speakers, they can either ignore the mistake or correct it. According to Levelt's model, the monitor may be interpreted as a stage in L2 speech production that allows speakers to compare their production and what would be correct according to their phonological knowledge.

Models of monitoring have also been put to test in SLA research (Kormos, 1999, 2000; van Hest, 1996) and suggested that self-repair contributes to pronunciation learning by making L2 learners aware of their limitations. Kormos and colleagues (Dörnyei and Kormos, 1998; Dörnyei and Scott, 1997) argued that Levelt's perceptual loop theory of L1 monitoring does not explain that a lot of errors in L2 pronunciation go unnoticed or undetected by L2 learners due to attentional deficits and limited-resources available during self-monitoring. Detection of errors in pronunciation is a requirement for phonological repairs to take place during motoring.

1.3.3. Imitation

In this study, we propose one type of articulatory training based on a vowel imitation task, based on the motor theory of speech perception (Liberman, Harris, Hoffman, & Griffith, 1957) and Levelt's model of speech production (1989, 1993; adapted by de Bot, 1992; Levelt et al., 1999), integrating self-monitoring and self-correction (see sections 1.3.1 and 1.3.2).

This articulatory training is based on the imitation of audiovisual vowel stimuli. Articulatory training assumes that audiovisual cues lead to vocal imitation, which triggers the articulatory movements that are necessary to pronounce L2 speech sounds accurately. It holds the view that imitation starts when the perception of audiovisual cues induces control of articulations to produce the target sound in the *imitator*.

The use of imitation in speech perception research has been acknowledged since the late 1950s, when the motor theory of speech perception (Liberman, Harris, Hoffman, & Griffith, 1957) presented evidence of *articulatory gestures* encoded in speech sounds and governed by "sequential motor actions" . The McGurk effect (McGurk & Macdonalds,

1976) presented experimental evidence that non-native perception also involves the perception of *gestures*, that is, the use of audiovisual cues in speech perception (See section 2.3). From the point of view of Articulatory Phonology (Browman & Goldstein, 1986, 1989, 1990), *gestures* are equivalent to simultaneous movements of the various articulator sets of the vocal tract: the lips, tongue, glottis, velum, etc (Browman & Goldstein, 1991). For example, to produce vowel gestures, the lips open at the same time as the tongue body produces a certain degree of constriction (e.g., a constriction at the hard palate for /i/, or in the pharynx for /a/). The tongue body is the major organ for the production of gestures, with the lips playing a minor role.

Imitation is deeply linked to L1 speech perception and production (Howard & Messum, 2007; Kuhl & Meltzofff, 1996; Kuhl, 2000; Yoshikawa et al., 2003). These studies suggest that children learn to pronounce through mimicry. Imitation implies a self-supervised matching-to-target process: the child (1) captures the output of others, (2) captures his own output, (3) compares the two and (4) moves his own output towards the adult norm as necessary. During the imitation process, the speaker uses the motor system to imitate a speech sound once the system associates it with a specific articulatory gesture.

Kuhl and Meltzoff (1982, p. 1140) reported that almost half of the 4-month-old infants in their study *"produced sounds that resembled the adult female's vowels"*. This is in line with Studdert-Kennedy's description of speakers as *"attentive listeners"* who learn by establishing a set of *"direct mappings between sound and gesture"* (Studdert-Kennedy, 1988, p. 191). The general assumption about the imitation process is that sound qualities can be copied: *"Infants learn to produce sounds by imitating those produced by another and imitation depends upon the ability to equate sounds produced by others with ones infants themselves produce"* (Kuhl, 1989).

There is evidence, for example, that infants only a few weeks old attempting to mimic the vocalizations of adults and 20-week old infants can imitate the point vowels [i], [a] and [u] (Kuhl & Meltzofff, 1996). There are also studies suggesting that imitation allows more advanced articulations of L2 sounds and that is a particularly apt method for studying the vowel system of L2 learners (Alivuotila, Hakokari, Savela, Happonen, & Alatonen, 2007; Chistovich et al., 1966; Messum, 2007; Vallabha & Tuller, 2004).

The articulatory training method in the present study adopts the proposals formulated by the NML-e that learning stems from the linkage between sensory and motor experience, whereby stored perceptual representations guide motor imitation and, consequently, the speaker's successive motor approximations to the target sound and gesture (Kuhl & Meltzoff, 1996). Our study was also inspired by previous research using the imitation of synthesized (Chistovich et al., 1966; Repp & Williams, 1985), self-produced (Vallabha & Tuller, 2004) and visual (Stadler et al., 1991)vowel stimuli by adult speakers.

In one of the first imitation studies, Chistovich et al. (1966) asked a phonetically trained Russian speaker to imitate 12 synthesized vowels along an [a-e-i] continuum and found between four and six vowel categories in the subject's productions despite having only three vowel phonemes in her L1. Repp and Williams (1985) asked two phonetically trained speakers to imitate 150ms synthetic vowels along the [æ-i] and [u-i] continua but found large inter-subject differences in their patterns of imitation responses. The production skills of the speakers did not match the stimuli most of the times. Repp and Williams tested speakers on the imitation of their own vowels, which were equidistant along [æ-i] and [u-i] continua. The variability of the self-imitations was remarkably similar to the imitations of synthetic vowels. These results suggested that the L2 learners' accuracy of the productions during imitation may be somehow related to their accuracy in perception.

Stadler et al. (1991) had subjects reproduce the dots located on a grid on a sheet of paper and found that their reproductions were biased towards the corners of the visual field. Following Stadler et al.'s finding, Wildgen (1991) proposed that vowel systems were organized in an analogous manner, with the peripheral vowels [i], [a] and [u] functioning as attractors. This sugests that the perceptual space of vowels is warped and the warping is influenced by L1 vowel categories (Kewley-Port & Atal, 1989; Iverson & Kuhl, 1995). Vallabha & Tuller (2004) also found systematic variation in the speakers' imitation of self-reproduced vowels in /hVd/ contexts. Vowel production was found to be related to the speaker's perceptual abilities.

1.3.4. Summary

The present study holds the view that L2 speech production requires the coordination and movement of the articulators so that the shaping of the vocal tract results in the production of the intended acoustic signal (Nieto-Castanon et al., 2005). This can be promoted and induced through articulatory training. In parallel, feedback sensory systems provide online monitoring so that rapid motor adjustments can be made (Guenther et al., 2006). Additionally, audiovisual inputs are also thought to play a role in feedback monitoring, as demonstrated by the McGurk effect (see section 1.2.6. of this chapter) (Bamiou et al., 2003). This comparison between predicted (audiovisual input) and actual movement of the articulators (output generated by the speaker) allows the rapid detection and self-correction of articulation errors (Guenther et al., 2006; Golfinopoulus et al., 2010), resulting in relatively efficient feedback processing (Eliades & Wang, 2008).

Levelt's production model is based on a long tradition of psycholinguistic research and on robust empirical evidence. It provides a solid framework for building articulatory vowel training and a satisfactory account of the speech production processes involved in learning to articulate L2 vowels. Imitation is relevant to this study as the articulatory training is based on a vowel imitation task. Following exposure to an audiovisual stimulus, imitation is the second stage of the training task, followed by self-monitoring, to allow learners to correct their own mistakes and initiate the speech production mechanism again.

1.4. Perception and production of English vowels by Catalan-Spanish speakers

In this study bilingual Catalan-Spanish learners of English were trained and tested on their perception and production of the whole set of SBE monophthongs (/iːɪeɜːæʌɑːɒɔːʊuː/). This section first describes the differences between the Catalan, Spanish and English vowels systems. Then it explains why the tense-lax vowel contrasts are especially difficult to acquire for this population by referring to some predictions set by speech learning models concerning cue-weighting. Finally, it reports on the findings in immersion and FI settings, and results from training experiments, with special emphasis on what constitutes the focus of this study: the effectiveness of phonetic training in L2 vowel learning.

1.4.1. The vowel systems of Catalan, Spanish and English

Previous research has clearly established that the vowel system of Southern British English (SBE) differs considerably from the Catalan and Spanish systems in size, as well as in the type of cues used for vowel discrimination and identification (see Figure 1.2). First of all, English has a larger number of contrasting vowel monophthongal categories than Catalan and Spanish (11 versus 5-7), which may explain that Catalan-Spanish learners of English do not weight spectral and durational acoustic cues in perceiving vowels as native English listeners do.

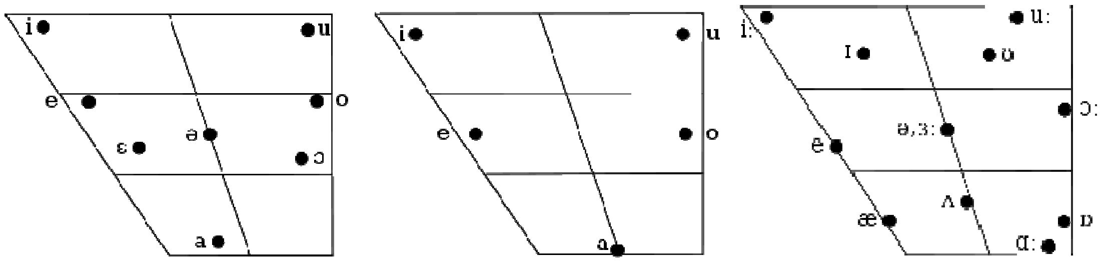


Figure 1.2. Vowel systems of Catalan (8 vowels), Spanish (5 vowels) and Southern British English (11 vowels).

1.4.1.1. Overview of the L1 vowel system of participants

The bilingual Catalan-Spanish population is somehow peculiar due to the potential overlap between their two slightly different native vowel systems (Bosch et al., 2000; Pallier et al., 1997; Sebastián-Gallés, Echeverría, & Bosch, 2005). Whereas Spanish has a five-vowel system (/a e i o u/) and no duration contrast (e.g. Flege 1989), Catalan has a seven-vowel system with an additional contrast in height, distinguishing higher-mid vowels /eo/ from lower-mid vowels /ɛɔ/, and has no duration contrast either. Spanish /e/ falls between the spectral qualities of the Catalan /e/ and /ɛ/ and, therefore, differences in language dominance have been found to predict different degree of accuracy in vowel discrimination along the Cat. /e/ - Sp /e/ - Cat /ɛ/ continuum (Mora, Keidel, & Flege, 2011; Mora & Nadeu, 2012). Whereas Spanish-dominant bilinguals have been shown to perform poorly in the discrimination of Catalan mid vowels (Pallier et al., 1997, 2001; Sebastián-Gallés & Soto-Faraco, 1999), Catalan-dominant or balanced bilinguals identify the /e/-like vowels at higher accuracy rates. The pervasive influence of Spanish has also been reported as responsible for less categorical perception of Catalan back mid /o/-/ɔ/ and, consequently, for less distinct realization of the Catalan mid vowels. Taken together, research has shown that Catalan mid-vowel contrasts are difficult for Spanish-dominant bilinguals (Sebastián-Gallés & Soto-Faraco, 1999; Pallier et al. 1997; Sebastián-Gallés et al.,

2005) and that young people in Catalonia tend to neutralize those contrasts due to the influence of Spanish (Recasens, 1993). All in all, bilingual Catalan-Spanish learners of English in Catalonia constitute a different population from the Spanish-monolingual learners often studied in previous research.

The native vowel system is believed to function as a *filter* through which L2 sounds are perceived (Trubetzkoy, 1939, 1969; Flege, 1981; Best, McRoberts & Goodell, 2001; Best & Tyler, 2007). L2 perception is also thought to determine L2 production up to a certain extent (See section 1.1.2. about the *perception leads production* hypothesis). Importantly, the interaction of the phonological systems of Catalan and Spanish (Sebastián-Gallés & Bosch, 2005) contributes to the complexity of the phonology of the participants under study and will inevitably influence L2 vowel learning in adulthood.

1.4.1. Commonalities and differences between the L1 and L2 vowel systems

Unlike the Catalan and Spanish vowels (Bosch & Sebastián-Gallés, 2003), the English vowels are not only distinguished by means of *height* and *frontness-backness* but by *duration* and *tenseness-laxness*, features that are not contrastive in the Spanish or Catalan vowel systems (Flege, 1989). Catalan/Spanish front /i/ closely approximates English /iː/ although it shares the duration feature with English /ɪ/, which also is closer in F1 to the Catalan/Spanish front /e/. The two high-front /iː/ and /ɪ/ vowels may perceptually overlap for L1 Catalan-Spanish listeners if both are mapped onto a single L1 Catalan-Spanish /i/ category.

None of the native vowel categories of Catalan-Spanish bilinguals is acoustically or articulatory identical to any English vowel category but some commonalities and differences between L1 and L2 vowels can be outlined:

- Whereas the Catalan-Spanish shared vowel system (Catalan and Spanish share /aeiou/ in their vowel systems) has two high vowels (/i/ and /u/), English has four (/iː/-/ɪ/ and /ʊ/-/uː/). Catalan/Spanish /i/ is spectrally closer to the English /iː/ than /ɪ/, although it more closely resembles the latter in duration (Stockwell & Bowen, 1965). Additionally, the F1 values of English /ɪ/ fall between the Catalan/Spanish /i/ and /e/, although its F2 is closer to Spanish /e/.

- In studies involving Catalan speakers (Cebrian, 2007), the English mid-front vowel /e/ (close to English /ɪ/ in F1) has been found to be particularly similar in F2 and, therefore, equated to Catalan/Spanish /e/ (Álvarez González, 1980; Escudero, 2005; Flege, 1991; Hojen & Flege, 2006). English /e/ falls between Catalan /e/ and /ɛ/, although the former probably occupies a space shared by the corresponding Catalan/Spanish /e/s.
- The English front-central-back low vowels /æʌɑː/, slightly different in F1 and substantially different in F2, correspond to /a/, the only open vowel in Catalan and Spanish.
- English /ɒ/ is similar to Catalan/Spanish /a/ with respect to F1 but its F2 is closer to Catalan/Spanish /o/ than to Catalan /ɔ/ in F1 (Cebrian et al., 2011). Similarly, the F2 values of English /ɔː/ approximate the Spanish /o/ and the (frequently merged) Catalan /o/-/ɔ/.
- Similar to the high-front vowels, English /ʊ/-/uː/ have generally been found to be acoustically closer to Catalan/Spanish /u/ in F1. However, it should be noticed that high-back vowel fronting (e.g. with /ʊ/ becoming particularly centralized) is currently a sound change in progress (Torgersen, 1997; Watt & Tillotson, 2001), in SBE (Hawkins & Midgley, 2005; McDougall & Nolan, 2007) and other varieties of English (Cox & Palethorpe, 2001; Fridland, 2008; Gordon et al., 2004).

- While English /ɜː/ has no correspondence with any Spanish vowel, its F2 is relatively close to Catalan /ɛ/, although the latter is considerably more fronted. It should be noticed that Spanish-dominant bilinguals with poor differentiation of Catalan /e/-/ɛ/ would not probably associate the central English vowel to any of their native categories. This was in fact suggested by Cebrian (2007), who tested Catalan listeners on their identification of English /iː/-/ɪ/-/ɜː/ on a continuum constructed in 11 vowel quality steps and 4 temporal steps and found they did not differ from NSs for /ɜː/. Catalan listeners showed native-like reliance on spectral cues in the identification of /ɜː/ regardless of its duration (100ms, 150ms, 200ms and 250 ms).

This review suggests that English vowels that are dissimilar to Catalan/Spanish vowels (e.g. /ɜː/) are not problematic for Catalan-Spanish speakers. English vowels which share the 'phonetic space' occupied by one Catalan/Spanish vowels (e.g. /iː/-/ɪ/, /æ/-/ʌ/-/ɑː/) are typically confused in perception (when there is no one-to-one correspondence), especially where duration is a part of the difference) (Coe, 2001).

1.4.3. Difficulties perceiving and producing English vowels and cue weighting

Difficulties in non-native vowel perception and consequently, lack of accuracy in L2 vowel production (see Section 1.1.2 about the *perception leads production* hypothesis), have been attributed to factors such as the size of the L1-L2 vowel inventories, the distribution of vowel categories in the vowel space (Iverson & Evans, 2009), and the acoustic cues used for vowel contrast in each of the languages (Bohn, 1995; Cebrian, 2006; Iverson et al., 2003; Iverson & Evans, 2007; McAllister et al., 2002).

Catalan-Spanish speakers experience great difficulty perceiving and producing English vowels accurately. Catalan and Spanish /i/ and /a/ occupy approximately the portion of the spectral space occupied by the English vowel contrasts /iː/-/ɪ/ and /æ/-/ʌ/, respectively. Consequently, English /iː/-/ɪ/ and /æ/-/ʌ/ are confused and typically assimilated to two single L1 categories, Catalan/Spanish /i/ and /a/, respectively.

A great deal of research conducted in immersion, FI and training settings has exclusively focused on the English vowel contrasts /iː/-/ɪ/ and /æ/-/ʌ/ and has reported similar problems in perception and production for both Spanish monolinguals (Escudero, 2000; Flege, 1991, 1995; Flege et al., 1997; García-Lecumberri & Cenoz, 1998; Iverson & Evans, 2007, 2009; Morrison, 2002) and Catalan-Spanish bilinguals (Aliaga-Garcia, 2011; Cebrian, 2002, 2006; Mora, 2008; Mora & Fullana, 2007; Rallo, 2009).

Expressed in PAM-L2 terms (Best, 1995; Best & Tyler, 2007), Catalan-Spanish bilinguals are predicted to assimilate English /iː/ and /ɪ/ to Catalan-Spanish /i/ via a category-goodness assimilation pattern. Therefore, one instance of the vowel pair may be more consistently assimilated to its native counterpart than the other. Whereas English /iː/ is commonly perceived as a good match for (or similar category to) Catalan/Spanish /i/, English /ɪ/ is initially considered a poor match. Following the SLM predictions (Flege, 1995), a new category is likely to be established for the least similar category (/ɪ/), but this does not imply quick success in discriminating the English /iː/-/ɪ/contrast. In fact, researchers hold that the acquisition of /iː/ and /ɪ/ appears to be particularly difficult for Spanish speakers (Bohn, 1995; Escudero & Boersma, 2004; Escudero, 2006; Koundarova & Francis, 2008; Morrison, 2008, 2009). According to the NLM-e predictions, English /iː/ is commonly perceived as the closest category to the native prototype, whereas /ɪ/ is not clearly mapped onto a single L1 category. The findings from previous studies conducted on Catalan learners of English reveal discrepancy regarding the perceptual mapping of /ɪ/. Whereas some studies have shown that English /ɪ/ is perceived as similar to Catalan /e/

(Rallo, 2005) by inexperienced learners, Cebrian (2006, 2009) has suggested that English /ɪ/ is most frequently assimilated to Catalan /i/ regardless of L2 experience.

Current research into L2 phonology suggests that one of the key issues to understand difficulties and inaccuracies in the acquisition of L2 vowels is acoustic cue weighting (e.g. how a listener weights the spectral and durational information when perceiving non-native vowel stimuli). Native English speakers have been shown to rely predominantly on spectral properties (F1 and F2) when identifying vowels, with duration playing only a secondary role (Hillenbrand, Clark, & Houde, 2000) (See section 1.2.5.1. for further information).-

Bohn (1995) found that while English speakers rely heavily on spectral information, Spanish and Mandarin speakers relied upon the duration of the English /iː/-/ɪ/ stimuli more than on the spectral information. Following Bohn's *Desensitization Hypothesis* (1995), the formation of a new category for English /ɪ/ will be probably first based on different duration values of /iː/ (i.e. short /iː/ vs. long /iː/) because Spanish learners of English have become linguistically "desensitized" to the spectral differences between /iː/ and /ɪ/ and vowel duration is probably more acoustically salient.

Escudero and Boersma (2004) argued that Spanish listeners employ a L1 mechanism that detects the statistical distribution of (non-L1) duration in English vowel productions and impedes vowel categorization on the basis of quality. Escudero (2000) attributed inter-subject variation in cue weighting to the dialectal heterogeneity of the Spanish population in her study (See section 1.2.5). Whereas native Scottish and SBE speakers produced a relatively large spectral difference and no or little duration difference between /iː/ and /ɪ/, Spanish learners of English made significantly less use of spectral cues and/or significantly more use of duration cues. This is consistent with their perception patterns: Spanish listeners made a category-goodness assimilation of English /iː/ and /ɪ/ to Spanish /i/, distinguishing the two English vowels on the basis of their duration differences. Morrison's

(2005) reanalysis of Escudero & Boersma (2004) presented a similar pattern of results for Spanish learners in spite of choosing a new metric for cue weighting (logistic regression coefficients) which differed from Flege et al. (2007) and Escudero & Boersma (2004)'s traditional reliance measures.

In general, all these findings point to the fact that Spanish listeners do not perceive English vowels accurately because (1) they over-attend to some secondary acoustic dimensions and (2) have difficulties directing their attention towards primary acoustic cues in vowel categorization. In other words, the results suggest that Catalan-Spanish speakers increase the perceptual distance between /iː/ and /ɪ/ along a particular dimension (i.e. duration).

According to the SLM, the accuracy with which non-native vowel segments are produced is limited by how accurately they are perceived, and an improvement in production is preceded by one in perception. Catalan-Spanish learners of English have been reported to initially be reliant on duration and assimilate English /iː/ to their native /i/ in view of their perceived similarity. Later, they become aware of the spectral differences between English /iː/ and /ɪ/ at some stage of acquisition (Escudero, 2000; Bohn, 1995). The gradually more accurate production of /ɪ/ may lead to the deterioration of its tense, long counterpart /iː/ (Cebrian, 2007; Flege, 1992) until the tense-lax contrast is be eventually perceived as a duration and spectral contrast (Escudero & Boersma, 2004). Morrison (2008, 2009) challenged the initial stage in the Spanish speakers' acquisition of English /iː/ and /ɪ/ hypothesized by Escudero (2000) (*duration-based perception*) and proposed a new earlier *multidimensional-category-goodness assimilation* stage. Following Morrison's (2009) hypothesis, Catalan-Spanish learners of English are predicted to label stimuli that were considered good examples of Spanish /i/ as English /ɪ/ and poor matches as English /iː/ (See section 1.2.5.). However, Morrison (2009) admits that additional research is needed to determine the influence of the L1 dialect on the learner's labelling and learning of English /iː/ and /ɪ/.

In line with cue-weighting studies, Cebrian (1996, 1999) found no significant spectral differences between /iː/ and /ɪ/ productions of Catalan learners of English. The findings revealed that mispronunciations of /ɪ/ were either acoustically closer to /iː/ (too low F1 and too high F2) or /e/ (too high F1 and too low F2). Interestingly, the results obtained by Flege et al. (1997) showed correlations between accurate production of /ɪ/ and poor productions of /iː/. This suggests that learners gradually reanalyze the tense-lax contrast with increased L2 experience and higher amount of L2 input. This will lead learners to redirect their attention away from duration and produce larger spectral differences between /iː/ and /ɪ/.

In agreement with SLM and PAM-L2, Cebrian (2006, 2007) compared accurate productions of the English vowels /e/-/ɪ/ to the inaccurate productions of English /iː/-/ɪ/. English /e/ was found to be accurately produced by Catalan speakers, and the fact that they used the most similar L1 categories in their production of English /e/ was probably unnoticed by English listeners. However, the results obtained for the productions of /iː/ and /ɪ/ did not match those obtained for /e/-/ɪ/. This lends support to Bohn's *Desensitization hypothesis*: Unlike English /e/, English /iː/ (similar to Catalan /i/) was not produced as expected for a near-identical vowel but patterned with the dissimilar vowel /ɪ/ (new category). It is hypothesized that the failure to produce /i/ as accurately as /e/ and /ɪ/ resulted from the categorization of the /iː/-/ɪ/ contrast as a temporal opposition and a subsequent reanalysis of /iː/as a temporal variation of /ɪ/ (e.g. short /ɪ/).

The results and the predictions reported above for Catalan/Spanish bilinguals coincide in part with studies examining the perceptual similarity between English and Catalan vowels (Cebrian et al., 2010, 2011). A perceptual assimilation task (PAT), in which L2 vowels were compared against L1 vowels, and a rated discrimination task (RDT) were used to compare L1 and L2 vowels (Cebrian et al., 2011). Altogether the results of the RDT paralleled those obtained in the PAT, indicating the following assimilation patterns (See Figure 2):

- English (/æ/, /iː/ and /e/ obtained the highest identification scores (90%) and goodness ratings (4.6 out of 7) as the closest Catalan vowels /a/, /i/ and /ɛ/, respectively. The English-Catalan vowel pairs /æ/-/a/, /iː/-/i/ and /e/-/ɛ/ obtained the lowest dissimilarity ratings, which did not differ from low ratings for the English same-category vowel pair /iː/-/iː/ (2.3). By contrast, the Catalan-English vowel pairs /ɛ/-/æ/ (6.4) and /e/-/e/ (4.8) received the highest dissimilarity ratings, closely followed by the Catalan-Catalan pair /e/-/ɛ/ (4.6). Consistent with earlier studies (Cebrian, 2006, 2009), the results indicate that English /æ/, /i/ and /e/ are readily and strongly assimilated to a Catalan vowel more than 90% of the time, and are therefore perceived as instances of the corresponding native categories (/a/, /i/ and /e/).
- English /ʌ/, /uː/, /ɔː/, and /ɪ/ were fairly consistently identified (80-90%) as Catalan /a/, /u/, /o/ and /i/, respectively, but obtained low goodness ratings (3.5-3.8). Catalan-English vowel pairs /a/-/ʌ/, /u/-/u/, /o/-/ɔː/, /i/-/ɪ/ were perceived in the mid range of the dissimilarity scale (3.2-3.6).
- English /ɒ/ and /ɑː/ obtained relatively close assimilation rates (78%) but different goodness of fit ratings as Catalan vowels. English /ɒ/ was identified as Catalan /ɔ/ 70% of the time (goodness rating: 4.1), while English /ɑː/ was perceived as Catalan

/a/ 78% of the time (goodness rating: 2.5). Similar to Cat-Eng /ɛ/-/e/, the dissimilarity ratings for Cat-Eng /ɔː/−/ɒ/(2.3) did not differ significantly from the rates for Cat-Eng /i/-/iː/, suggesting that English /ɒ/ is also heard as an instance of a native category. By contrast, the Catalan-English /a/-/ɑː/ pair was heard to be relatively dissimilar (4.5).

- English /ɜː/ obtained the lowest identification score (9-30%) as the Catalan vowels /e/, /ɛ/, /a/, /ɔ/ and /o/, and the lowest goodness rating (1.4-2.7), which confirmed that English /ɜː/ was not consistently assimilated or perceived as any instance of a native vowel category, but was highly dissimilar from any Catalan vowel. Similarly, the English vowel /ʊ/ was not strongly clustered with others.

		English stimuli /i	ɪ	e	æ	ʌ	ɜ	ɑ	ɒ	ɔ	u/
Catalan responses	i	96 (4.6)	82 (3.2)								
	e		15 (3.4)	7 (3.3)			30 (2.0)				
	ɛ			91 (4.7)			24 (1.7)				
	ɔ					11 (4)	9 (2.7)	20 (3.9)	70 (4.1)	10 (2.6)	
	o						9 (2.1)		28 (4.4)	86 (3.6)	
	u										93 (3.8)

Figure 1.3. Assimilation patterns of English vowels to Catalan vowels (in percentages) and goodness of fit (in parenthesis) (adapted from Cebrian et al., 2011).

In Cebrian et al. (2011), lower goodness ratings obtained for English vowels in the PAT consistently corresponded to higher dissimilarity scores in the RDT. However, English /ɪ/ was predominantly assimilated to Catalan /i/ in the PAT but was perceived as more dissimilar from Catalan /i/ than from Catalan /e/ in the RDT. Discrepancies matched Cebrian's earlier findings (2006, 2009) showing that English /ɪ/ was mostly assimilated to Catalan /e/. This finding lends support to models of L2 speech learning whose predictions for difficulty are based on the notion of cross-linguistic similarity and non-native use of cue weighting.

Mora & Fullana (2007) replicated earlier findings of learners' overuse of temporal cues (Flege, 1995; García-Lecumberri & Cenoz, 1998). Their findings also showed non-native implementation of spectral differences in the two vowel contrasts /iː/-/ɪ/and /æ/-/ʌ/, irrespective of AOL and L2 experience. Spanish learners produced the target vowels without significant spectral differences and significantly longer durations than native speakers.

In contrast, Flege et al. (1997) and, more recently, Iverson & Evans (2007, 2009) showed evidence that Spanish speakers made the same use of spectral and duration cues as speakers with large native vowel systems (i.e. German) despite their smaller vowel systems and their poor L2 vowel perception and production. Nevertheless, Spanish speakers obtained lower identification scores and smaller improvement after training than German speakers, with more crowded vowel spaces (for which the SLM predicted a higher degree of L1 interference), despite having more room left for new vowel categories to develop.

Altogether thee difficulty that Catalan-Spanish speakers experience in attending to both temporal and spectral cues and their over-reliance on duration may be due to the fact that single-category assimilation may have desensitized them for the perception of spectral cues (Bohn & Flege, 1990; Bohn, 1995; Cebrian, 2006; Escudero & Boersma, 2004; Morrison, 2008).

1.4.4. Research on L2 vowel perception and production

At present, findings from vowel studies with L2 learners do not provide conclusive evidence about the exact relationship between L2 vowel perception and production. They have yielded evidence both for and against the *perception leads production* hypothesis (Chan, 2001; Flege & Eefting, 1987; Ingram & Park, 1997. The complexity of the perception-production link partly lies in the fact that it is particularly difficult to compare vowel perception and production due to differences in the experimental methods used in testing the two dimensions. There are various factors that may account for disparity of results in vowel perception and production: consonantal contexts in which vowels are elicited, the English dialect of the vowel stimuli, the number of response categories or labels in perceptual tasks, and the methods used to elicit production data.

Previous studies reporting on the effect of L2 experience on vowel production (i.e. Flege, 1995; Flege et al., 1997; Flege & Bohn, 1992; Munro, 1993) suggest there is a close link between the perception and production of L2 vowels for both experienced and inexperienced learners. Other studies have demonstrated a mismatch between vowel perception and production as a function of L2 experience. Strange (1995) challenged the *perception leads production* hypothesis, holding that perceptual difficulty with vowels may persist after production has been mastered. Bever (1981) stated that L2 perception and production may be uncorrelated in more experienced learners. Flege et al. (1997) demonstrated that inaccurate productions of English /iː/-/ɪ/ by Spanish speakers correlated weakly with their native-like vowel perception. Only modest correlations between vowel perception and production were found in Flege (1999) for highly experienced learners of English. Similarly, Cebrian (2007) reported poor production accuracy for /iː/ and found a negative correlation in the success with which learners of English produced /iː/ and /ɪ/: at some stage the best /ɪ/ producers were found to be the worst /iː/ producers. An interpretation of this finding suggests that target-like production of the new L2 vowel sound may come at the cost of the quality in the production a neighbouring L2 sound.

However, the assumption that perception leads production may explain why most phonetic training, as well as most FL teaching and pronunciation tasks for L2 vowels are based on perception tasks rather than perception.

Various researchers have investigated the relationship between changes in L2 vowel perception and production in adulthood. For example, a study examining perceptual responses to English vowels (Fox et al., 1995) showed that the vowel space of proficient listeners was more target-like than that of the non-proficient listeners. This suggests that the perceptual dimensions used in vowel identification can be gradually modified with increased L2 proficiency. The multidimensional scaling (MDS) analyses in Fox et al. (1995) revealed two differences between proficient and non-proficient Spanish listeners. Flege et al. (1997) demonstrated that adults will eventually gain accuracy in L2 vowel perception and production as they gain experience, as assessed through intelligibility rating and acoustic analyses comparing relative experienced and inexperienced adult L2 learners. Variability in the production data might be explained by the fact that age-based limitations are higher for L2 production than for L2 perception. In relation with the predictions made by Flege (1991) and Flege et al. (1997), Fullana & MacKay (2003) did not find significant effects of AOL or L2 experience in a FI context on the Spanish speakers' /i/ identification scores. However, native judges tended to identify older learners' productions at higher correct rates than those of younger learners. This in turn suggested that experience played a more modest role in the improvement of L2 vowel perception as compared to L2 vowel production.

In view of these findings, further research on changes in L2 vowel perception and production in adulthood is clearly needed. Hojen and Flege (2006) raised the questions of how much and what type of L2 input learners must receive to achieve native-like perception of L2 vowels. Considering that "early" Spanish learners of English showed near-native discrimination of the most difficult English vowel contrasts in Hojen and Flege (2006), special attention must be given to what type of training L2 learners must be given to accurately perceive and produce L2 vowels.

1.4.5 Summary

Two main conclusions emerge from the review of the research on non-native vowel perception and production. Most research shows that learners (1) have important difficulties in distinguishing L2 vowel contrasts that are not employed in their L1 and (2) need to overcome interference from secondary cues that are detrimental to L2 vowel perception and production. The present study pr edicts, based on previous literature, that training may help L1-Catalan/Spanish learners acquire L2 English vowel categories by increasing their attention to previously unattended primary (spectral) cues and by inhibiting their over-use of duration in vowel categorization and production.

The next section reviews research on L2 phonetic training to date, with consideration to changes in L2 speech perception and production.

1.5. Phonetic training

There is early empirical evidence that adult L2speakers do not permanently lose either their perceptual sensitivities to distinguish non-native vowel categories or their ability to control articulatory gestures and model vowel productions. Some studies show that some L2 speakers can successfully learn to perceive new L2 sounds in a native-like manner (i.e. Bongaerts et al., 1997; Escudero, 2006; Iverson & Evans, 2007; Jamieson & Morosan, 1986; Logan & Pruitt, 1995; Markham, 1997; Nishi & Kewley-Port, 2008; Ylinen et al., 2009).

The central hypothesis in this thesis is that adult Catalan-Spanish speakers can acquire even the most difficult non-native vowel phonemes if they receive *effective* training. Phonetic training studies constitute a promising area of research in favour of plasticity in L2 speech perception and production.

The findings in the literature indicate that the perceptual spaces of Catalan-Spanish learners of English are miss-tuned for acquiring the whole set of SBE monophthongal vowels. Catalan-Spanish speakers often overweigh duration cues, which are secondary to L2 vowel categorization, in perceptual and/or production tasks and fail to attend to critical spectral cues (See Section 1.4 for a review). However, training studies demonstrate that it is likely that lower-level perceptual processes are modified via adequate training procedures, with varying degrees of success, despite the loss of perceptual sensitivity imposed by age-related biological limitations in adulthood (Iverson et al., 2003).

The aim of this section is to review research on L2 phonetic training to date, with special emphasis on areas of interest such as:

- Changes in adult speech perception and production
- The influence of task variables such as: training tasks, type of training, stimuli, stimuli and talker variability, scope of training set and feedback administered.
- Generalization to novel stimuli, talkers and words
- Long-term retention of outcomes

1.5.1. Overview of phonetic training studies

Regardless of the challenge that L2 speech learning poses for L2 learners, different training studies conducted so far have established that it is possible to develop target-like perception and production of 'difficult' L2 sounds (i.e. Aliaga-Garcia& Mora, 2007; Aliaga-

Garcia 2011; Aliaga-García & Mora, 2008; Bradlow et al., 1999; Flege, 1995; Jamieson & Morosan, 1989; Lively et al., 1994; Logan & Pruitt, 1995; Iverson & Evans, 2007; Nishi & Kewley-Port, 2008; Ylinen et al., 2009).

Previous training studies have included a variety of subjects and target sounds, such as Americans training on exotic sounds (Catford & Pisoni, 1970) and Zulu (Best et al., 1988), Japanese and Americans training on Hindi (Pruitt, 1995), speakers of Canadian French (Jamieson & Morosan, 1986), Chinese (Flege, 1989), Japanese (Bradlow et al., 1999) and many others training on English sounds.

Over the past four decades, several important advances have been made toward establishing effective laboratory training procedures for improving the perception of difficult L2 sounds (i.e. Akahane-Yamada, 1996; Jamieson, 1995; Pisoni et al., 1994; Logan & Pruitt, 1995; Pisoni & Lively, 1995). Perceptual training studies conducted so far have mainly investigated English consonant and vowel contrasts, with the exception of a few suprasegmental studies (e.g. Wang et al., 2003). The case of Japanese adults learning the English /r/-/l/ contrast has been one of the main foci of research within the area (i.e. Logan et al., 1991) because of its extreme difficulty for Japanese speakers (i.e. Goto, 1971; Sheldon & Strange, 1982; Yamada & Tohkura, 1992) and its resistance to learning after discrimination training (Strange & Dittman, 1984). The outcome of these studies is that perceptual learning also led to an improvement in production (i.e. Bradlow et al., 1999).

1.5.2. Training techniques in perceptual training

As regards the methodology implemented, auditory training studies can be roughly divided into those following: Cue Enhancement –which includes adaptive training, perceptual fading, cue inhibition techniques (e.g. Jamieson & Morosan, 1986, 1989; Iverson et al., 2005)–, or High Variability methods (e.g. Aliaga-García, 2010, 2011; Iverson et al., 2003, 2004; Lengeris & Hazan, 2010; Nishi & Kewley-Port, 2007) simulating exposure to a more natural type of stimulus presentation and cue distribution, as in this study.

1.5.2.1. Cue-manipulation training

Most of the /r/-/l/ training procedures have suggested that *enhanced* stimuli can help learners develop selective attention to the crucial phonetic cues of certain L2 sounds. The Perceptual Fading technique makes use of maximally contrastive enhanced stimuli (Jamieson & Morosan, 1986; Iverson et al., 2005; McCandliss et al., 2002; Pruitt et al., 2006). All Enhancement and Secondary Cue Variability types of training (Iverson, Hazan, & Bannister, 2005) are also based on cue-manipulation methods.

All these methods are based on the enhancement of listeners' attention to category-relevant dimensions but include gradual or steady manipulation of acoustic cues. Adaptive training starts with stimuli with exaggerated values (distant from non-manipulated acoustic dimensions) and gradually reduces the enhancement of acoustic cues to resemble natural stimuli, as perceptual acuity improves over the course of training.

Evidence of *generalization* of the knowledge acquired by means of synthesized stimuli to improvement in listening to natural speech has been found in a large number of training studies (Bradlow et al., 1997; Fox & Maeda, 1999; Jamieson & Morosan, 1986; Hardison, 2003; Hazan et al., 2005; Nobre-Oliveira, 2007; Pruitt et al., 2006; Yamada et al., 1996; Wang et al., 1999; Wang, 2002; Wang & Munro, 2004). However, more research on this issue is needed to find out whether training with enhanced stimuli is more effective than training with natural stimuli.

1.5.2.2. Auditory High-variability phonetic training

Generally speaking, consonant training studies outnumber L2 vowel training studies (see Table 1.2). While previous studies have shown that vowel synthesis can also be useful to enhance subtle but crucial L2 properties (Wang, 2002; Wang & Munro, 2004), the most commonly used training protocol for vowel training has emphasized the importance of exposing learners to naturally produced high variable stimuli or minimal pairs in multiple environments. HVPT has proved particularly effective in promoting "robust" L2 vowel categories by means of a variety of natural stimuli and multiple talkers and contexts. HVPT is especially effective when the difficulties with L2 sounds lie in the use of L2 phonetic cues which are unused or weighted differently in the L1, or are based on the integration of multiple cues.

HVPT has also proved effective to improve both the perception of L2 consonants (i.e. Hazan et al., 2005; Lively et al., 1993; Logan & Pruiit, 1991; Pruitt et al., 2006) and L2 vowels (Iverson & Evans, 2009; Lengeris & Hazan, 2010; Nishi & Kewley-Port, 2007, 2008). It has shown (1) robust learning and generalization to novel talkers and tokens (i.e. /r-l/: Lively et al., 1993; Logan et al., 1991; Pruitt et al., 2006; Yamada, 1995), followed by (2) long-term retention of gains months after the completion of training (Bradlow et al., 1999; Iverson & Evans, 2009; Lively et al., 1994) and (3) transfer to speech production for both consonants (Bradlow et al., 1997; Hazan et al., 2005) and vowels (Lambacher et al., 2005; Lengeris & Hazan, 2010).

We can identify three major generalizations regarding the success of HVPT to train Japanese speakers on English vowels (Lambacher et al., 2005; Nishi & Kewley-Port, 2007), English speakers on the Japanese vowels (Hirata et al., 2007; Tajima et al., 2008) and Spanish speakers on English vowels (Aliaga-Garcia, 2007, 2011; Iverson & Evans, 2007):

First of all, the success of HVPT demonstrates that vowel perception displays sufficient plasticity in adulthood to undergo substantial modification through laboratory training alone. The major strength of HVPT for perceptual learning is that it produces highly **generalized learning** (e.g. improvement is generalized to novel items and talkers).

Carlet & Cebrian (2015) have recently highlighted the efficiency of short-term HVPT to improve L2 vowel perception and promote "robust learning" and category formation.

Secondly, the HVPT method produces immediate as well as **long-term changes** in the underlying phonetic system. The impact of this method to perceptual learning has been established experimentally (Lively et al., 1993; Magnuson et al., 1995). However, it still remains an empirical question whether HVPT is more effective in promoting long-term improvement in L2 production than low-variability training methods (Bradlow et al., 1997, 1999; Lively et al., 1994).

Finally, perceptual HVPT training studies (i.e. without specific pronunciation practice) have shown that improvement extends beyond the perceptual domain to produce changes in articulation. On the one hand, long-lasting improvement in both L2 perception and production is consistent with the theories that posit integrated perception and production systems. Bradlow et al.'s (1999) findings demonstrate the effectiveness of the HVPT for the acquisition of fine phonetic details of the trained contrasts. Furthermore, the retention and transfer of knowledge three months after the training across perception and production domains implies *"a close link between speech perception and production during perceptual learning of novel phonetic contrasts"* (Bradlow et al., 1999: 983).

More recently, Carlet (2017) has shown the efficacy of two HVPT methodologies (based on discrimination and categorical AX discrimination). Discrimination HVPT proved to be more suited to enhance learners' perception of both attended and unattended target sounds, whereas identification HVPT seemed to be more efficient in promoting generalization and retention. Importantly, identification HVPT was considered superior to discrimination HVPT for training L2 vowels, due to the fact that only identification HVPT was found to lead to improvement in production for trained vowels. It is worth noting that identification training had a positive impact in both L2 perception and production even in the absence of production training.

HVPT supports the motor theory of speech production (Liberman et al., 1967;) and the direct realist approach to speech perception (Best, 1995), which share the idea that listeners perceive sounds in terms of articulatory gestures (see Sections 1.3). HVPT can be effective in promoting associations between 'intended phonetic gestures' (input administered) and actual vocal productions (output generated by the learner). The fact that perceptual training alone also produces changes in speech production may be due to learners' on-line monitoring of their own output, as we hypothesize to be the case in the ART method in the present study.

1.5.2.3. Audiovisual (AV) phonetic training

Audiovisual types of HVPT (Hardison, 1999; Ortega-Llebaria et al., 2001; Hazan & Sennema, 2005, 2007) have proved more effective than auditory-only (A) training in (a) increasing the discriminability of the target contrast (provided the visual cues are sufficiently salient), in (2) leading to long-lasting improvement in L2 sound categorization (even in environmental degradation), and in (3) transferring perceptual gains to L2 production (Bradlow et al., 1997).

When Flege (1989) first attempted to test the effectiveness of using visual feedback to train L2 vowel production to one Spanish learner of English, scarce and unsystematic information concerning the use of visual cues for training existed. Work with hearing-impaired participants (Osberger & Kent, 1983) suggested that, whereas audiovisual training had little effect on F2 because the visual articulatory differences were hard to see, differences associated to jaw movement and F1 were probably more useful for vowel training. Prior work by Massaro (1987) had shown that the use of audiovisual cues had a positive effect on perception. Osberger (1987) had just used audiovisual information to teach hearing-impaired children to produce /iː/-/æ/. Povel & Wansink (1986) had used visual information derived from the acoustic output to represent tongue position (i.e. a dot on a computer screen varying its position according to the spectral quality of the vowel as produced by learners). However, Flege (1989) designed a vowel training (/iː/-/ɪ/ and /æ/-/ɑː/) with visual feedback administered through a glossometer (Fletcher & Hasegawa, 1983) which measured and displayed the speaker's tongue position as they attempted to match visual tongue targets for the English vowels. After a ten-session audiovisual training, the Spanish learner produced /ɪ/ and /ɑː/ more authentically, and was able to contrast /iː/ and /ɪ/, but visual feedback did not lead to significant changes for the other vowels. In more recent studies, the effect of visual cues has been found to depend on the L1 background and on the visual salience of the phonetic contrast. For instance, Hazan & Sennema's (2007) audiovisual exposure to articulatory gestures through the CSLU toolkit (Cole et al., 1999) –a language-training application consisting of an animated conversational agent called Baldi (Massaro & Cohen, 1998)–led to a greater improvement in the learners' pronunciation of /r/-/l/, even without specific pronunciation training. However, the same study revealed that "*mere exposure to visible articulators did not appear to have the same effect on production than combined exposure to visible articulators and to acoustic information.*"

More recently, Kartushina et al. (2015) has proved that production training with immediate, trial-by-trial, visual feedback about the learners' articulation (tongue position and mouth openness) alongside the native speakers' articulation of the target vowel improves L2 perception. Importantly, the experiment by Kartushina et al. proved that as little as 1h of production training with visual feedback is effective in improving the production accuracy of four trained vowels. However, although trainees exhibited significantly more accurate productions of these vowels, they did not reach native-like performance levels. It is speculated that the amount of training and the low-variability stimuli used may have been insufficient to promote native-like performance on any of the trained vowels. Further work is needed to validate the efficacy of this new "articulatory feedback training method", both in the context of L2 speech sounds that are easily perceived and those poorly perceived that require completely new articulatory patterns.

1.5.3. *Fullset vs. subset in vowel training*

While consonant training procedures have largely used small problem-focused sets of consonants, vowel training studies have either focused on small *subsets* (e.g. with difficult vowels: Akahane et al., 1997; Carlet, 2017; Nobre-Oliveira, 2007) or widened its scope to include a *fullset* of vowels (e.g. Iverson & Evans, 2007, 2009; Lengeris & Hazan, 2010). This has raised questions about (1) the overall efficacy of widening or narrowing the scope of training and (2) the generalizability of gains after a problem-focused type of training.

The majority of the vowel training studies have trained learners on less than five vowels. For example, studies training Japanese speakers on three AE vowels (/æ, ɑː, ʌ/) (Akahane et al., 1997) and five AE vowels (/æ, ɑː, ʌ, ɔː, ɝ/) (Lambacher et al., 2005; Sperbeck et al., 2005) showed that learners could simultaneously learn up to five L2 vowel categories through perception training and improvement was extended to novel tokens spoken by novel speakers. However, training including a smaller vowel subset showed no improvement for untrained vowels. Strange et al. (2001) showed that the perceptual assimilation of 11 AE vowels to L1-Japanese vowels depended on consonantal contexts, which raised the question of whether these vowels were similarly trainable. Taken together, these findings suggest that training using only a small *subset* of vowels may not help learning a complete set of L2 vowels (Nishi, 2007; Nishi & Kewley-Port, 2007). In other words, the trained vowel set might have inevitably trained learners to focus their attention on the cues that are relevant only for the trained vowels.

Nishi (2007a) investigated for the first time the effect of the size of the training vowel set on Japanese speakers' identification of AE vowels. Nishi compared the overall training efficacy of using one small problem-focused vowel *subset* (/ɑːʌʊ/) to a *fullset* covering the entire AE vowel space in a variety of consonantal contexts. Nishi's findings (2007a) cautioned against the risk of using small *subsets* for vowel training and stated that the narrow scope of the training introduced complications in the learning process. According to Nishi, small vowel *subsets* might have induced learners to unevenly learn to ignore cues that are related to untrained vowel categories. Therefore, learners trained on small *subsets* may fail to learn the complete L2 vowel set. As an extension of their previous studies, Nishi & Kewley-Port (2007b) revealed that the trainees following 9V-9V protocol significantly outperformed learners following hybrid protocols (i.e. 9V-3V and 3V-9V with a problem-focused component at the beginning or by the end of the training).

Nishi's findings highlighted the possible negative influence of using smaller training sets or, alternatively, the problems of training first on a subset with difficult vowels and then on a larger set. Nishi tentatively concluded that human language learning requires an adequate amount of complexity that enables multiple-cue judgment rather than memorizing the association between a category and exemplars. Small training subsets require exemplar-based judgment, by which the task simply becomes a memory retrieval process.

Several recent vowel training studies have included a large set of vowels (i.e. Aliaga-García, 2011; Iverson & Evans, 2007, 2009; Lengeris & Hazan, 2010), whereas others have investigated the effect of perceptual training on the perception of three "difficult" English vowel contrasts /iː/-/ɪ/ and /e/-/æ/, /ʊ/-/uː/, e.g. by Brazilian EFL students (Nobre-Oliveira, 2007), or the effect of identification and discrimination training on the perception and production of /iːɪæʌɜː/ by Catalan-Spanish learners of English (Carlet, 2017)

The present training study includes the whole set of SBE monophthongal vowels: (/iːɪeɜːæʌɑːɒɔːʊuː/. This general research agenda will also reveal information regarding the extent to which the adult phonetic system is plastic and thus capable of undergoing linguistically meaningful modifications.

1.5.4. *Perceptual vs. production training*

In spite of the inconsistent findings in cross-language research investigating the complex relationship between L2 perception and production, studies exploring cross-modal training effects (comparing perception vs. production training effects) suggest there is a close connection between the perception and production domains. Retention and transfer of improvement in perception to production domains implies there is in fact a close link between perception and production during perceptual learning of L2 phonetic categories. HVPT has been shown to produce long-term modifications in both perception and production of difficult non-native phonetic contrasts (i.e. Bradlow et al., 1997, 1999).

Studies comparing auditory versus articulatory training effects are scarce and have been conducted in short-term laboratory training conditions without training on a large set of non-native vowels. The effects of articulatory training were found to be greater than those of perception in the early 1970s. Catford and Pisoni's (1970) early research showed that articulatory training improves the perception as well as the production of 'exotic' sounds. Similarly, Weiss (1992) and Matthews (1997) showed that articulatory training had a significant effect on Chinese and Japanese speakers' perception of L2 sounds, respectively. Sheldon and Strange (1982) also found significant effects of the production drills used in training on the learner's production abilities, but not on their perceptual abilities. Likewise, Linebaugh (2007) provided evidence that the *perception leads production* hypothesis (Bradlow et al., 1995; Flege, 1995; Rauber et al., 1995; Rochet, 1995) may be too strong an assertion because explicit instruction on the articulation of English /ʌ/ and /ɒ/ had a significant effect on the Spanish learner's perception and production of those vowels.

In agreement with studies reporting the positive effects of perceptual training on the perceptual and production domains, Gómez-Lacabex (2009) and Gómez-Lacabex et al. (2008) found significant training effects on Spanish speakers' identification of English /ə/. Although differences in performance after auditory vs. articulatory training were very small, the researchers noted the methodological difficulty of isolating production training from perception training in FI settings. The fact that production training exerted a positive influence on perceptual abilities supports the view that there is a facilitating relationship between perception and production training. That is, production training exerts a positive influence on the learners' perception and perception training helps the learners' production. Importantly, Gómez-Lacabex et al. (2008) is the only study assessing the effectiveness of perception and articulatory training for Spanish learners' acquisition of English vowels, but the different training methods compared are not fully methodologically controlled. First, auditory and articulatory training sessions had different durations (20-minute auditory training sessions and 40-minute articulatory training sessions). Second, different types of feedback were administered (feedback was administered in groups in auditory training sessions; only articulatory training included individual feedback). Finally, auditory and articulatory training sessions consisted of different type of stimuli (auditory training used auditory stimuli and articulatory training used audivisual stimuli). All these made the two types of training less comparable.

Perception training was not completely isolated from production training methods in their study, as both types of training were not isolated experiments but framed within an EFL course: *"A student might be trained on a specific sound auditorily during the training session and decide to focus or try to practice this sound in a non-specific speaking task demanded in the English course the next day. Likewise, a student receiving articulatory training and feedback one day may decide to listen carefully to his/her teacher's pronunciation of that particular sound in subsequent lessons."* (Gómez-Lacabex et al., 2008: 158). Further research including fully comparable training methods with large vowel sets is needed.

1.5.5. *Recent trends in training research*

The literature on phonetic training reveals a variety of methods (including signal enhancement, visible articulation cues, high variability stimulus set, identification tasks and adaptive training, discrimination task, visual articulatory feedback training) conducive to improvement in the perception and/or production of L2 sounds. As a consequence, plasticity is not considered to be strictly bound by age-related constraints but simply diminished in adulthood. Some of these methods have been widely applied to training language-impaired speakers (e.g. perceptual fading: Tallal et al., 1998).

Although L2 learners' performance will normally remain substantially below that of native speakers', phonetic training can be successful in effecting gains in learners' performance, which is observable through (a) improvement in phonetic categorization; (b) generalization and transfer of learning to untrained stimuli, untrained talkers, (c) transfer of learning to other perceptual/productive domains, and (3) long-term retention of gains.

More recently, the debate has centered over the question of what type of learning is promoted by the so-called "successful" HVPT approaches and what factors account for different rates of success in learners. Perceptual HVPT has generally proved effective because (1) it changes attentional weights causing L2 learners to attend to relevant acoustic cues and ignore irrelevant cues (Lively et al. 1993; Logan et al. 1991), (2) it generalizes better than training without talker or stimulus variability and, therefore, (3) it leads to more *robust* learning. The degree of *robustness* of learning has often been measured through the learner's *generalization* of identification abilities to untrained words or novel contexts (Carlet, 2017; Lively et al., 2004)

However, some studies (Iverson et al., 2005; Hattori & Iverson, 2009; Iverson & Evans, 2009) have suggested that HVPT may not lead to actual cue (re)weighting. While consonant training is particularly effective for improving learners' categorization of target sounds (e.g. Japanese learners of English became more consistent at labelling English /r-l/

after training), there is not enough evidence that cue weighting changes due to training. Iverson and Evans (2009) found that, despite improving in L2 vowel identification accuracy, Spanish and German speakers did not improve in their L1-L2 vowel space mapping after perceptual vowel training. Their tentative conclusion is that HVPT may be more effective than low-variability training in promoting improvement in vowel identification (e.g. because learners learn to consistently apply L2 categories to real speech). However, it cannot be concluded that HVPT changes category representation. Evidence that category representations or phoneme boundaries are not changed due to training comes from the fact that there is no improvement in vowel discrimination after training. For example, Lengeris and Hazan (2010) showed that auditory HVPT improved learners' categorization of the English synthetic /iː/-/ɪ/ continuum but not of /æ/-/ʌ/. Their findings in vowel discrimination did not match the positive findings in vowel identification. Interestingly, perceptual training led to improvement both in perception and production, in "*much less overlap of English vowels, especially in the high-front area of /iː/ and /ɪ/, the mid-front/central area of /ɜː,e, aː, ʌ/)*" (Lengeris & Hazan, 2010, p. 3765). This suggested that Greek learners of English learned to differentiate English vowels in both speech perception and production.

While some studies (Francis et al., 2000, 2007; Francis & Nusbaum, 2002; Heeren & Schouten, 2008) demonstrate that perception training can induce learning of the relevant cues in perception as well as changes in cue weighting for consonant sounds, Iverson and Evans (2009) hold that HVPT may be an effective tool to improve L2 categorization (e.g. identification performance) but this cannot provide a full simulation of the kinds of changes in phonetic perception (such as cue re-weighting) that occur during long-term exposure to phonetic categories, e.g. during stay abroad and in immersion settings.

Language neuroimaging research has recently applied modern imaging techniques to phonetic training. Callan et al. (2003) was the first FMRi study which revealed localized brain changes as a result of training a difficult L2 sound consonant contrast. Japanese learners of English undergoing ID training on /r-l/, /b-g/ and /b-v/ contrasts (with visual feedback) displayed neural activation in regions of their brain which governed lip and tongue movement during speech production. These results point out that auditory-articulatory mapping facilitates the perception of difficult L2 consonant contrasts. From this findings, it follows that both perception and production-based tasks should be essential components in L2 segmental phonetic training.

Behavioural training effects have also been reflected in recent mismatch negativity (MNN) studies (e.g. Tremblay et al., 1997; Ylinen et al., 2010). Tremblay et al.'s (1997) study provided insightful knowledge about the area where neural changes occurred as a result of perceptual learning, supporting the view that L2 perception is malleable. The left hemisphere displayed greater activity than the right hemisphere during ID training. In other words, changes in MMN in one hemisphere provided an objective measure of how effective training can be in the perception of non-native consonant contrasts at a pre-attentive level (e.g. neurophysiological changes in a short period of time), and showed that the *"learned behaviour transfers to new acoustic conditions"* (Tremblay et al., 1997, p. 3771). These findings are consistent with previous findings (Kraus et al., 1995) and confirmed that training leads to changes in neuronal activity. Training in the discrimination of two similar-sounding synthetic speech stimuli of the phoneme /da/ with visual feedback increased the number of neurons used to respond to stimuli in areas of the auditory cortex.

The behavioral data of Ylinen et al. (2010) reinforces previous findings that the perception of a new sound contrast can be acquired by adult listeners and this trained ability generalized to novel stimuli. They showed that HVPTtraining consisting of vowels with modified and unmodified duration was effective. After training, Finnish learners of English attended to spectral cues in /iː/-/ɪ/ identification more reliably when the duration cue was unreliable. They found evidence in enhanced MNN brain responses that training produced significant changes in the learners' cue weighting (e.g. in the processing of the spectral cues) at a pre-attentive level. Brain responses are also good evidence that HVPT perceptual training with increasing variability in duration leads to more robust changes in cue weighting.

Vowel Training studies	Type of training	Technique	Participants	Target vowels	Stimuli	Improvement in Perception/Production
Akahane-Yamada et al. (1997, 1998)		HVPT	Japanese	/æ, ɑː, ʌ/ AE	A	perception
Strange et al. (2001)		HVPT	Japanese	11 AE vowels	A	perception
Kingston (2003)		HVPT	English (AE)	German front rounded /ø/-/œ/	A	perception
Wang-Munro (2004)		Perceptual fading (synthetic > natural)	Chinese	3 vowel contrasts	A	perception
Lambacher et al. (2005)		HVPT	Japanese	AE /æ, ɑː, ʌ, ɔː, ɝ/	A	perception and production
Sperbeck et al. (2005)		HVPT	Japanese	AE /æ, ɑː, ʌ/	A	perception
Nishi (2007)		HVPT	Japanese	/ɑː ʌ ʊ/ vs. fullset	A	Perception (fullset training > subset training)
Nobre-Oliveira (2007)		HVPT (natural) vs. Cue enhancement (synthetic)	Brazilian	/iː/-/ɪ/, /e/-/æ/, /ʊ/-/uː/	A	Perception (synthetic > natural words)
Aliaga-Garcia (2007)		HVPT	Catalan/Spanish	SBE /iː/-/ɪ/, /æ-/ʌ/, /ʊ/-/uː/	A	Perception, cue weighting
Hirata et al. (2007)		HVPT	English	Japanese vowels	A	perception
Iverson & Evans (2007, 2009)	**Auditory**	HVPT	Spanish, German	Fullset SBE	A	perception and production (long-term gains)
Nishi & Kewley-Port (2007, 2008)		HVPT	Japanese	Vowel subset vs. Fullest 9V-9V/9V-3V/3V-9V	A	perception
Tajima et al. (2008)		HVPT	English	Japanese vowels	A	Perception: Only trained tokens
Iverson, Pinet, & Evans (2009, 2011)		HVPT	French	English vowels	A	Perception (all vowels) & Production (some vowels)
Ylinen et al. (2010)		Cue enhancement	Finnish	/iː/-/ɪ/	A	Perception, cue weighting
Lengeris & Hazan (2010)		HVPT	Greek	Fullset SBE	A	perception and production
Koundarova & Francis (2011)		Cue enhancement vs. cue inhibition vs. HVPT	Spanish	AE /iː/-/ɪ/	A	Perception, cue weighting (less use of duration; higher use of spectrum)
Pereira (2013)			Chilean		A/AV/V	Perception and production
Carlet & Cebrian (2014, 2015)		HVPT	Catalan/Spanish	/i/, /ʌ/	A	Perception
Wong (2013, 2015)		HVPT	Cantonese	/e/, /æ/		Perception and production
Hu W et al. (2016)		duration-equalized vowel perception training	Chinese			Perception, cue reweighting
Carlet (2017)		HVPT	Catalan-Spainsh	/iː/, //ɪ/, /æ/, /ʌ/, /ɜː/	A	Perception and production, generalization, retention
Flege (1989)		Imitation, Articulation	Spanish	/iː/-/ɪ/, /æ-/ʌ/-/ɑː/	AV	perception and production of /iː/-/ɪ/ *(weak)*
Linebaugh (2007)	**Articulatory**		Spanish	/ʌ/ and /ɒ/	A	perception and production
Kartushina et al. (2015, 2016)		visual articulatory feedback training	French	Danish /ɔ/, Russian /ɨ/	V	production
Gómez-Lacabex et al. (2008, 2009)	**Auditory vs. Articulatory**	HVPT	Spanish	/ə/ (vowel reduction)	A	perception
Aliaga-Garcia (2011)		HVPT	Catalan/Spanish	fullset (11 SBE vowels)	A/AV	Perception and production

Table 1.2. Summary of vowel training studies classified according to type of training: perception training, articulatory training and training studies comparing auditory and articulatory training.

1.6. Summary and conclusions

The objective of this chapter was two-fold. First, to review studies reporting on adult L2 learners' difficulties in the perception and production of L2 phonetic contrasts, and secondly, to review the relatively small number of vowel training studies to date.

The conclusion from this review is that the perceptual system of adult L2 learners is sufficiently malleable to allow for the acquisition new phonetic contrasts, provided efficient training methods are used. Until now, research on phonetic training has been mainly limited to auditory or audiovisual perceptual training using natural or synthetic stimuli in a variety of contexts and with a variety of talkers. The findings so far suggest that perceptual HVPT holds great promise as an effective method for the development of phonological categories with a wide range of populations, ranging from L2 learners to cochlear implant users and language-impaired children.

The studies reviewed also point to a very encouraging scenario in the design and application of perceptual training procedures for L2 learners, given the fact that improved perception at post-test tends to correlate positively with performance in production tasks (e.g. Aliaga-Garcia et al., 2010; Hazan & Sennema, 2007; Lengeris & Hazan, 2010). However, only a small number of training studies have explored the effectiveness of using articulatory training to improve the perceptual and articulatory abilities of the L2 learner and those were mainly concerned with non-native consonants. Research on training methods other than perceptual training is still scarce (e.g. Hattori & Iverson, 2009), especially for vowel training. For example, Gómez-Lacabex et al.'s explored the two training methods (2007, 2009) but training was limited to /ə/. To my knowledge, none of the previous vowel studies have compared the impact of **auditory versus articulatory training** audiovisual high-variability including a **large vowel set** on the learners' **perception and production of L2 vowels**.

The present investigation aims at extending this field of research by comparing the improvement in L2 vowel perception and production following two HVPT methods including the full set of English vowels, namely Identification (ID)training and Articulatory (ART) training.

PART II

THE STUDY

Chapter 2

Objectives and Research Questions

This chapter provides a brief summary of the first chapter (Section 2.1) followed by the main objectives (Section 2.2) as well as research questions (Section 2.3) that have motivated this research study and, subsequently, the main hypotheses are formulated (2.4).

2.1. Introduction

The present study is concerned with the acquisition of the English vowel system, more specifically stressed monophthongs (/iːɪeɜːæʌɑːɒɔːʊuː/) by Catalan-Spanish learners of English, and the effectiveness of phonetic training in effecting gains in the perception and production accuracy of these L2 sounds.

The view of the literature in Chapter 1 suggests that L1-Catalan/Spanish learners of L2 English have difficulty perceiving and producing vowels accurately and such difficulties are mainly explained by patterns of cross-language mapping of L2 vowels to L1 vowel categories and by learners' inappropriate use of phonetic cues (over-reliance on duration). Such difficulties in phonological acquisition are assumed by the main L2 speech learning models to have a perceptual basis (*perception leads production* in L2 speech learning). Hence, most training paradigms, whether using auditory or audiovisual materials, including HVPT, have emphasized training on the perception of speech sounds. More recently, however, the efficiency of phonetic training methods based on articulation (articulatory training) have been investigated. Overall this research shows that whether training is based on perception only or includes articulatory training, gains are obtained both in perception and production. A variety of training methods (including signal enhancement, visible articulation cues, high-variability stimulus set, identification task and adaptive training) have proved to be conducive to improvement in L2 speech perception and production and have served as a productive testing ground for general principles of learning and claims about adult neural plasticity.

This study contributes to the field of phonetic training by addressing several gaps in the literature:

(1) This study examines the effect of **articulatory training**, a training methodology that is scarce in the literature. The number of perception training studies showing carry-over effects of improvement in perception to production (e.g. Bradlow et al., 1999; Iverson & Evans 2007, 2009; Lambacher et al., 2005) outnumbers the number of articulatory studies showing improvement in production and perception (e.g. Hattori & Iverson, 2009). In the last decade, research on phonetic training has only focused on a variety of perceptual training methods. This study aims at assessing the perception-production relationship such that improvement in L2 vowel perception and production following AV Identification training is compared against improvement following AV Articulatory training including the whole set of English vowels.

(2) It contributes to the literature on **vowel studies**. The number of consonant training studies outnumbers the number of vowel training studies (e.g. Iverson & Evans, 2007; Nishi & Kewley-Port, 2008; Ylinen et al., 2009). Specifically, audiovisual training

studies have focused on consonants rather than vowels due to their relatively low visual salience (e.g. Hazan et al., 2007, 2009).

(3) It **compares auditory and articulatory training** methods. Studies comparing auditory and articulatory training are scarce and the few studies testing the effectiveness of these two different methods are mostly consonant studies (e.g. Catford & Pisoni, 1970).

(4) It studies the acquisition of all English stressed monopththongs (/iːɪeɜːæʌɑːɒɔːʊuː/ rather than focus on specific vowel targets. Vowel studies comparing auditory versus articulatory training effects have not included a **large set of vowels**, and have been conducted in short-term laboratory training conditions or in FI settings without completely isolating perception from production training (e.g. Gómez-Lacabex, 2009).

The present study examines the effects of phonetic training on two groups of 32 Catalan-Spanish learners of English each receiving a different type of AV HVPT during a five-week period, auditory training and articulatory training, respectively. The L2 learners and a control group of 20 untrained learners were pre- and post-tested through multiple-choice (ID1-task1) and forced-choice vowel categorization (ID2-task2) tasks, a vowel discrimination task (DIS-task3) and a vowel production task (Task4) including the full set of SBE vowels.

2.2. Main objectives

The present study addresses the question of whether improvement in vowel perception and production (including the full set of English vowel monophthongs (/iːɪeɜːæʌɑːɒɔːʊuː/) is best promoted by perception- or production-based AV HVPT.

Further more specific objectives include testing the effectiveness of AV training. Although visual cues to vowel identification are not as salient as cues to consonant identification (e.g. features related to place of articulation), L2 learners may be influenced by visual cues such as lip-rounding degree of opening of the oral cavity and possibly duration in the perception of L2 vowels.

With the aim of extending research in the phonetic training field, this study addresses important questions as to whether it is possible to obtain greater benefits on the perception and production of L2 vowels following ID AV training or, alternatively, ART AV training.

Second, given the NNSs' difficulties directing attention to primary acoustic cues, the study compares the impact of auditory and articulatory training on the relative weighting of acoustic cues. In other words, it seeks to find out whether auditory or articulatory HVPT with natural vowel stimuli can be effective to decrease learners' attention towards secondary duration cues and, consequently, to help them weigh quality cues in a more native-like manner.

Finally, the results of this study have pedagogical implications in that it seeks to make a contribution to EFL teaching practice in the Catalan educational system by suggesting ways of improving L2 learners' perception and/or production of English vowels.

The research questions are presented in the following section.

2.3. Research Questions

This paper addresses two main research questions, each subdivided into several sub-questions:

Research Question 1 (RQ1): *Main effect of training* (T1 vs. T2)

Will AV HVPT lead to changes in the Catalan-Spanish speakers' perception and production of English vowels? That is, is AV HVPT an effective method to improve L2 vowel perception and production?

RQ1.1: To what extent will learners improve in the identification of natural L2 vowels?

RQ1.1.1: Will HVPT be effective in improving mean *correct vowel identification* (ID1-task1)?

RQ1.1.1a: Will the effect of training and identification gains vary as a function of *vowel set*?

RQ1.1.1b: Will the effect of training and identification gains vary as a function of *vowel*?

RQ1.1.1c: Will the effect of training and identification gains vary as a function of *stimulus presentation* (auditory vs. audiovisual)?

RQ1.1.1d: Will improvement in vowel identification be generalized to *new words* and *untrained talkers*?

RQ1.1.1d: Will the effect of training and identification gains vary as a function of *stimuli presentation* (fixed context vs. context variability)*?*

RQ1.1.2: Will HVPT be effective in reducing *mean error dispersion*?

RQ1.2: To what extent will learners improve in the identification of synthesized English vowels (with manipulated duration) (ID2-task2)?

RQ1.2.1: Will HVPT lead to improvement in *mean percent correct identification* of synthesized vowels? Will HVPT be effective in reducing reliance on duration cues and increasing attention to spectral cues for accurate identification of duration-manipulated *tense* and *lax* vowels?

RQ1.2.1a: Is *vowel* a factor explaining the effect of HVPT on correct identification of duration-manipulated tense/lax vowels, and identification gains?

RQ1.2.1b: Will the effect of training and identification gains identification obtained for tense and lax vowels vary as a function of *manipulation of duration*?

RQ1.2.2: Will HVPT lead to a decrease of *duration effect score*?

RQ1.3: To what extent will learners improve in the discrimination of L2 vowel contrasts (DIS-task3)?

RQ1.3a: Will the effect of training on vowel discrimination vary as a function of *vowel set*?

RQ1.3b: Will the effect of training on vowel discrimination vary as a function of *vowel contrast*?

RQ1.3c: Will the effect of training vary as a function of *test condition* (fixed context vs. context variability)?

RQ1.4: To what extent will learners improve in the production accuracy of L2 vowels (Task 4)?

RQ1.4.1: Will HVPT lead to changes in *spectral distances* between contrasting vowels?

RQ1.4.1a: Will *spectral distance changes* vary as a function of *test condition* (fixed context vs. context variability)?

RQ1.4.1b: Will *spectral distance changes* vary as a function of *vowel set*?

RQ1.4.1c: Will *spectral distance changes* vary as a function of *vowel*?

RQ1.4.2: Will HVPT lead to changes in *tense-lax duration ratios*?

RQ1.4.3: To what extent will HVPT affect vowel *height, frontness* and *duration*?

Research Question 2 (RQ2): *Effect of type of training* (ID vs. ART)

Will there be differential gains in L2 vowel perception and production as a function of *type of training* (Identification vs. Articulatory training)?

RQ2.1: Which training method (Identification or Articulatory training) will effect larger gains in L2 vowel identification (ID1-task1)?

RQ2.1a: Are ID and ART types of training differently effective in improving the identification accuracy of the three *vowel sets?*

RQ2.2b: Are ID and ART types of training differently effective in improving the identification accuracy of distinct *vowels?*

RQ2.2c: Will the effect of training and identification gains vary as a function of *stimulus presentation* (auditory vs. audiovisual)?

RQ2.2d: Will training method (Identification or Articulatory training) will be more effective in promoting improvement on *trained* words and talkers, and generalization *new words,* and *untrained talkers*?

RQ2.2e: Are ID and ART types of training differently effective in improving identification accuracy in different *test conditions* (fixed context vs. context variability)*?*

RQ2.2f: Which training method (Identification or Articulatory training) will be the most effective in reducing *mean error dispersion*?

RQ2.2: Are ID and ART types of training differently effective in increasing attention to spectral cues for accurate identification of duration-manipulated tense and lax vowels (ID2-task2)?

RQ2.2a: Are ID and ART types of training differently effective in improving the identification accuracy of duration-manipulated *tense vowels?*

RQ2.2b: Are ID and ART types of training differently effective in improving the identification accuracy of duration-manipulated *lax vowels?*

RQ2.2c: Will training will be more effective in reducing the effect of *manipulation of duration*?

RQ2.2d: Will training method (Identification or Articulatory training) will be more effective in decreasing the *duration effect* score?

RQ2.3: Which training method (Identification or Articulatory training) will effect larger gains in L2 vowel discrimination (DIS-task3)?

RQ2.3a: Are ID and ART types of training differently effective in improving the discrimination accuracy of the three *vowel sets?*

RQ2.3b: Are ID and ART types of training differently effective in improving the discrimination accuracy of distinct *vowel contrasts?*

RQ2.3c: Are ID and ART types of training differently effective in improving discrimination accuracy in different *test conditions* (fixed context vs. context variability)?

RQ2.4: Which training method (Identification or Articulatory training) will effect larger gains in L2 vowel production (Task 4)?

RQ2.4a: Which training method (Identification or Articulatory training) will effect larger spectral distance changes?

RQ2.4b: Which training method (Identification or Articulatory training) will effect larger gains in tense-lax duration ratios?

RQ2.4c: Which training method will effect larger changes in vowel *height, frontness* and *duration*?

2.4. Hypotheses

In relation to **RQ1.1** and **RQ2.1**, improvement in vowel identification (ID1-task1) from pre-test to post-test is predicted to occur after both types of HVPT. However, the extent to improvement is expected to vary across training groups and types of segments (e.g. vowel sets and vowels) due to the nature of the participants' predicted L1-L2 vowel similarities. Besides, improvement on trained words is expected to be greater than in new words, although improvement on untrained words and untrained talkers might also occur as a consequence of the high-variability nature of the two types of training administered. Due to the audiovisual nature of the training, improvement in audiovisual perception is expected to exceed auditory perceptual abilities despite overall improvement in both types of stimuli.

Assuming there is improvement in vowel identification from pre-test to post-test after both types of training, the expectation is that ART training may lead to perception gains (as in Kartushina et al., 2015). Although ART training does not include a training task specifically based on vowel identification with vowel sets as multiple responses as the ID training does, it must be noticed that ART training includes auditory feedback based on self-perception, comparison with NS productions, repetitions and continuous articulatory adjustments based on the trainees' own perception of errors and differences. Regarding the effectiveness of the training methods, it is possible that both types of training will be similarly effective in improving vowel identification. Both ID and ART training might similarly promote generalization to new words, contexts and talkers.

Regarding **RQ1.2** and **RQ2.2**, improvement in the identification of synthesized vowels (with manipulated duration) (ID2-task2) is expected to occur after two types of HVPT including all English monophthongs that might force participants to use more native-like spectral cues for achieving more accurate perception (in the case of ID training) or more native-like production (in the case of ART training). After training including high variability of natural vowels, contexts, words and talkers (rather than synthesized stimuli), trainees are expected to apply what they have learned to the tokens that have been durationally manipulated (in support of Fox & Maeda (1999). This study hypothesized that HVPT may shift non-native listeners' perceptual strategy from heavy reliance on duration cues to less heavy reliance, while placing greater reliance on spectral cues. However, it is uncertain which type of training will promote lesser reliance on duration cues and more attention to spectral cues for accurate identification of duration-manipulated tense and lax vowels.

Regarding **RQ1.3** and **R2.3**, it is possible that ID and ART trainees will obtain better discrimination results at post-test (RQ1.3), and it is possible that both types of training will be similarly effective in improving discrimination performance (RQ2.3). Given previous studies claiming the superiority of identification training over discrimination training for improving perception (e.g. Jamieson & Morosan, 1985), especially for improving identification of trained words (Carlet & Cebrian, 2015; Carlet, 2017), it is expected that identification training –which requires the comparison of presented stimuli with pre-existing memory representation of sounds and the labeling of sounds– will improve the subjects' discrimination abilities to discern differences between two physically present categories. On the other hand, assuming that production-based training leads to improvement in perception, ART trainees are expected to improve discrimination too. Both types of training might similarly promote better vowel discrimination at post-test.

Lastly, assuming there is improvement in perception from pre-test to post-test, the expectation is that some positive changes will occur in vowel production after HVPT (**RQ1.4** and **RQ2.4)**. Despite inconstant findings in the literature regarding training effects on production accuracy (Aliaga-Garcia & Mora 2007; Nobre-Oliveira 2007; Wang 2008; Lengeris 2008; Lacabex & Lecumberri 2010; Aliaga-Garcia 2013), there is also previous support that perception training may lead to production gains (Rato & Rauber, 2015) even if to different degrees as a function of vowel and vowel set, whether to a full or limited extent (Hazan et al., 2005; Lambacher et al., 2005; Iverson et al., 2012; Thompson, 2011; Thomson & Derwin, 2014). It seems that high-variability ID training may be effective in raising awareness of vowel spectral dissimilarities, which in turn could transfer to production gains. On the other hand, it is hypothesized that ART training with repetition self-corrective feedback will probably lead to some positive outcomes in production as it provides learners with opportunities to compare their production of L2 sounds to that of native speakers (the target). To sum up, it is believed that both ID and ART training will have somewhat a positive effect on spectral distances. As regards the relative efficiency of each type of training, ART is expected to be more effective than ID training in leading to spectral changes. However, but it is uncertain to what extent any type training will affect the subjects' reliance on the duration cue for vowel production. On the one hand, it may be possible that if training redirects the trainees' attention to spectral differences, they will consequently rely less on duration for vowel production. On the other hand, duration may be affected to a greater extent than spectral distances after training.

Finally, we hypothesize that identification and production-based HVPT may not be modality-specific as previous evidence has suggested. We expect to find some transfer of production to perception as well as some transfer of improved production to perception.

Summary

This chapter has discussed the aims and the main research questions of the present study. Two research questions were posed, effects of audiovisual high-variability training and differential effects of the type of training (auditory or articulatory training). Each research question was divided into four main areas concerning the four tests administered at pre- and post-test.

Chapter 3

Methodology

This chapter presents the experimental study. In order to answer the main research question (RQ1; Section 2.3), a pretest-posttest experiment was designed that included a battery of perceptual and production tasks to assess improvement in vowel perception and production following identification (ID and articulatory (ART) training. Perception improvement was assessed through two identification tasks consisting of natural and synthesized stimuli, respectively, and a discrimination task with natural stimuli. Improvement in production accuracy was assessed through a delayed repetition task with natural stimuli.

In this chapter we describe the experimental research design (section 3.1), the participants (section 3.2), speech materials (including the training and testing stimuli, tests, types of training and measures) (section 3.3). Finally, an analysis section is included to describe how the different perceptual and production tasks have been analyzed to measure changes in vowel identification, discrimination and articulation after training (section 3.4).

3.1 Experimental design: a pre-test/post-test experiment

The main question addressed in the current study is whether audiovisual identification (ID) and articulatory (ART) high-variability training (HPVT) would have a positive effect on the learners' perception and production of English vowels, and which type of training would be more effective. For this reason, the general design of this study included (see Table 3.1) (e.g. Logan et al., 1991): (1) a pre-test, (2) a five-week training phase when learners received either identification or articulatory training (3) a post-test.

The vowel perception and production performance of (trained) participants was measured before and after 10 training sessions using a battery of perceptual and production tasks. The purpose of having a control (untrained) group completing the same pre- and post-test tasks was to verify whether each of the training groups' performance improved more than that of the control group at post-test, which would help evaluate the effectiveness of the identification and articulatory training methods. A native-speaker control group provided a baseline for the pre-test and post-test tasks used to assess the effectiveness of training. This baseline was also used as reference acoustic values forgauging learners' amount of improvement.

A subset (*N*=64) of the L2 learners (8 males and 56 females; *mean age*= 22.3; *range*= 19-50) enrolled in a phonetic training programme and were given course credit for their participation. They were randomly assigned to two groups, one received identification training (ID, *N*=32) and the other received articulatory training (ART, *N*=32). The remaining L2 learners (*N*=20) served as (volunteering) controls, that is, they performed the pre- and post-tests but received no training (Table 3.1).

The Southern-British English (SBE) speakers completed the same perception and production tests as the L2 learners, and thus provided baseline data. They were paid for their participation and were also given a reward in the form of small refreshment upon completion of all the tests. The L2 learners were tested and/or trained in the University of Barcelona, in Barcelona, whereas the SBE baseline group were tested at University College London (UCL).

	Participants	**Pre-test (T1)**	**Training**	**Post-test (T2)**	**Predicted Proficiency at T2**	***N***
Learners	Experimental group I	✔	Identification (ID)	✔	Second best	32
	Experimental group II	✔	Articulatory (ART)	✔	Second best	32
	Control group	✔	-	✔	Poorest	20
Native SBE speakers	Baseline group	✔	-	✔	Best	10
Total						84

Table 3.1. An overview of the pretest/posttest experimental design with four groups of participants: two experimental groups (ID and ART training), a NNS control group and a NS baseline group.

3.2. Participants

The subjects in the present study were a group of adult Catalan-Spanish advanced learners of English (*N*=84; 12 males, 72 females) as a foreign language (L2 learners) in their second and third year of a degree in English studies at the University of Barcelona. A group of SBE speakers (*N*=10; 5 male, 5 female) provided baseline data. All participants reported having normal hearing and having no speech-related dysfunctions.

The learners were Catalan-Spanish bilinguals living in Barcelona and neighbouring areas. They were exposed to English at university on a daily basis and had a relatively homogeneous advanced proficiency level. They reported having begun learning English when they were 9-11 years old (*AOL*= 10 years). As it is common in the Catalan school context, the focus of the English as a foreign language teaching they had received was on reading and grammar, and all reported having difficulties with English pronunciation. However, they had had one prior semester of formal study of English phonetics at the time they were tested but had not studied English abroad for longer than 2 months. Most of them reported being highly motivated and interested in improving their pronunciation, according to a background language questionnaire they completed before testing.

The SBE speakers lived in London except for one male speaker of the same Southern variety of English residing in Barcelona but having spent most of his life in the London area. All SBE speakers were monolingual (except one) and had grown up in the south of England. The only bilingual English/Spanish native speaker had in fact little knowledge of Spanish, no knowledge of Catalan, and reported using English most of the time (>95% use) on a daily basis as he was an English teacher in Barcelona. The rest were Master students or English teachers in London, with no exposure to native Spanish pronunciation. Native English controls were aged between 31 and 47 years old (mean 36.5 years).

3.3. Speech Materials

Speech materials included (a) **testing materials** presented to controls and trained learners before and after the training period and (b) **training materials**. The English vowel stimuli in the training materials were identical for ID and ART training groups. Half of the training stimuli were used in the testing stimuli (*trained* words), the other half of the testing stimuli were *untrained* words, which were included in order to test for training generalization effects to novel native-speaker voices, contexts and new tokens, as well as robustness of learning. All the stimuli in the pre- and post-tests and training tasks were recorded in the same manner in an anechoic chamber at UCL (London).

3.3.1. Training stimuli

Training stimuli were produced by 10 native English talkers from Southern England (5 female) with SBE pronunciation. They were asked to read a randomized list of words twice, presented one at a time on a computer screen to avoid list-reading intonation. They were instructed to read the words carefully at normal speed and on a falling intonation. The word list reading was preceded by a short practice section, consisting of a few practice items, and the researcher verified the recorded stimuli to ensure that the stimuli had been produced as intended.

The speaker was positioned at ear level at a distance of 1 metre from the computer screen and 2 metres from the video-camera.The speaker's face was set against a blue-background and illuminated with a key and a fill light (see Figure 3.1). Unlike other training studies showing the speaker's face fully visible within the frame (e.g. Hazan et al., 2005; Massaro, 1998), in this study only the nose-mouth region was made visible. This was to done to prevent trainees from getting distracted by speakers' eye movements and to enhance their focus of attention on the visible movement of the articulators (i.e. the jaw, lips, degree of openness and tongue movement).

The utterances and the native speaker faces were video-recorded using a Canon XL-1 DV camrecorder, and a Bruel and Kjaer type 4165 microphone, in an attenuated chamber in the Phonetics Laboratory at UCL. The resulting videoclips were edited so that the start and end frames of each token, lasting 3 seconds, showed a neutral facial expression. The words at the beginning of the list were excluded to avoid list-reading effects. The most intelligible best tokens from each talker were selected for the experiment on the basis of auditory judgment and spectrographic analysis.

Figure 3.1. Screenshot of one audiovisual training stimuli showing the talker's face.

The stimuli chosen for the audio-visual training corpus comprised 173 different CVC words comprising the whole set of English monophthongs in a large variety of consonantal contexts (b_d, b_t, d_d, d_n, f_l, h_l, k_n, l_n, m_d, p_k, p_t, s_d, t_k, t_n, w_t) (Table 3.3.1) and pronounced by 10 SBE speakers. Therefore, the vowel stimuli used can be said to be phonetically highly variable due to the high number of contexts and speakers they came from. Different CVC contexts created minimally contrastive sets of real words for all the target English vowels. Vowels were grouped in vowel subsets established on the basis of perceptual similarity and confusability. Vowel set 1 was comprised by four high-front vowels: /i: ɪ e ɜ:/; vowel set 2 consisted of four low vowels: /æ ʌ ɑ: ɒ/; and vowel set 3 included three back vowels: /ɔ: ʊ u:/) (Table 3.2). The multiple-choice responses during the training corresponded to one of these: Vowel set 1: *bead-bid-bed-bird*; vowel set 2: *mac-muck-Mark-mock*; vowel set 3: *Paul-pull-pool*) (e.g. Iverson & Evans, 2007, 2009).

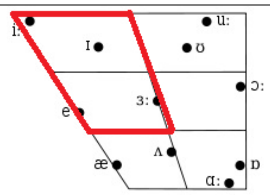
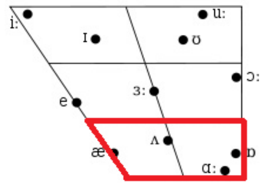
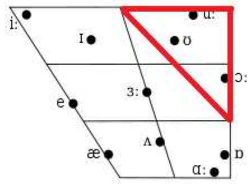
Vowel subset		vowels	CVC contexts	tokens
(1) High front		iː	16	64
		ɪ	16	
		e	16	
		ɜː	16	
(2) Low		æ	16	64
		ʌ	16	
		ɑː	16	
		ɒ	16	
(3) Low back		ɔː	15	45
		ʊ	15	
		uː	15	
Vowel stimuli (Total)				173

Table 3.2. Training set of natural vowel recordings.

3.3.2. Testing stimuli

The testing corpus consisted of 83 CVC stimuli containing the 11 English vowels in different contexts (/b_n, b_s, d_l, f_b, f_d, g_d, h_d, h_m, k_p, l_d, l_k, m_l, p_m, r_t, s_d, s_k, s_ts, ʃ_d, ʃ_t, t_g, t_t, w_d, w_f, w_l/) pronounced by 2 novel SBE speakers (1 male, 1 female). The 83 CVC stimuli consisted of 47 *trained* words and 36 *untrained* words (Table 3.3). Trained words converged in the two sets of stimuli (training and testing stimuli) but differed in talker, since pre- and post-tests aimed at comparing the learners' performance on *trained* and *untrained words.*

The testing stimuli were recorded, segmented, edited and selected following exactly the same procedures used for the training stimuli (See previous section 3.3.1). Finally, the 83 video files (audiovisual stimuli) were converted to sound, creating a new subset of 83 auditory CVC stimuli so that each testing stimulus could be presented auditorily and audiovisually in the perceptual and production tasks. CVC stimuli were used to test vowel perception in *context variability*.

Natural /h/-V-/d/ words (e.g. /hiːd, hɪd/, /hæd/-/hʌd/) from the testing stimuli set had a twofold purpose. On the one hand, these were used in the same way as other words in the identification, discriminination and delayed repetition tasks to test vowels in the *fixed context* condition. On the other hand, they were also used to build an identification task with synthesized vowels (with manipulated duration) (ID-2), for which /iː ɪ æ ʌ ɑː ʊ uː/ tokens were lengthened and shortened using Praat in order to assess to what extent learners relied on duration to contrast tense-lax vowel pairs.

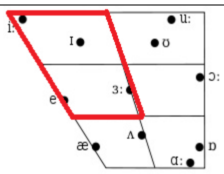
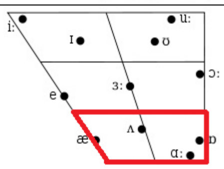
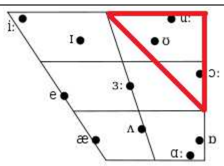
Vowel subset		**Vowels**	**CVC Trained contexts**	**CVC Untrained contexts**	**words (total)**
(4) High front		iː	4	3	28
		ɪ	4	3	
		e	4	3	
		ɜː	4	3	
(5) Low		æ	4	3	28
		ʌ	4	3	
		ɑː	4	3	
		ɒ	4	3	
(6) Low back		ɔː	5	4	27
		ʊ	5	4	
		uː	5	4	
			trained words	**untrained words**	
Vowel stimuli (Total)			47	36	83

Table 3.3. Testing set of natural vowel recordings

3.3.3. Testing tasks and procedures

The present study assessed L2 learners' improvement between testing times in (1) the identification of vowels, (2) the identification of synthesized vowels (with manipulated duration), (3) the discrimination of vowel contrasts and (4) the production of natural vowels. ID and ART training groups, as well as the control group and the SBE baseline group took part in the same tests. The perception and production tasks, and the testing procedure, are described below.

The test battery included three perception tasks and one production task which participants performed in the following order (see Table 3.4):

(1) Identification of natural English vowels (ID1-TASK1)

(2) Identification of synthesized English vowels (manipulated duration) (ID1-TASK1I)

(3) Discrimination of natural English vowels (DIS)

(4) Delayed repetition task (imitation) (DR)

Domains	Tests	Vowel stimuli	Word type	Stimulus presentation	Test condition	Speakers	vowels	Task duration
Perceptio	Identification task I	natural	trained untrained	AV/A	Varied CVC	2 (1m, 1 f)	11	30 min
	Identification task II	synthesized	untrained	A	Fixed (/hVd/)	1 (f)	7	10 min
	Discriminatio task	natural	untrained	A	Varied / fixed	8 (4 m, 4 f)	11	15 min
Productic	Delayed Repetition tas	natural	trained untrained	AV/A	Varied / fixed		11	30 min
Pretest and Posttest (total dur.)								1h 30 min

Table 3.4. An overview of the battery of perception and production tasks used at pre-test and post-test.

3.3.3.1. Perception tasks

Subjects were pre- and post-tested on the identification of auditory and audiovisual stimuli (ID1-TASK1), duration-manipulated audiovisual stimuli (ID1-TASK1I), and on the discrimination of vowel contrasts. Stimuli included trained tokens as well as novel tokens produced by new talkers to test for generalization effects of training.

3.3.3.1.1. Identification task I (natural stimuli) (ID1-task1)

Identification task I included the whole English vowel set. It was a multiple-choice categorization task lasting 30minutes. The stimuli set consisted of 264 CVC words, which included 15 trained and 15 untrained words distributed into auditory and audiovisual blocks with 3 tokens x 11 vowels x 2 repetitions each (corresponding to 2 new SBE talkers, 1 male and 1 female) (see Table 3.5).

Condition	**Word type**	**vowels**	**Tokens**	**Talkers**		
A	trained	11	3	2	66	132
	new	11	3	2	66	
AV	trained	11	3	2	66	132
	new	11	3	2	66	
Stimuli / Trials (total)		11	3	2		264

Table 3.5. Multiple-choice categorization task (ID1-task1) with auditory (A) and audiovisual (AV) conditions, and trained vs. new natural tokens.

ID1-TASK1 therefore provided a measure of trainees' improvement in how accurately they could identify L2 vowel stimuli in terms of L2 vowel categories while ignoring variability dues to context and talker. Their vowel identification scores provide a measure of how well-developed (target-like) and robust their L2 vowel representations are and whether they improve as a result of training.

In order to assess learners' reliance on acoustic and/or visual cues for vowel recognition, this identification test included two types of stimuli presentation in a variety of contexts: audiovisual (AV) and auditory (A), and two counterbalanced presentation orders (A-AV and AV-A) (e.g. Hazan et al., 2005; Iverson & Evans, 2007).

In order to test generalization effects, the test included a *fixed context* (/h_d/) and a *context variability* test condition. These stimuli were presented in fully randomized blocks.

Participants were required to watch a 3-second videoclip showing the talkers' mouth (AV condition) or, alternatively, hear the stimuli (A condition), and then immediately click/select one word out of 3-4 options displayed vertically on the computer screen (e.g. *feel-fill-fell-furl*, *cat-cut-cart-cot*, *Paul-pull-pool*). Together with each response word, another common word was also given (see Figure 5), in case the response word was unfamiliar to the participants (e.g. Iverson & Evans, 2007; Iverson, Pinet, & Evans, 2011; Lengeris, 2009). A practice set preceding the task included 5 trials to ensure participants understood the task structure and functioning.

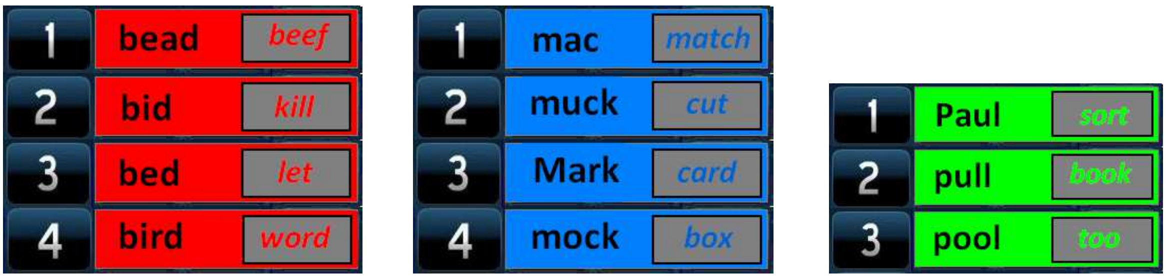


Figure 3.2. Button responses distributed in vowel sets in the multiple-choice categorization task (ID1-task1).

3.3.3.1.2. Identification task II (synthesized stimuli) (ID1-TASK1I)

Identification task II was a 20-minute forced-choice ID task with 280 synthesized /hVd/ stimuli. These included 5 repetitions of 8 different vowels produced by a female speaker, each presented with 7 different durations (80, 116.7, 153.3, 190, 226.7, 263.3, 300 ms). The vowel /ʌ/ was used twice, in order to test two low vowel contrasts: /æ/-/ʌ/, and /ʌ/-/ɑː/ (Table 3.6).

Vowel contrast		Vowels	Duration continua	Duration steps	Repetitions	Tokens
C1	High-front	/hiːd/-/hɪd/	2	7	5	70
C2	Low	/hæd/-/hʌd/	2	7	5	70
C3	Low	/hʌd/-/hɑːd/	2	7	5	70
C4	High-back	/hʊd/-/huːd/	2	7	5	70
Total		/iː, ɪ, æ, ʌ, ɑː, ʊ, uː/7	8	7	5	280

Table 3.6. Forced-choice categorization task of synthesized vowels (ID2-TASK2-task2) based on 8 vowel continua consisting of 7 duration steps (80, 116.7, 153.3, 190, 226.7, 263.3, 300 ms).

Duration was manipulated using Praat 4.2.21 software (Boersma & Weenink, 2004). Randomized stimuli were presented auditorily (A) in four randomized blocks, one for each vowel contrast (/hiːd/-/hɪd/, /hæd/-/hʌd/, /hʌd/-/hɑːd/ and /hʊd/-/huːd/, with 70 trials per block. Stimuli and blocks were presented in different random orders to every participant.

The duration-manipulated stimuli used in this task were drawn from 7 synthetic vowel duration continua: two high-front vowel continua (C1: /hiːd/,/hɪd/), four low vowel continua (C2: /hæd/-/hʌd/; C3 /hʌd/-/hɑːd/) and two high-back vowel continua (C4: /hʊd/-/huːd/), which meant a total of 7 duration-manipulated versions of the same /hVd/ word.

The eight continua were created using Praat. The vowel durations of /hiːd/, /hɪd/, /hæd/, /hʌd/, /hɑːd/, /hʊd/ and /huːd/ were first measured (Table 3.7). Further, the duration of each vowel in each member of a minimal pair was manipulated so that it corresponded to the duration of the other member of that minimal pair, and viceversa. Thus, each stimulus had 7 versions, none of which had the original vowel duration value. For example, the duration of /ɪ/ was manipulated so that half of the stimuli on the continuum corresponded to the normal or exaggerated duration of /iː/, and viceversa.

Vowel continua	Original natural stimuli		Unmodified duration (ms)	Manipulation of duration (ms)
C1	High-front	V1 tense /hiːd/	446.5	80, 116.7, 153.3, 190, 226.7, 263.3, 300
		V2 lax /hɪd/	235.7	80, 116.7, 153.3, 190, 226.7, 263.3, 300
C2	Low	V3 tense /hæd/	331.2	80, 116.7, 153.3, 190, 226.7, 263.3, 300
		V4 lax /hʌd/	221.8	80, 116.7, 153.3, 190, 226.7, 263.3, 300
C3	Low	V5 lax/hʌd/	221.8	80, 116.7, 153.3, 190, 226.7, 263.3, 300
		V6 tense /hɑːd/	412.9	80, 116.7, 153.3, 190, 226.7, 263.3, 300
C4	High-back	V7 lax /hʊd/	194.8	80, 116.7, 153.3, 190, 226.7, 263.3, 300
		V9 tense/huːd/	355.0	80, 116.7, 153.3, 190, 226.7, 263.3, 300

Table 3.7. Natural and modified duration (ms) of the vowels used in the identification task II, and vowel continua.

In each continuum, the formant frequencies were unmodified but duration was modified in 7 steps (80, 116.7, 153.3, 190, 226.7, 263.3, 300 ms) for the eight /hVd/ tokens. The values 80ms (Step 1) and 300ms (Step 7) were chosen as endpoints and stood for "exaggerated" duration in relation to prototypical lax (e.g. /ɪ/) and tense vowels (e.g. /iː/), respectively, to allow for 7 duration steps (36.7ms per step). Both the endpoints and the intermediate steps were implemented by adding or subtracting 36.7 ms to subsequently manipulated stimuli.

This identification task was designed to determine whether training had a positive effect on L2 learners' cue weighting. The main aim was to assess learners' vowel identification when the duration cue was made ambiguous (e.g. when /iː/ was /ɪ/-like in duration and vice-versa). The weighting of duration as a cue was tested on the four minimal pairs (see Table 3.8.)

Participants heard one duration-manipulated vowel stimulus in a fixed context (/h_d/) at a time and had to identify one of two response options by using the keyword to click on a button with the response words written in English orthography (Figure 3.3),one on the left, another on the right (e.g.*he'd* vs. *hid*). The task was preceded by a short practice period including four trials to familiarize participants with the procedure.

Figure 3.3. Button responses per block in the synthesized vowel identification task (ID2-TASK2-task2).

3.3.3.1.3. Discrimination task (natural stimuli) (DIS)

In this AX discrimination task participants were presented with 128 CVC word pairs based on 16 vowel contrasts produced by 8 untrained SBE talkers (4 male, 4 female) (See Table 38). 25% of the trials were false-alarm trials. The vowel pairs selected varied in their degree of potential difficulty in discrimination according to previous vowel identification data (Cebrian, 2006; Fabra & Romero, 2012; Iverson & Evans, 2007; Mora, 2005; Rallo Fabra, 2005) and perceptual assimilation data (Cebrian et al., 2011). Two testing conditions were included, CVC context variability and fixed context (/h_d/).

Participants were presented with the two members of a vowel contrast with a 3-second inter-trial interval and and were instructed to respond as quickly and as accurately as possible. Participants heard one minimal pair at a time and labelled it as *same* or *different*. They pressed a key corresponding to the response option *same* (left button) or *different* (right button) (Figure 3.4). If no response was given, the next trial was presented after 3 seconds.

Vowel contrast	Vowel contasts	Blocks/contexts	Repetitions	Stimuli (total)
C1	/iː/-/ɪ/	2	2	
C2	/ɪ/-/e/	2	2	
C3	/e/-/ɜː/	2	2	
C4	/ɪ/-/ɜː/	2	2	
C5	/e/-/æ/	2	2	
C6	/ɜː/-/æ/	2	2	
C7	/æ/-/ʌ/	2	2	
C8	/ʌ/-/ɑː/	2	2	
C9	/æ/-/ɑː/	2	2	
C10	/ɜː/-/ʌ/	2	2	
C11	/ɜː/-/ɑː/	2	2	
C12	/ɒ/-/ɑː/	2	2	
C13	/ʌ/-/ɒ/	2	2	
C14	/ɒ/-/ɔː/	2	2	
C15	/ʊ/-/uː/	2	2	
C16	/ɔː/-/uː/	2	2	
Total	16	2	2	128

Table 3.8. AX Vowel Discrimination task (DIS-task3) of 16 natural vowel contrasts.

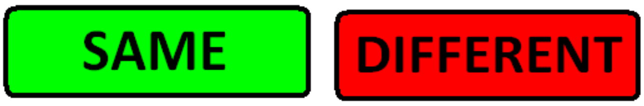


Figure 3.4. Button responses in the AX vowel discrimination task (DIS-task3).

This discrimination task was included to ascertain whether participants had improved in their ability to phonetically distinguish between members of potentially difficult vowel contrasts after training had to say whether the two stimuli were the *same* or not. In order for participants to respond correctly, they had to be able to perceive the difference between the contrasting pair of words. Therefore, the focus was on the perceptual discrimination of two L2 similar sounding sounds rather than identify L2 sounds in terms of their L2 sound representations.

3.3.3.2 Vowel Production task: delayed repetition task

The production task was based on a 30-minute delayed repetition task consisting of 160 CVC words in isolation which included both trained and untrained words that participants had previously encountered in previous perception tasks. Words were distributed into two randomized blocks or conditions with 80 stimuli each. Participants were presented with auditory (A) or audiovisual (AV) stimuli produced by 2 untrained SBE talkers (1 male, 1 female) (Table 3.9).

Stimuli presentation	**vowels**	**Blocks (test conditions)** Fixed context (/h_d/)	CVC contexts	**Repetitions**	**Tokens**
A vs. AV	/iː/	1	1	2	4
	/ɪ/	1	3	2	6
	/e/	1	3	2	6
	/ɜː/	1	4	2	7
	/æ/	1	3	2	6
	/ʌ/	1	3	2	6
	/ɑː/	1	4	2	7
	/ɒ/	1	3	2	6
	/ɔː/	1	2	2	5
	/ʊ/	1	1	2	4
	/uː/	1	2	2	5
Total	11	11	29	2	80
x 2 (A/AV) *blocks*					160

Table 3.9. Delayed repetition task with auditory (A) vs. audiovisual (AV) presentation blocks, trained vs. untrained words and fixed vs. context variability conditions.

An introduction to the production task was given to the participants beforehand in the form of four practice trials produced by a native speaker whose voice was not included in the test trials. Each participant was presented with AV/A CVC stimuli produced in isolation, randomized in four blocks, and was instructed to wait for a sound and visual signal (a beep and a picture of a microphone) ocurring 3000 ms after the stimuli presentation to repeat it (Figure 3. 5). Participants were recorded individually as they repeated each of the trial twice in a computer room using a headset with microphones. Stimuli were recorded directly on computers at a 48-kHz sampling rate and the individual recordings were automatically stored on the hard disk.

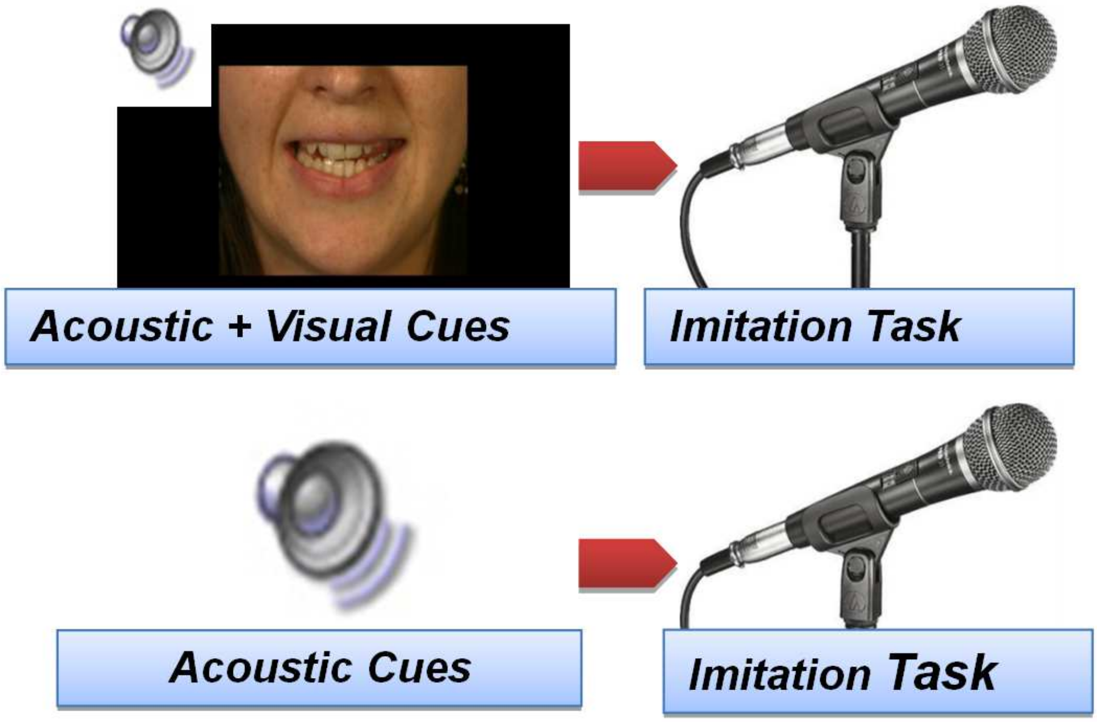


Figure 3.5. Screen design in each of the blocks of the Delayed Repetition production task (task4).

The purpose of this task was to determine whether trained participants produced English vowels more accurately at post-test, and assess how much trainees weighted (primary) spectral and (secondary) temporal cues when they articulated vowels, more specifically if the spectral distances between contrasting vowels were sensitive to change. Vowels were articulated in fixed context and context variability conditions, when they were presented with auditory or audiovisual stimuli for delayed repetition..Once recorded and stored digitally, the recording for all subjects' words in the pre- and post-test were transferred to a laptop, where they were acoustically analyzed using Praat.

3.3.4. *General testing procedures*

Pre- and post-tests took place at the University of Barcelona. All participants were tested in a single session in computer rooms consisting of 20 PCs in quiet conditions, with stimuli played over high-quality headphones (Soyntec Netsound 500) at a user-controlled comfortable level. The same PCs were used to run the four different tests in a session lasting about 1 hour and a half. All perceptual tasks were administered using DMDX software to collect written responses (e.g. key presses). Participants' utterances (in the vowel production task) were also gathered using DMDX and the headset microphones.

Before each of the tasks begun, there was a practice section consisting of at least four trials structured identically to the actual experiment. The practice period was used to familiarize participants with the program, the procedure and adjust the listening level. The subjects never received feedback. Stimuli could not be replayed.

In each test, there was a pause between sections to give listeners a short break. The three perception tests took approximately 60 min to complete. The production test took about 30 min to complete. Participants took a 15-min break between the two parts.

At the end of the pre-test, participants completed a language background questionnaire.

3.3.5. Measures

3.3.5.1. Perception measures

The participants' perceptual phonological competence was assessed by computing mean percent correct vowel identification scores (*percent correct identification*) for natural vowels (ID1-TASK1) for each subject, *type of training* (ID vs. ART), *word type* (trained vs. untrained), *stimulus presentation* (AV vs. A), *vowel set* (high-front, low and high-back) and *vowel* (/iː, ɪ, e, ɜː, æ, ʌ, ɑː, ɒ, ɔː, ʊ, uː/), at pre- and post-test (T1 vs. T2). Changes in the identification of natural vowels (ID1-task1) after training were computed as mean percent identification *gains,* obtained by subtracting pre-test from post-test identification scores. The distribution of errors among response options was measured through a measure of degree of *error dispersion*. This was a score from 1 to 3, depending on the number of wrong responses recorded for every trial and subject. This measure was computed at pre-test and post-test too for *type of training*, word type, *stimulus presentation*, *vowel set* and *vowel. Error* *dispersion* is an independent measure that quantifies the number of error categories per response (e.g. errors in /iː/ identification distributed among /ɪ/ and /e/ at pre-test) which can be used to distinguish response patterns in vowel identification as quantifies for the spreading of the errors over response categories. A measure of pre-test/post-test *differences in error-dispersion* was also obtained, *error dispersion change*, as a measure change (i.e.increase/decrease) in error dispersion per vowel category and stimulus response.

In order to assess improvement in the identification of synthesized vowels (ID2-TASK2), and in particular L2 cue weighting after training, mean percent correct vowel identification scores (*percent correct identification*) were computed for each subject, *type of training* (ID vs. ART), *synthesized vowel* (/i, ɪ, æ, ʌ, ɑ, ʊ, u/), and *manipulation of duration* (stimulus number: step 1, 80; step 2, 116.7; step 3, 153.3; step 4, 190; step 5, 226.7; step 6, 263.3, step 7, 300 ms), at pre- and post-test (T1 vs. T2). An overall percentage of /iː, ɪ, æ, ʌ, ɑː, ʊ, uː/ identification was derived from the proportion of /iː, ɪ, æ, ʌ, ɑː, ʊ, uː/ responses in order to assess the degree of categoriality and general sensitivity to changes in vowel duration along the 4 vowel continua (C1: /hiːd/-/hɪd/, C2: /hæd/-/hʌd/, C3: /hʌd/-/hɑːd/, C4: /hʊd/-/huːd/). To measure the degree of reliance on duration as a cue in vowel identification, a *duration effect score* (DES) was computed. This measure is the difference in mean *percent correct* categorization at the two endpoints of the duration continua, 300 ms and 80 ms. DES was calculated separately for the 8 vowel continua. Pre-test/post-test changes in DES were also calculated by subtracting pre-test from post-test DES scores.

The participants' perceptual phonological competence was also assessed by computing a mean *percent correct* measure of correctly discriminated vowel pairs or overall discrimination (*percent correct discrinination*) for each subject, *type of training* (ID vs. ART), *test condition* (fixed context vs. context variability), vowel set (Vowel set 1: high-front; Vowel set 2: low vowels; Vowel set 3: back vowels) and *vowel contrast* (/iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/, /ɪ/-/ɜː/; /e/-/æ/, /ɜː/-/æ/, /æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɜː/-/ʌ/, /ɒ/-/ɑː/, /ʌ/-/ɒ/; /ɒ/-/ɔː/, /ʊ/-/uː/, /ɔː/-/uː/), at pre- and post-test (T1 vs. T2). Mean percent discrimination *gains* were also obtained by subtracting pre-test from post-test scores. In addition to *percent correct* scores, mean *d-prime* scores of vowel discrimination (MacMillan and Creelman, 1991) were also obtained. *d-prime*(d' = z(H) - z(F)) takes into account the proportion of hits (H: proportion of correct DIS of the minimal pairs) and false alarms (F: proportion of incorrect DIS of distractor same-word pairs) providing an unbiased measure of subjects' discrimination ability.

Training effects on vowel perception were expected to result in overall better identification and discrimination ability (higher identification and discrimination scores), and lower error-dispersion in vowel identification. Post-test results were expected to show greater reliance on vowel quality than duration in vowel categorization (and smaller DES) at post-test.

3.3.5.2. Production measures

Vowel production accuracy was assessed by means of acoustic measurements (e.g. Lambacher et al., 2005; Lengeris, 2009) rather than native speaker perceptual judgments (e.g. Hattori & Iverson, 2009; Iverson et al., 2008; Lengeris, 2008; Lengeris & Hazan, 2009).

Formant measurements were taken to assess accuracy in learners' vowel quality before and after training. Duration measurements were taken to assess the learners' reliance on duration as a cue to contrast vowels at pre- and post-test. We wanted to assess the effect of training, particularly different types of training on both vowel quality and duration.

The duration of each vowel was measured manually from spectrogram and waveform displays with the aid of Praat speech analysis software (Boersma & Weenik, 2009). The duration of a vowel was defined as the period between the onset of vocal cord vibration to the offset of periodic energy visible in the spectrogram. Formant frequency measurements were taken by manually placing the cursor in the middle of the steady state-period of the vowel spectrogram and running a script that would extract the average F0, F1, and F2 formant frequency measurements from a 50-ms window centered around the cursor. The LPC smoothed spectrum display was used only to confirm the location of the steady state formant frequencies estimated by eye from the spectrogram. Frequency values in hertz (Hz) were converted into Bark (B) and normalized using a vowel-intrinsic method (Syrdal & Gopal, 1986) in order to compensate for possible speaker-dependent differences such as gender differences and variations in vocal tract size. The Bark-difference method was chosen because of the possibility of filtering out physiological differences while retaining sociolinguistic differences. Data normalization procedures are described in the following section 3.3.5.3.

To further assess changes in vowel articulation across training groups, additional spectral measures were used. Analogous to other studies in the literature (e.g. Flege, Bohn, & Jang, 1997; Mora, Keidel, & Flege, 2015) mean Euclidean distances or *spectral distances* (SD) between contrasting vowels were calculated, using normalized formant frequency data (Bark) (Flege, Bohn, & Jang, 1997): /iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/, /æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɒ/-/ɑː/, /ʌ/-/ɒ/, /ɒ/-/ɔː/, /ʊ/-/uː/. Subjects showing greater benefits of training were predicted to show less overlap between pairs of English vowels, which would also be expected to result in relatively larger SD scores between contrasting vowels as well as larger differences in vowel *height* (B1-B0) and/or *frontness*-backness (B2-B1), but not necessarily differences in *duration*.

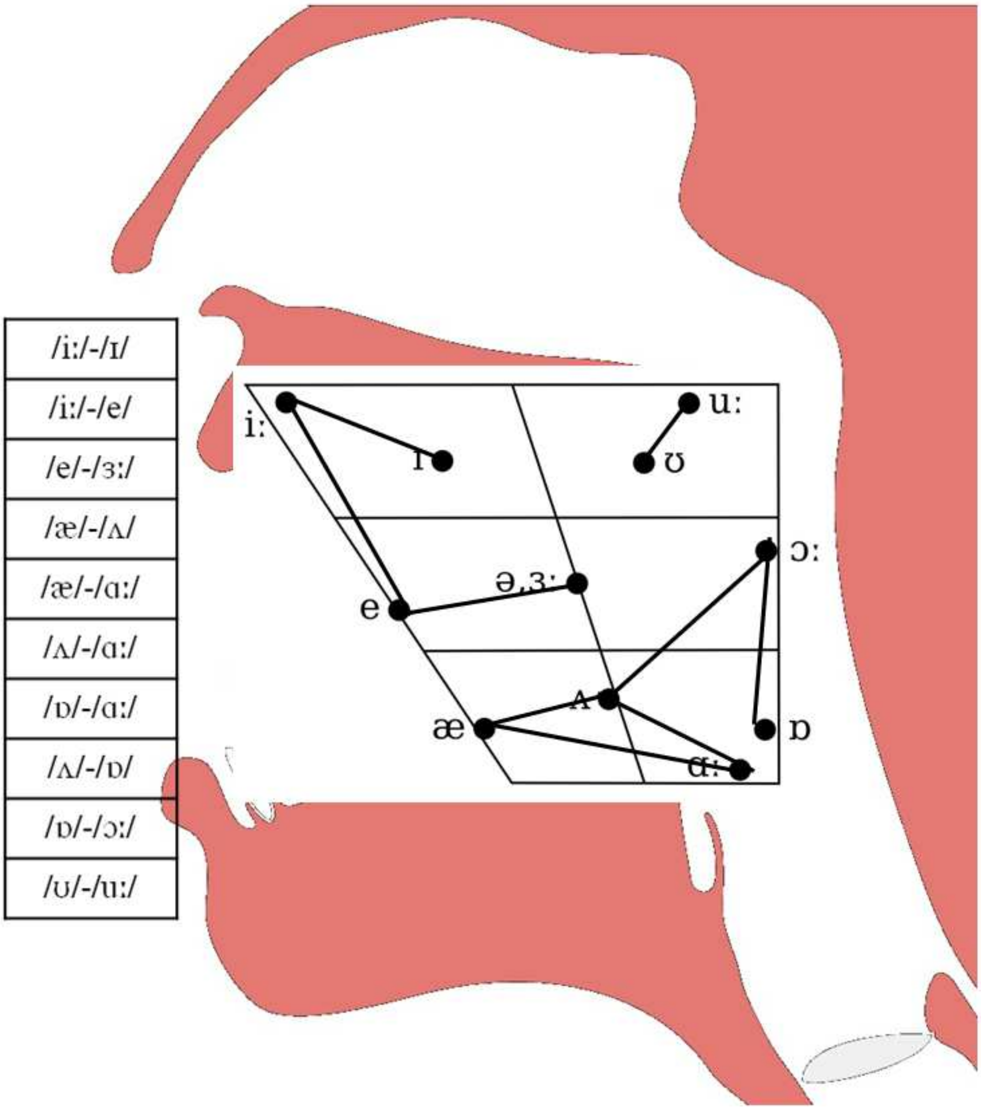


Figure 3.6. Vowel diagram with 10 vowel spectral distances (Bark).

With obtained duration measures for all vowels. Tense-lax *duration ratios* (DR) were calculated for six vowel contrasts:/iː/-/ɪ/, /iː/-/e/, /e/-/ɜː/, /æ/-/ɑː/, /ʌ/-/ɑː/, /ʊ/-/uː/. To obtain this proportional measure, a tense-lax duration ratio (DR), the duration of the tense vowel (/iː/, /ɜː/, /ɑː/, /uː/) was divided by the duration of the lax vowel (/ɪ/, /e/, /æ/, /ʌ/).

3.3.5.3. Data normalization procedures

In order to eliminate variation caused by physiological differences among speakers and minimize the differences between male and female participants, the bark-distance vowel normalization model was used (Syrdal & Gopal's, 1986; Traunmuller, 1997; Adank, Smits & van Hout, 2004; e.g. Baker & Trofimovich, 2005), to provide speaker-independent estimates of vowel height and frontness-backness.F0, F1, F2 frequency values in Hz were converted to Bark (B) using the formula $Z_i = 26.81/(1+1960/F_i) - 0.53$, where $F_i$is the frequency value in Hz for a given formant and Z is the frequency in B. This vowel-intrinsic method has been reported as successful in removing effects of gender while retaining the vowel main effects and interactions revealed by ANOVAs conducted on transformed Hertz values (Bohn & Flege, 1990, 1992).

The acoustic measures derived from the Bark-converted frequency values were vowel *height* and *frontness-backness*, estimated by subtracting the mean B0 (Bark-converted fundamental frequency or f0) from the mean B1 values (Bark-converted first-formant frequency or F1) (B1-B0) and B1 from B2 values (Bark-converted second-formant frequency or F1) (B2-B1), respectively. The vowels were then plotted in the acoustic space with B2-B1 values on the X-axis (where lower values represent back vowels and higher values represent front vowels) and B1-B0 values on the Y-axis (where lower values represent high vowels and higher values represent low vowels) (Baker & Trofimovich 2005; Bohn & Flege, 1990; Baker & Trofimovich, 2005; Flege, Bohn & Jang, 1997).

3.3.6. High-variability Audiovisual Training on Vowels: Stimuli and procedure for training

Separate groups of Catalan-Spanish learners of English (*N*=64) received different types of AV HVPT (natural CVC words from multiple talkers), namely identification (ID) (*N*=32) and articulatory (ART) training (*N*=32), on the 11 English RP monophthongal vowels.

Method

Our training method is **a variant of the HPVT approach** developed by Logan et al. (1991) and which has been subsequently used and proved effective for EFL learners with different L1s (Livelyt et al. 1994, Akahane-Yamada et al. 1997, 1998; Strange et al., 2001; Lambacher et al., 2005; Nishi, 2007; Nobre-Oliveira, 2007; Iverson & Evan, 2007, 2009; Aliaga-Garcia, 2007)**;** Lengeris & Hazan, 2010; Hu W et al., 2016). Pioneering HVPT studies (Logan, Lively, & Pisoni, 1991; Lively, Logan, Pisoni, 1993) using multiple exemplars of different words produced by different talkers have shown significant improvement in the identification of nonnative speech contrasts, and have also proved effective for reallocating the learners' attention to the critical acoustic cues in L2 vowel identification. Subsequent HPVT studies have shown that improvement generalizes to new stimuli and talkers, and that perceptual gains are retained over three months (Bradlow et al., 1999; Pisoni et al., 1994; Carlet & Cebrian, 2015). Moreover, gains can also be transferred to speech production (Bradlow et al., 1997; Iverson & Evans, 2007, 2009; Lengeris & Hazan, 2010; Hu W et al., 2016). More recently, Ylinen et al. (2009) has shown that HVPT can also be effective to help learners focus their attention on the relevant acoustic cues for accurate vowel perception and production.

Number of training sessions

The HVPT training comprised **10 computer-based sessions** (e.g. Ylinen et al., 2011) **over 5 weeks**, which lasted **50-60 minutes**, although in most vowel training studies training the number of training sessions was inferior. In vowel training research including the *full set* of vowels, the training consisted of only 4 sessions (Carlet & Cebrian, 2014; 5 (Iverson & Evans, 2009; Carlet , 2017) and 6 weekly sessions which lasted 45 min. (Iverson & Evans, 2007; Lengeris, 2009), or 8 45-min sessions over 1-2 weeks (Iverson et al., 2011), with one different NS talker in each session, as it is typical of HPVT. Vowel training studies with a small *subset* of English vowels (e.g. Lambacher et al., 2005) comprised 6 weekly training sessions. In Nishi and Kewley-Port (2007) there were 9 training sessions dealing with 3- or 9-vowel sets. In Gomez-Lacabex, Garcia-Lecumberri and Cooke (2007), the unstressed vowel /ə/ was the main focus during 12 training sessions.

The entire course of our training was completed over 5 weeks, with 2 weekly sessions which typically lasted about 60 min. Although one different talker per sessions is typical of HVPT (e.g. Logan et al. 1991; Iverson et al. 2011), there were **two different talkers per each session** (i.e. NS1 blocks 1-3, N2 blocks 2-4).

Input quantity

Each training session guaranteed exposure to a minimum of 176 trials (4 tokens x 11 vowels x 4 repetitions or blocks) and maximum 528, depending on the number of repetitions provided in the feedback section. All sessions included the three *subsets* of vowels (high front vowels /iː ɪ e/, high back vowels /ɔː ʊ uː/, and low vowels /æ ʌ ɑː/) as pronounced by 2 different NS talkers (i.e. NS1 blocks 1 and 3; NS2 blocks 2 and 4). That would mean that participants would have been presented with 1760-5280 AV stimuli for identification or repetition by the end of the training.

Full set of vowels

Stimulus words included the full set of English monophthongs (as in Aliaga-García, 2011; Iverson & Evans, 2007, 2009; Lengeris & Hazan, 2010). Some researchers have reported the negative influence of using smaller training sets on later learning (Nishi, 2007) and some recent training experiments have suggested the advantage of training on a full set (Iverson & Evans, 2007, 2009; Lengeris & Hazan, 2010). The idea behind including a large set of vowels for our study, rather than a small set of "difficult" vowels, is that HVPT on a *fullset* is believed to be more efficient for the following reasons:

(1) A more complex type of training requires multiple-cue judgment rather than exemplar-based judgment.

(2) It increases learners' ability to generalize improvement to new (untrained) tokens, contexts and talkers.

(3) It prevents uneven learning of untrained vowels.

(4) It resembles more real-world learning, full of irrelevant variability.

Audiovisual input

Vowel stimuli used in training were presented audiovisually (e.g. Hazan & Sennema, 2007; Pereira, 2013). Therefore, visual information about lip-shape, tongue position and openness was provided in a high variety of vowel stimuli, contexts and talkers. Although there only few AV training studies for non-native vowels (Flege, 1989; Ortega-Llebaria et al., 2001), there is evidence suggesting that L2 learners can be heavily influenced by lip-rounding and openness features.

Training procedures

Both ID and ART training sessions were designed and run using DMDX software. The 10 sessions were administered twice a week in a quiet computer room, with stimuli played over headphones at a user-controlled comfortable level, and computers (PC) producing the stimuli and collecting responses. Vowel production recordings were also made in these quiet PC rooms with a unidirectional headset microphone Soyntec550.

The different training procedures varied depending on whether participants underwent ID or ART training. Each of the training methods is described in the following sections.

2.3.4.1. Identification (ID) training

Identification (ID) training was based on a vowel identification task similar to the *Identification task I (ID1-TASK1)* described in section 3.3.3.1.1.and sessions consisted of different parts.(see Figures 3.8-3.9):

(1) AV presentation of stimuli and identification

On each trial, subjects heard a stimulus word and clicked on three of four minimal-pair alternatives (depending on vowel set). For example, they could hear *beat* and be asked whether it sounded like *beat, bit, bet or Bert.* The stimulus word was presented audiovisually before the response alternatives were shown, so that responses were open set rather than primed by the response alternatives visible (e.g. Iverson & Evans, 2007).

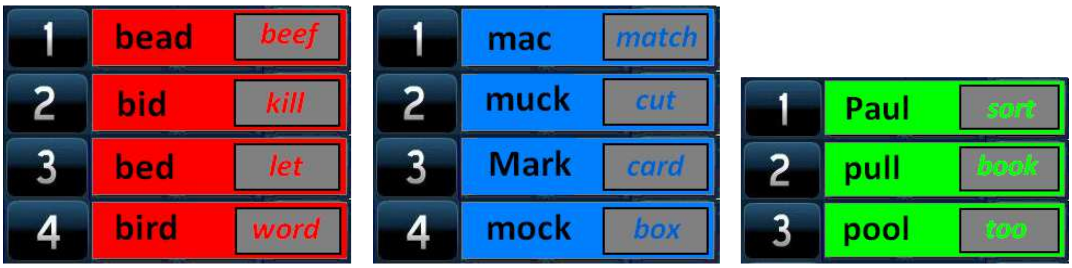


Figure 3.7. Button responses distributed in vowel sets in ID training task (ID training).

(2) Feedback

Immediate acoustic and written feedback was provided after each response. The type of feedback ensured learners would correct their own mistakes and made the training sessions adaptive to the learners' individual needs and difficulties.

Correct responses were signalled with a cash register sound and then participants heard the word again as the correct response/button was simultaneously highlighted. Errors were accompanied with a "cross" sign on top of the "wrong" button pressed and a tone with a descending pitch showed the incorrect response.

(3) **Repeated presentation of stimulus in case of misidentification**

Negative feedback obliged participants to try once again. Participants immediately heard the incorrect response immediately followed by the correct word played again and a written message saying *"Try again"* indicated the stimulus word would be heard again and the alternative responses would be displayed again. Trainees could hear each stimulus a maximum of four times.

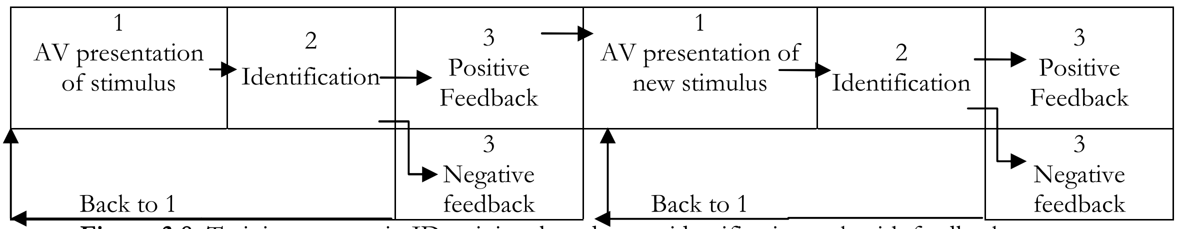


Figure 3.8. Training process in ID training, based on an identification task with feedback.

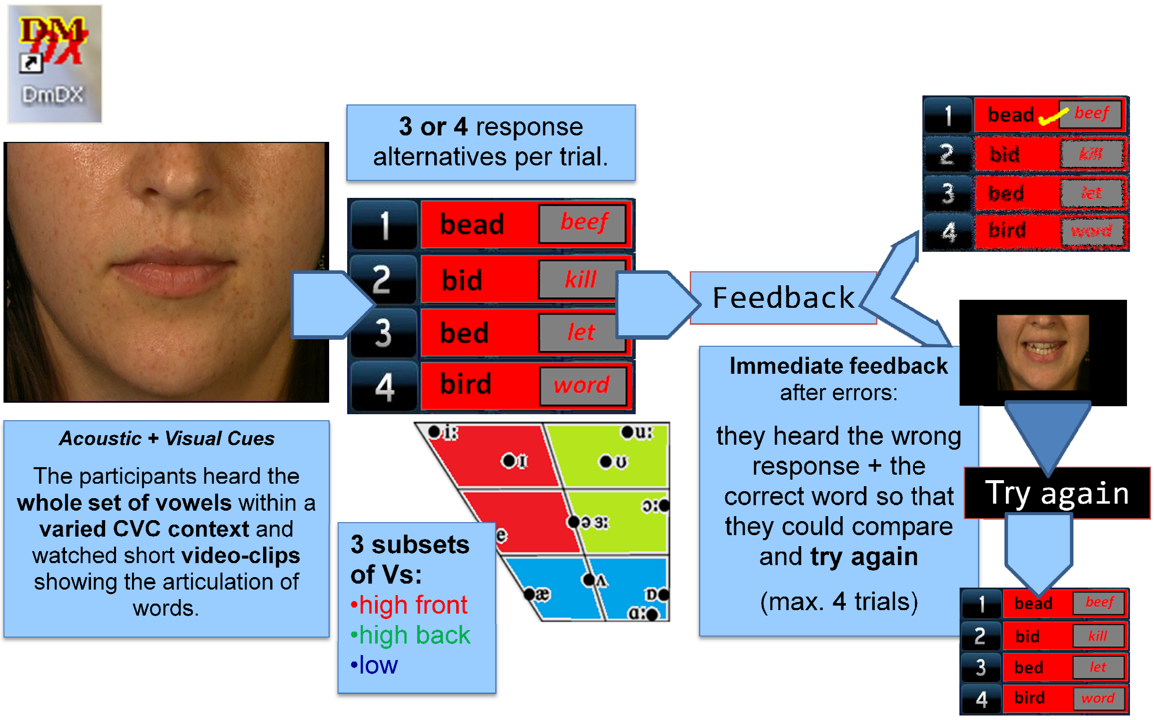


Figure 3.9. Screenshots of the ID training interface.

3.3.6.2. Articulatory (ART) training

Articulatory (ART) training was based on an imitation task similar to the production task described in section 3.3.3.2.which consisted of different phases (see Figures 3.10-3.11):

(1) AV presentation of stimuli and delayed repetition

Participants were presented audiovisual stimuli so that they could focus both on acoustic properties of sound and articulatory gestures for more accurate vowel articulation.

(2) Feedback between sections based on self-monitoring

Each of the four blocks was followed by a feedback section. Feedback ('*Listen and compare*') was based on self-monitoring and self-correction with the aim of activating the learners' feedback sensory systems so that rapid motor adjustments could be made

(Guenther et al., 2006) after detecting salient differences between the NS production and their own output.

During that section, they heard their own responses ('*This is you…*'), compared them with the model ('*This is the native speaker…*') and then corrected themselves as many times as necessary ('*Try again.*'). The number of trials was limited to four times. This comparison between predicted output (AV input) and actual production (output generated by the speaker) would allow them to rapidly detect articulation errors during the next attempts to accurately imitate the vowel (Guenther et al., 2006; Golfinopoulus et al., 2010).

When they were satisfied with their own productions, they pressed *"Next word"* and continued listening to the other words.

(3) Repetition for Self-correction

When they were not satisfied after judging their own productions, they pressed *"Try again"*. Then they were presented with the same stimulus and they proceeded to imitate it, and then judge their own productions.

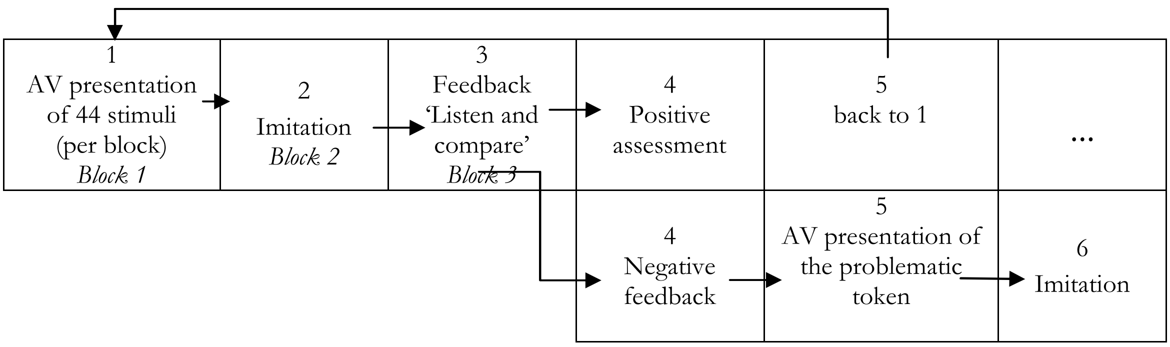


Figure 3.10. Training process in ART training, including feedback based on self-monitoring and self-repairs.

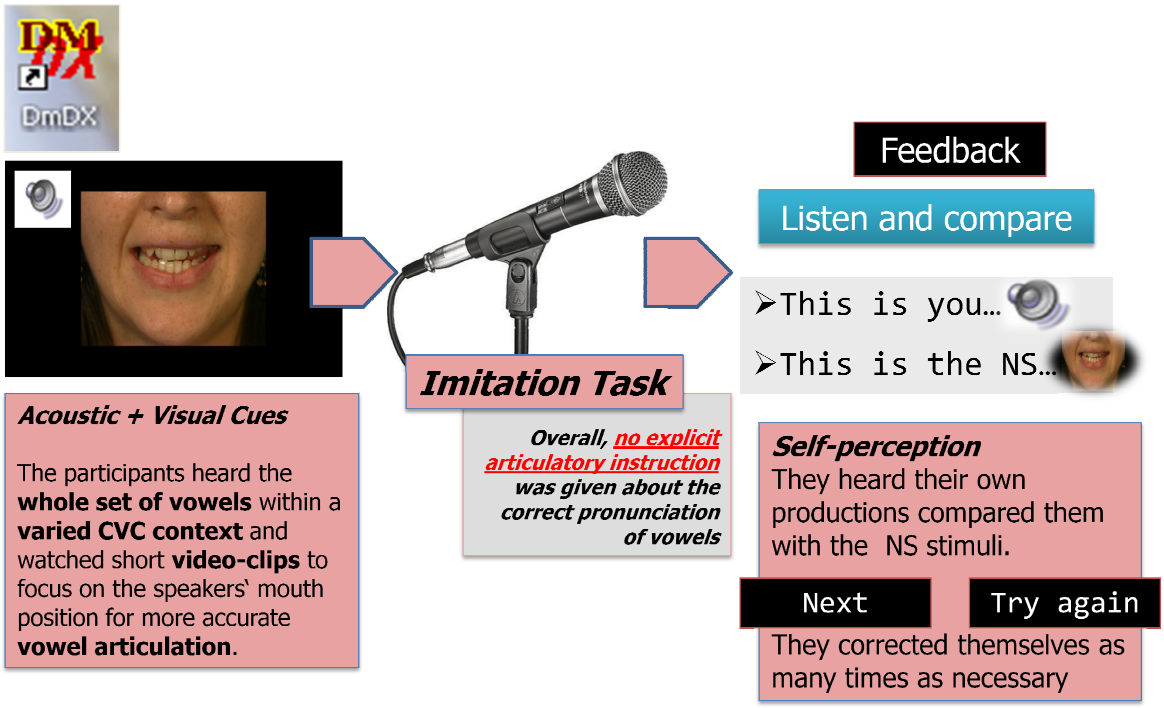


Figure 3.11. Screenshots of the ART training interface.

3.4. Main statistical analyses

3.4.1. English vowel perception: effect of training and type of training on vowel perception (RQ1.1 and RQ2.1)

RQ1assessed whether *training* was effective in improving vowel perception resulting from higher accuracy and gains in vowel identification (RQ1.1), identification of synthesized vowels (RQ1.2) and vowel discrimination (RQ1.3). RQ2.1 assessed which *type of training* was more effective and would lead to larger gains in vowel identification and discrimination.

First, in order to compare experimental and control groups on vowel perception at pre-test, a series of one-way ANOVAs with *subject group* (ID, ART, control) as the between-subjects factor, were run on percent correct identification of natural vowels (*identification* *accuracy-IDI),* mean percent correct identification of synthesized vowels *(identification accuracy-ID-2)* and mean percent correct discrimination *(discrimination accuracy-DIS)* of the three subject groups at T1. These analyses would confirm the comparability of subjects groups prior to training.

Next, paired-samples t-tests were also run to statistically compare the pre-test and post-test perception measures of the control group, which were hypothesized not to differ significantly due to the absence of treatment. These would validate the improvement and gains obtained for the ID and ART groups on vowel perception.

To explore the overall effects of training and type of training on vowel perception, and assess the differential effects of the type of training (ID vs. ART) (RQ2.1), a series of mixed between-within ANOVAS with effect of *training* (T1 vs. T2) as within-subjects factor and *type of training* (ID vs. ART) as between-subjects factor, were conducted on a series of perception measures derives from each of the tests. These would reveal, first, if the trainees' post-test English vowel identification and discrimination had improved with respect to pre-test and, second, and would evaluate the differential perception gains of the distinct training methods. Further within-subject factors included in the ANOVAS were *stimulus presentation* (A vs. AV) or factors to test for generalization effects and robustness of learning, such as *talker condition* (trained vs. untrained) and *word type* (trained vs. new), *manipulation of duration* (S1, S2, S3, S4, S5, S6 and S7) and *test conditions* (fixed context vs. context variability).

To explore the overall effects of HPVT (RQ1.1) and the effects of type of training (RQ2.1 on the identification of natural vowels, the following dependent variables were submitted to the ANOVAS mentioned above:

- *Mean correct identification* considering the full set of SBE vowels (/i ɪ e ɜː æ ʌ ɑː ɒ ɔː ʊ uː/), for each *vowel sets* (vowel set 1: /i ɪ e ɜː/, vowel set 2: /æ ʌ ɑː ɒ/, vowel set 3: /ɔː ʊ uː/), for each *vowel* (/iː/, /ɪ/, /e/, /ɜː/, /æ/, /ʌ/, /ɑː/, /ɒ/, /ɔː/, /ʊ/ and/uː/), and corresponding *gains*.

- Mean *error dispersion* scores and *error dispersion difference* calculated for each *vowel*: (/iː/, /ɪ/, /e/, /ɜː/, /æ/, /ʌ/, /ɑː/, /ɒ/, /ɔː/, /ʊ/ and/uː/).

To assess the effects of training (RQ1.2) and the effects of type of training (RQ2.1) on the perception of synthesized vowels, and assess the learners' degree of reliance on the duration cue, the following measures were submitted to a series of mixed between-within ANOVAS:

- *Mean correct identification* of all synthesized vowels, (/i ɪ æ ʌ ɑ ʊ u/), of each *vowel,* (/i ɪ æ ʌ ɑ ʊ u/), of tense (/i ɑ u/) vs. lax vowels (/ɪ æ ʌ ʊ /), and corresponding *gains.*
- *Duration effect score* (DES) obtained for each *vowel* and changes in *DES*.

In order to assess the effects of training (RQ1.3) and type of training (RQ2.1) on vowel discrimination, the ANOVAS were run on these measures:

- Mean correct discrimination of 16 vowel contrasts (/C1 /iː/-/ɪ/, C2 /ɪ/-/e/, C3 /e/-/ɜː/, C4 /ɪ/-/ɜː/, C5 /e/-/æ/, C6 /ɜː/-/æ/, C7 /æ/-/ʌ/, C8 /ʌ/-/ɑː/, C9 /æ/-/ɑː/, C10 /ɜː/-/ʌ/, C11 /ɜː/-/ɑː/, C12 /ɒ/-/ɑː/, C13 /ʌ/-/ɒ/, C14 /ɒ/-/ɔː/, C15 /ʊ/-/uː/, C16 /ɔː/-/uː/), each *vowel set*: (high-front C1-C4), low (C5-C13) and back vowel contrasts (C4-C16).

Further analyses included a series of paired-samples run on ID and ART training groups to further assess the amount of gains in vowel identification and discrimination, for each vowel or vowel contrast in each test condition. To verify between-group differences (ID vs. ART) in post-test accuracy scores and perception gains, and compare the performance of each training group at post-test, a series of independent-samples t-tests were conducted on each perception measure.

3.4.2. English vowel perception: effect of training and type of training on vowel perception (RQ1.4 and RQ2.2)

RQ1.4 assessed whether HVPT was effective in improving vowel production, resulting in a more target-like change of weighting of spectral and duration dimensions from pre-test to post-test, i.e. leading to increased spectral distances, decreased duration distances, and significant changes in the height and frontness dimensions.

RQ2.2 assessed which type of training was more effective in improving vowel production and was better able to increase attention to spectrum and decrease attention to duration.

To test the comparability of experimental and control groups on Catalan-Spanish learners' vowel production at pre-test, first one-way ANOVAs *subject group* (ID, ART, control) as the between-subjects factor, compared mean *spectral distance*, *duration ratio* of tense and lax vowels, normalized *height* (B1-B0), *frontness* (B2-B1) and *duration* values of the subject groups at T1. These analyses would confirm the comparability of subjects groups prior to training.

Paired-samples t-tests were also run to statistically compare the pre-test and post-test production measures of the control group, which were hypothesized not to differ significantly due to the absence of a treatment. These would validate the improvement and gains obtained for the ID and ART groups on vowel production.

To explore the overall effects of HVPT on vowel production (RQ1), and the differential effects of the type of training (ID vs. ART) (RQ2), mixed between-within ANOVAS with effect of *training* (T1 vs. T2) as within-subjects factor and *type of training* (ID vs. ART) as between-subjects factor were conducted on each of the production measures to reveal, first, if the trainees' post-test English vowels differed acoustically from their pre-test vowels and, secondly, which were the differential gains of the distinct training methods, and which vowels differed acoustically between subject groups at post-test.

To explore the overall effects of HPVT (RQ1.4) and the effects of type of training (RQ2.4) on the Catalan-Spanish learners' vowel production, the following dependent variables were submitted to statistical analyses:

- The *spectral distance* scores (SD) derived from 10 vowel distances (/iː/-/ɪ/,/ɪ/-/e/, /e/-/ɜː/, /æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɒ/-/ɑː/, /ʌ/-/ɒ/, /ɒ/-/ɔː/, /ʊ/-/uː/), as a measure of the amount of overlapping in the L2 vowel space of trainees, who would be significantly raising or lowering the their tongue position after training.
- *Changes in Spectral duration* (SD) derived from pre-test and post-test SDs calculated for the ten vowel contrasts, as a measure of the amount of spectral changes (increase or decrease) from pre-test to post-test.
- *Duration* (measured in msec) values of 11 English vowel monophthongs (/iː/-/ɪ/, /e/, /ɜː/, /æ/, /ʌ/, /ɑː/, /ɒ/, /ɔː/, /ʊ/, /uː/).
- Vowel frequency measurements (F0, F1, F2) converted to Bark (B) and then normalized by computing the difference between Bark converted F1 and F0 (B1-B0) and between Bark-converted F2 and F1 (B1-B0) for vowel **height** and **frontness** obtained from /iː/-/ɪ/, /e/, /ɜː/, /æ/, /ʌ/, /ɑː/, /ɒ/, /ɔː/, /ʊ/ and /uː/.

Furthermore, a combination of paired-samples were conducted separately for ID and ART training groups to further assess the amount of changes in vowel production, i.e. pre-test/post-test difference for each vowel, vowel contrast in each test condition after training. To verify between-group differences (ID vs. ART) in post-test accuracy scores and production gains, and compare the performance of each training group at post-test, a series of independent-samples t-tests were conducted on each production measure.

Chapter 4

Results I: Perception

In this chapter, the analysis and results of the perception tasks is provided. The results are presented in three different sections corresponding to the three tasks the participants completed: identification of vowels (ID1-task1), identification of synthesized vowels (with manipulated duration) (ID2-task2) and vowel discrimination (DIS-task3). Within each section we provide results on the effects of *training* (RQ1) and the effects of the *type of training* (RQ2):

Research Question 1 (RQ1): *Main effect of training* (T1 vs. T2)

Will AV HVPT lead to changes in the Catalan-Spanish speakers' production of English vowels? That is, is AV HVPT an effective method to improve L2 vowel production?

Research Question 2 (RQ2): *Effect of type of training* (ID vs. ART)

Will there be differential gains in L2 vowel production as a function of *type of training* (Identification vs. Articulatory training)?

Within each of the sections, the first part is dedicated to the effects of training on perceptual performance (RQ1). More specifically, it examines the overall effectiveness of training on L2 vowel perception by describing the performance of three subjects groups (ID training, ART training, control). Secondly, the results of the analysis comparing the effects of identification and articulatory training methods, ID and ART, on vowel perception are reported (RQ2).

Results are presented in the following order: (1) comparability of subject groups at pre-test in vowel perception accuracy, (2) effects of *training* and *type of train*ing (ID vs. ART) on perception accuracy, and (3) effect of training and stimulus variables on perception accuracy. For all statistical tests the alpha level was set at .05.

4.1. Identification of vowels (ID1 – task1)

The **Sub-RQ1.1** examines to what extent learners will improve in the identification of natural L2 vowels from pre-test to post-test. This section presents the effects of training on mean *correct vowel identification* (RQ1.1.1) and *mean error dispersion* (RQ1.1.2), as well as the differential effects of identification (ID) and articulatory (ART) types training on each of these identification accuracy measures (RQ2.1).

4.1.1. Comparability of subject groups at pre-test

The ID (M= 62.32, SD= 19.08), ART (M= 67.28, SD= 19.39), and control groups (M= 63.16, SD= 17.20) obtained similar pre-test perceptual identification accuracy scores. A one-way ANOVA confirmed that the three groups did not significantly differ from one another at pre-test in vowel identification accuracy ($F(2, 81)=0.62$, $p= .543$), suggesting that the groups were indeed comparable in terms of their identification skills before training (see Figure 4.1).

The NS control group performed at ceiling in vowel identification (see Table 4.1). A one-way ANOVA showed that the effect of subject group (ID, ART, NNS control, NS control) on pre-test vowel identification scores was significant, ($F(3, 89)=7.81$, $p< .001$). A

Tukey posthoc test confirmed that the differences between the NS control group and each of the NNS groups (ID, ART, control) were statistically significant ($p < .001$).

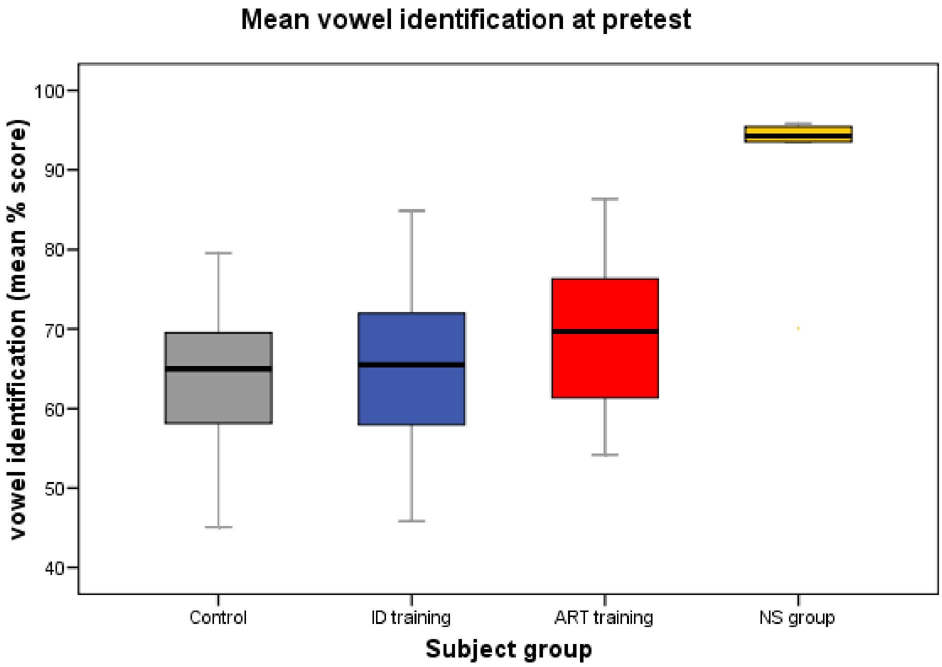


Figure 4.1. Average percent correct identification of English vowels for ID training, ART training, NNS control and NS control groups at pre-test.

Vowel Identification at pre-test	Subject groups			
	ID training	ART training	NNS control	NS
Mean vowel identification	61.55 (18.37)	67.28 (19.09)	63.16 (16.75)	98.11 (8.22)
/iː/	60.55 (21.22)	62.89 (21.64)	56.04 (17.91)	91.92 (3.14)
/ɪ/	58.72 (22.36)	66.93 (25.87)	62.92 (24.35)	99.48 (1.47)
/e/	80.34 (24.14)	84.64 (23.70)	81.66 (22.19)	99.48 (1.47)
/ɜː/	73.05 (24.98)	79.17 (24.13)	76.04 (24.96)	99.48 (1.47)
/æ/	50.11 (25.61)	71.22 (27.46)	59.17 (23.16)	100.0 (0.0)
/ʌ/	65.56 (20.23)	62.50 (22.65)	48.54 (23.78)	98.96 (2.95)
/ɑː/	67.97 (25.14)	72.13 (23.44)	57.08 (21.17)	99.48 (1.47)
/ɒ/	57.03 (21.93)	59.76 (20.37)	52.50 (18.11)	85.42 (4.45)
/ɔː/	59.37 (19.66)	60.81 (19.45)	59.79 (20.33)	92.71 (9.11)
/ʊ/	45.96 (15.24)	48.83 (15.85)	53.54 (13.76)	82.50 (4.98)
/uː/	44.66 (17.56)	48.83 (19.46)	47.92 (13.35)	93.44 (7.36)

Table 4.1. Percent correct identification scores for the individual vowels standard deviations by ID training, ART training, NNS control and NS baseline groups.

4.1.2. Effects of training and type of training on mean percent correct vowel identification

The **Sub-**RQ1.1.1 examined whether audiovisual HVPT resulted in a significant improvement in the identification of English vowels from pre-test to post-test in the trainees. If the training was effective, the training groups should reveal a significant increase in the vowel identification scores at post-test, while the control group should not exhibit such improvement. The sub-question RQ2.1 aimed to compare the effectiveness of the two training methods in improving accuracy in vowel identification.

The descriptive statistics for pre-test and post-test identification accuracy and improvement scores (gains) for each group and vowel are shown in Table 4.2. As shown in Figure 4.2, the training groups increased in mean accuracy (improvement of 14.87%) and the standard deviations decreased, whereas the control group remained almost unchanged from pre-test to post-test.

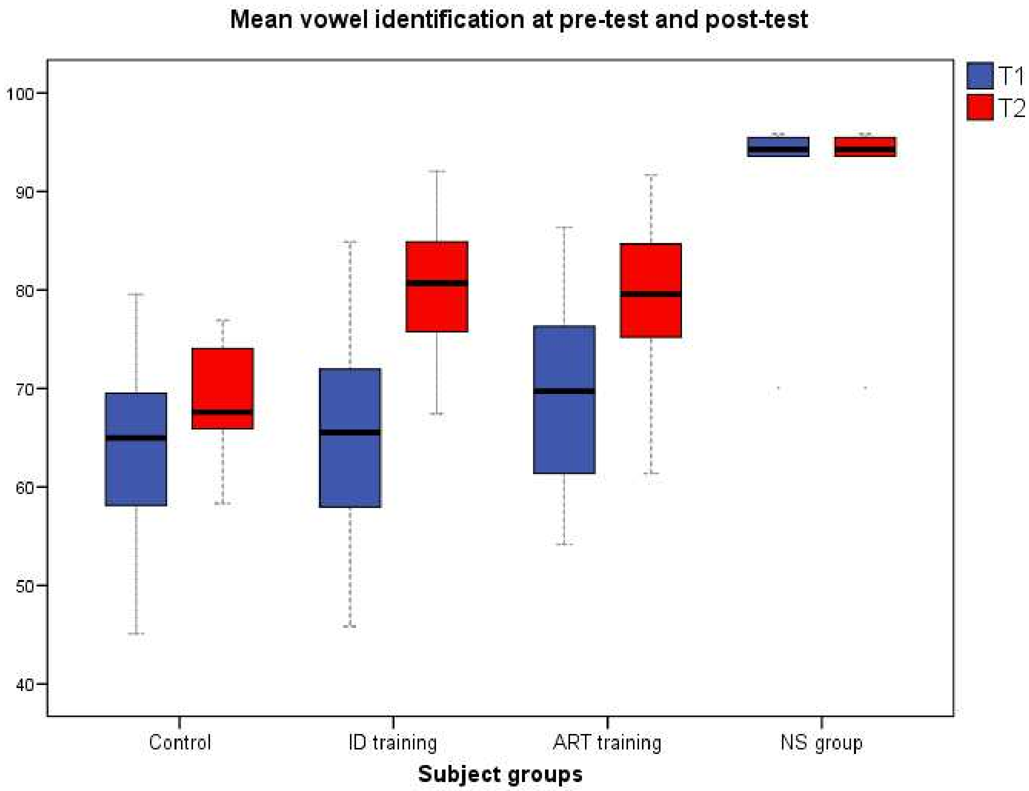


Figure 4.2. Average percent correct identification of English vowels for ID training, ART training, NNS control and NS control groups at pre-test and post-test.

Vowel identification	ID training (*N*=32)			ART training (*N*=32)			Control group (*N*=20)			NS (*N*=10)
	T1	**T2**	**gains**	**T1**	**T2**	**gains**	**T1**	**T2**	**gains**	
Mean	61.55	79.18	*17.63*	67.28	80.16	*12.88*	63.16	68.73	*5.57*	98.11
/iː/	60.55 (21.22)	76.17 (10.59)	15.36 (19.87)	62.89 (21.64)	75.00 (11.64)	12.37 (21.70)	56.04 (17.91)	61.88 (22.28)	5.62 (20.73)	91.92 (3.14)
/ɪ/	58.72 (22.36)	80.08 (11.82)	21.35 (23.47)	66.93 (25.87)	80.08 (15.94)	13.02 (25.53)	62.92 (24.35)	62.50 (20.28)	0.20 (12.28)	99.48 (1.47)
/e/	80.34 (24.14)	97.79 (4.36)	17.32 (23.78)	84.64 (23.70)	94.66 (7.94)	10.16 (25.00)	81.66 (22.19)	90.0 (9.21)	8.3 (20.50)	99.48 (1.47)
/ɜː/	73.05 (24.98)	93.88 (8.79)	20.44 (22.98)	79.17 (24.13)	94.14 (9.39)	15.36 (25.08)	76.04 (24.96)	86.04 (13.94)	10.00 (19.28)	99.48 (1.47)
/æ/	50.11 (25.61)	76.17 (16.77)	26.06 (24.85)	71.22 (27.46)	78.52 (22.13)	7.30 (32.55)	59.17 (23.16)	64.58 (17.20)	6.67 (0.43)	100.0 (0.0)
/ʌ/	65.56 (20.23)	78.64 (13.75)	13.08 (17.79)	62.50 (22.65)	80.60 (16.24)	18.10 (24.65)	48.54 (23.78)	57.29 (16.49)	-0.83 (13.49)	98.96 (2.95)
/ɑː/	67.97 (25.14)	83.72 (14.29)	15.75 (24.77)	72.13 (23.44)	83.07 (14.89)	10.94 (22.42)	57.08 (21.17)	62.92 (13.95)	5.42 (20.81)	99.48 (1.47)
/ɒ/	57.03 (21.93)	76.04 (11.93)	19.01 (22.10)	59.76 (20.37)	72.53 (9.52)	12.76 (18.42)	52.50 (18.11)	60.21 (17.39)	7.70 (20.24)	85.42 (4.45)
/ɔː/	59.37 (19.66)	71.48 (11.05)	12.11 (20.26)	60.81 (19.45)	67.84 (10.75)	7.03 (20.51)	59.79 (20.33)	67.08 (8.54)	7.29 (20.45)	92.71 (9.11)
/ʊ/	45.96 (15.24)	57.16 (10.27)	11.20 (16.85)	48.83 (15.85)	55.34 (12.26)	6.51 (16.22)	53.54 (13.76)	54.58 (14.04)	0.62 (14.76)	82.50 (4.98)
/uː/	44.66 (17.56)	67.06 (12.31)	22.40 (20.13)	48.83 (19.46)	66.15 (12.47)	17.32 (19.20)	47.92 (13.35)	56.46 (11.88)	10.21 (21.14)	93.44 (7.36)

Table 4.2. Percent correct identification scores for the individual vowels and standard deviations in the pre- and post-test by ID training, ART training, NNS control and NS baseline groups.

A paired-samples t-test was run on the pre- and post-test identification scores of the control group to determine whether there was a statistically significant difference in vowel identification accuracy between the two testing times. Results indicated that post-test scores were not significantly higher (gains: M= 5.57, SD= 15.53) than pre-test scores ($t(19)$= -1.60, p= .125).

A two-way ANOVA with mean percent correct vowel identification as the dependent variable, effect of *training* (T1 vs. T2) as the within-subjects factor, and *type of training* (ID vs. ART) as the between-subjects factor, was conducted to determine if the training had been effective in increasing mean identification accuracy, and if differences between testing times in vowel identification were due to the type of training the trainees had received. There was a statistically significant effect of *training* on mean vowel identification ($F(1, 62)$= 52.10, p< .001, η2= .457). However, the effect of *type of training* did not reach statistical significance ($F(1, 62)$= 0.26, p= .615, η^2= .004). The interaction between *training* and *type of training* did not reach significance ($F(1, 62)$= 1.54, p= .220, η^2= .024). These results confirmed that vowel identification scores were significantly higher at post-test than they were at pre-test, suggesting that the training was generally effective in increasing vowel identification accuracy. All trainees showed significant improvement in vowel identification (ID training: 17.63%, ART training: 12.88%). Up to this point, the type of training made no significant difference in the results. ID and ART types of training were both similarly effective in improving vowel identification accuracy.

In the following sections we examine how other variables (vowel set, vowel, test condition, word type, talker condition) might have affected vowel identification accuracy and how this changed as result of training and type of training from pre-test to post-test. Subsequent ANOVAs included within-subject factors such as *vowel set* (**high-front, mid, low vowels),** ***vowel*** (/iːɪeɜːæʌɑːɒɔːʊuː/) and stimuli properties that might affect accuracy in vowel perception, such as *stimulus presentation* (auditory vs. audiovisual), *word type* (trained vs. new), and *talker condition* (trained vs. untrained).

4.1.3. Effects of training and type of training on identification of vowel sets

The interest in the analysis of percent correct identification by vowel sets lies in the fact that our HVPT was based on the use of vowel sets chosen based on difficulty (Nishi & Kewley-Port, 2007). One of the research questions was whether identification gains would vary as a function of *vowel set* (RQ1.1.1a). RQ2.1a explored whether ID and ART types of training were differently effective in improving the identification accuracy of the three *vowel sets.*

Results for accuracy in the identification of vowel sets (high-front, low, back) at pre-test and post-test, and identification gains, are presented on Table 4.3. Figure 4.3 shows the perception accuracy for each vowel set and each training group before and after training. Generally, the post-test identification scores were higher than the pre-test scores for the three vowel sets, and accuracy in vowel perception varied as a function of vowel set. Trainees improved on the identification of high-front (gains 13-18%), low (gains 12-19%), and back (gains 10-15%) vowels regardless of the type of training. Both ID and ART training groups showed a similar pattern of identification results: the high-front vowels (/iːɪ e ɜː/) got the highest accuracy in identification, followed by low vowels (/æʌɑːɒ/), whereas back vowels (/ɔːʊuː/) were perceived at the lowest accuracy, at pre-test (see Table 4. 3). At post-test, high-front vowels (M_{ID}= 86.98, M_{ART}= 85.97) were still perceived at higher accuracy rates than low (M_{ID}= 66.42 M_{ART}= 78.68) and back (M_{ID}= 65.23, M_{ART}= 63.11) vowels.

	Vowel identification					
	ID training (*N*=32)			ART training (*N*=32)		
Vowel set	**T1**	**T2**	**gains**	**T1**	**T2**	**gains**
High-front /iːɪeɜː/	68.16 (21.08)	86.98 (5.76)	18.81 (19.84)	73.40 (21.31)	85.97 (9.03)	12.56 (22.14)
Low /æʌɑːɒ/	59.69 (19.97)	78.64 (8.37)	18.98 (18.55)	66.41 (20.62)	78.68 (9.69)	12.27 (20.75)
Back /ɔːʊuː/	50.00 (15.69)	65.23 (6.86)	15.23 (16.82)	52.82 (15.40)	63.11 (8.29)	10.29 (15.72)

Table 4.3. Percent correct identification scores and standard deviations for the three vowel sets (high-front, mid, low) by ID training and ART training groups.

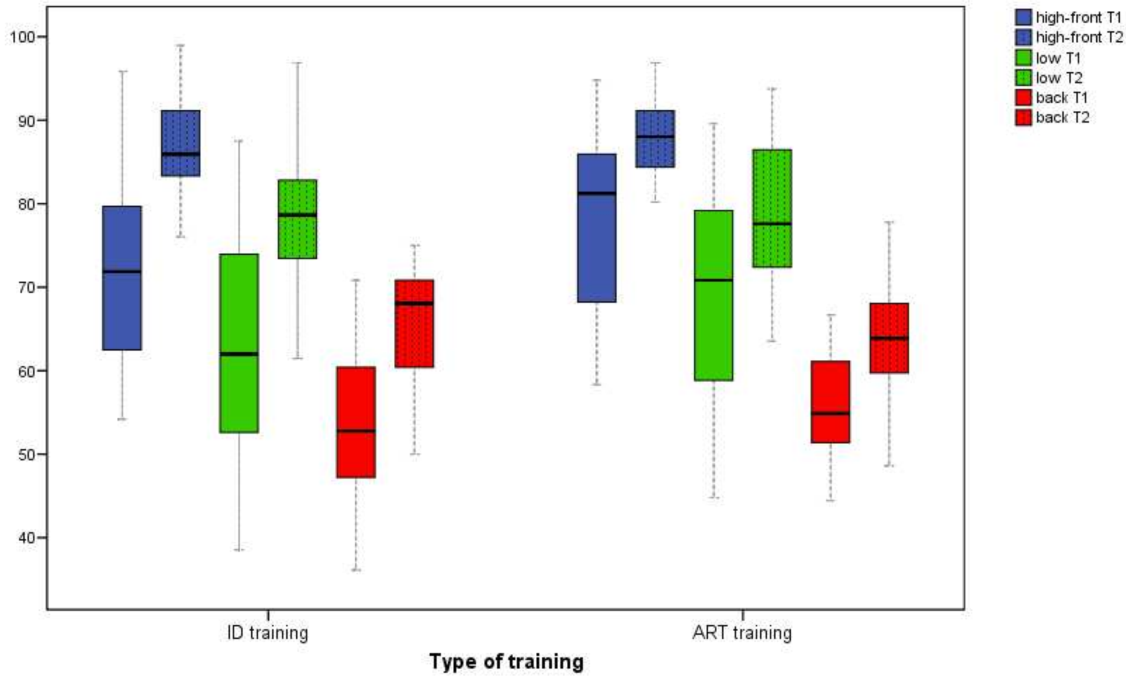


Figure 4.3. Percent correct identification scores for the three vowel sets (high-front, mid, low) by ID training and ART training groups.

A three-way repeated-measures ANOVA with effect of *training* (T1 vs. T2) and *vowel set* (3) as within-subjects factors, and *type of training (ID vs. ART)* as between-subjects factor, was conducted to determine whether training was effective in improving the identification of three vowel sets (RQ1.1.1a) and if the two types of training had been equally effective in improving the identification accuracy of the three *vowel sets* (high-front, low and central vowels) (RQ2.1a).

The ANOVA yielded a significant main effect of *training* for perception accuracy ($F(1, 62)=41.43$, $p<.001$, $\eta^2= .401$) and *vowel set* ($F(2, 124)=275.64$, $p<.001$, $\eta^2= .816$), and a significant interaction between *training* and *vowel set* ($F(2, 124)= 3.80$, $p< .05$, $\eta^2= .058$), indicating that the effects of training on identification accuracy varied as a function of vowel set. In fact, high-front vowels were perceived at higher accuracy rates than low and back vowels, both at pre-test and post-test, regardless of the type of training. The effect of *type of training* on identification accuracy did not reach significance, ($F(1, 62)= 0.59$, $p= .446$, $\eta^2= .009$). The interactions between *vowel set* and *type of training* ($F(2, 124)= 74.51$, $p< .05$, $\eta^2= .023$) and *training x vowel set x type of training* ($F(2, 124)= 6.66$, $p< .05$, $\eta^2= .005$) were not statistically significant. These results indicate that the type of training received did not determine the degree of accuracy for any of the vowel sets at pre-test or the amount of perception gains in high-front, low and back vowels.

To further compare ID and ART training groups, separate paired-samples t-tests were conducted to compare pre-test and post-test identification scores of high-front, low and back vowels. Tests revealed that the three vowel sets underwent significant pre-test-post-test changes after ID and ART training, so improvement in identification accuracy was significant at the $p< .05$ for all comparisons as a result of training (see Table 4.4). Vowel sets determine the degree of accuracy in perception, so at post-test trainees still performed better with high-front and low vowels than with back vowels.

Vowel Identification	Paired comparisons			
	ID training (*N*=32)		**ART training (*N*=32)**	
	df t p-value	gains	df t p-value	gains
Vowel set				
High-front /iːɪ e ɜː/	t(31)=-5.36, p< .001*	*18.81 (19.84)*	t(31)=-3.21, p< . 005*	*12.56 (22.14)*
Low /æʌɑːɒ/	t(31)=-5.79, p< .001*	*18.98 (18.55)*	t(31)=-3.34, p< . 005*	*12.27 (20.75)*
Back /ɔːʊ uː/	t(31)=-5.12, p< .001*	*15.23 (16.82)*	t(31)=-3.70, p< . 005*	*10.29 (15.72)*

Table 4.4. Results of paired-samples t-tests on the pre-test and post-test identification scores of three vowel sets (high-front, mid, low vowels) and accuracy gains obtained by ID and ART training groups.

In order to assess whether the size of gains was significantly different as a function of vowel set, an ANOVA with *type of training* (ID vs. ART) as a between-subjects factor and *vowel set* (high-front, mid, low) as a within-subject factors was run on ID and ART groups' gains. The test yielded a significant effect of *vowel set* on gains ($F(2, 124)=12.04$, $p<.001$, $\eta^2= .163$), and the interaction between *vowel set* and *type of training* did not reach significance ($F(2, 124)=00.20$, $p= .817$, $\eta^2=.003$). The effect of *type of training* on identification gains was not statistically significant ($F(1, 62)=1.73$, $p= .193$, $\eta^2= .027$). These results confirmed that vowel set could be used to explain the amount of perceptual gains. After ID and ART training, the largest size of gains were obtained for high-front and low vowels (18-19% gains after ID training; 12-13% gains after ART training), but the type of training was not the factor explaining differences in amount of gains.

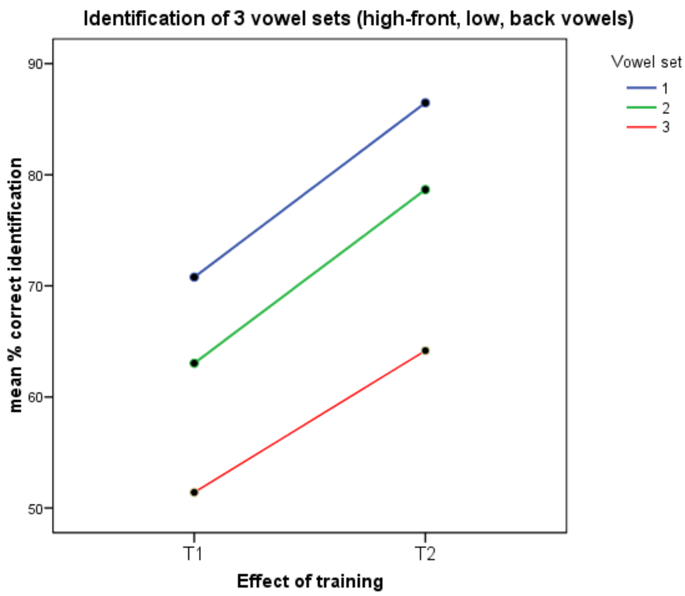


Figure 4.4. Mean percent correct identification at pre-test and pos-test and gains for each vowel set (1= high-front, 2=low, 3=back).

These overall results show, one the one hand, that there was a significant improvement in vowel identification of 10-18% for the three vowel sets (high-front, low and back vowels). However, these results indicate, on the other hand, that the size of the effect of training depended on vowel set (high-front, mid, low) rather than the type of training (ID vs. ART), which did not have a significant effect on post-test identification accuracy of such vowel sets. In other words, the type of training did not determine differences in perception at post-test or the fact that certain vowels were easier to identify at post-test. The explanation could be that *vowel set* seems to play a role in the learners' ability to identify English vowels correctly and in the overall training effect. As reflected in the pattern of means described above (see Figure 4. 4), high-front vowels were identified more accurately than low and back vowels before and after training. After each type of training, the pattern of identification scores remained similar to pre-test despite overall improvement: the relative difficulty that trainees had in identifying back vowels (M_{T2}=61-65%) at post-test indicated that they were somehow more resistant to training than high-front (M_{T2}=85-86%) and low vowels (M_{T2}=78%) despite improvement. Conversely, training seemed to have a more visible impact on the high-front and low vowel sets. The size of gains was larger for high-front and low vowels than back vowels after ID and ART training.

4.1.4. Effects of training and type of training on identification of different vowels

Results for accuracy in the identification of vowels at pre-test and post-test, previously presented on Table 4.2, show that trainees improved on the identification of the eleven vowels (RQ1.1.1b). Figures 4.5 and 4.6 show the perception accuracy for each vowel produced after ID and ART training, respectively (RQ2.2b).

Some English vowels were more accurately perceived than others after training. An inspection of the post-test performance obtained for each vowel revealed that the highest percentage of gains was obtained for /uː/ (mean improvement of 14.87 % points; ID: 22%, ART: 17%) despite the participants achieving the lowest post-test correct identification scores in that vowel, together with /ʊ/(see Figures 4.5-4.6). The best post-test identification performance was obtained for the high-front vowels /e/ ($T2_{ID}$= 97.79, $T2_{ART}$= 94.66) and /ɜː/ ($T2_{ID}$= 93.88, $T2_{ART}$= 94.14), for which trainees reached ceiling performance. A notable improvement was seen for /ɑː/ ($T2_{ID}$= 83.72, $T2_{ART}$= 83.07) and /ɪ/ ($T2_{ID}$= 80.08, $T2_{ART}$= 80.08), which were less frequently confused after training.

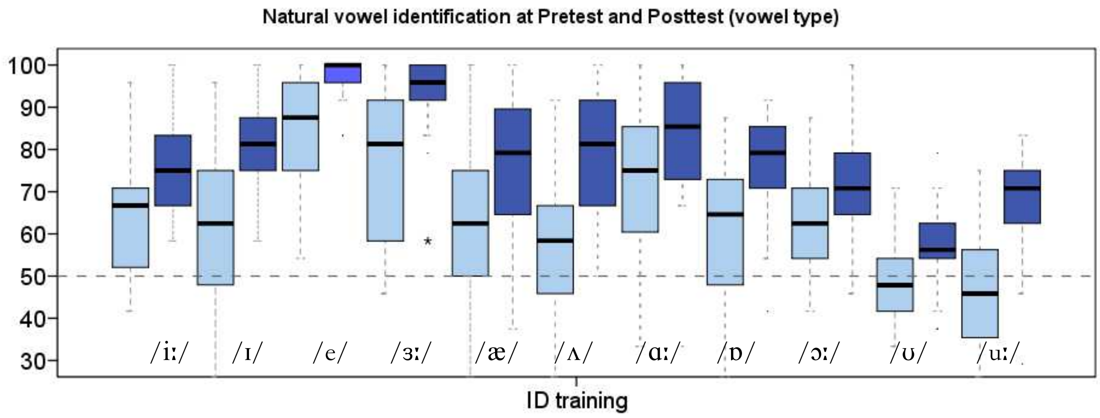


Figure 4.5. Boxplots of identification accuracy across all subjects in the ID training group, presented separately for each vowel at pre- (light blue) and post-test (dark blue).

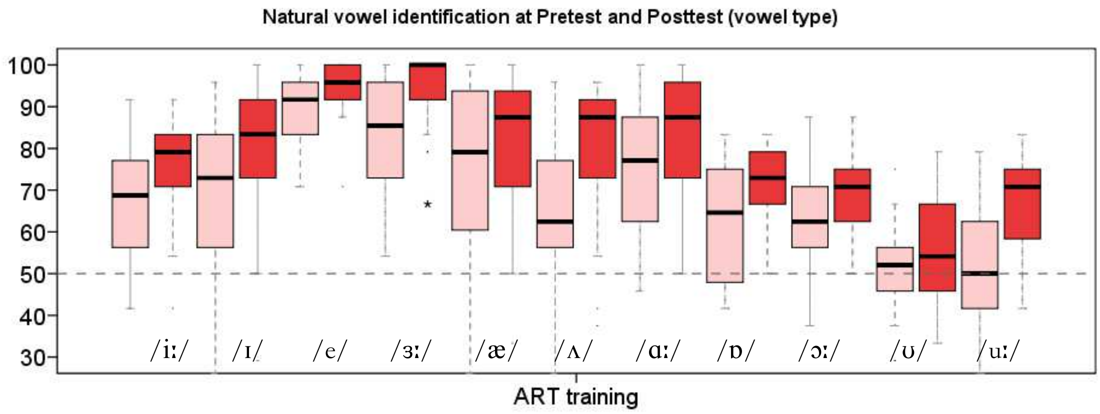


Figure 4.6. Boxplots of identification accuracy across all subjects in the ART training group, presented separately for each vowel at pre- (light red) and post-test (dark red).

Pre-test and post-test differences in identification scores demonstrated that Catalan-Spanish speakers improved relatively modestly (improvement of 9%) for two vowels (/ɔː/ and /ʊ/), while they improved twice as much for three (/uː/, /ɜː/, /ɪ/)

(improvement of 18%) and considerably more for the rest (improvement of 12%). The post-test data also revealed that the high back vowel contrast (/ʊ/-/uː/) was the most frequently confused at post-test.

A three-way repeated-measures ANOVA with *effect of training (T1 vs. T2)* and *vowel* (11) as within-subjects factors, and *type of training (ID vs. ART)* as between-subjects factor, was conducted to determine if the training had been equally effective in improving the identification accuracy of the 11 vowels (/iːɪeɜːæʌɑːɒɔːʊuː/) (RQ1.1.1b) and whether differences could be due to type of training (RQ2.2b).

The results revealed that there was a significant effect of *training* ($F(1, 62)=41.46$, $p<.001$, $\eta^2= .401$), a significant effect of *vowel* ($F(10, 620)=82.57$, $p<.001$, $\eta^2= .571$), and a significant interaction between *training* and *vowel* ($F(10, 620)=5.38$, $p<.001$, $\eta^2= .080$). The effect of *type of training* ($F(1, 62)=0.30$, $p=.584$, $\eta^2= .005$) and the *vowel x type of training* interaction ($F(10, 620)=201.65$, $p=. 381$, $\eta^2= .017$) did not reach significance. The *effect of training x vowel type x type of training* interaction ($F(10, 620)=31.58$, $p=.968$, $\eta^2= .006$) was not statistically significant. The significant two-way interaction suggested that the effect of training on accuracy differed as a function of the target vowel.

Subsequent paired-samples t-tests were conducted on pre-test and post-test identification scores of each vowel for ID and ART training groups. These revealed that most pre-test identification scores differed significantly from those obtained at post-test at the $p< .05$ level (see Table 4.5).

Vowel Identification	Paired comparisons					
	ID training (*N*=32)			**ART training** (*N*=32)		
	df t p-value	**gains**		**df t p-value**	**gains**	
iː	t(32)=4.34, *p*< .001*	15.36 (19.87)		t(30)=3.24, *p*< . 005*	12.37 (21.70)	
ɪ	t(32)=5.46, *p*< .001*	21.35 (23.47)		t(30)=2.72, *p*< . 05*	13.02 (25.53)	
e	t(32)=4.13, *p*< .001*	17.32 (23.78)		t(30)=2.27, *p*< . 05*	10.16 (25.00)	
ɜː	t(32)=5.12, *p*< .001*	20.44 (22.98)		t(30)=3.38, *p*< . 005*	15.36 (25.08)	
æ	t(32)=3.68, *p*< .005*	26.06 (24.85)		t(30)=1.55, *p*= . 131	7.30 (32.55)	
ʌ	t(32)=7.52, *p*< .001*	13.08 (17.79)		t(30)=3.99, *p*< . 001*	18.10 (24.65)	
ɑː	t(32)=3.56 *p*< .005*	15.75 (24.77)		t(30)=2.79, *p*< . 05*	10.94 (22.42)	
ɒ	t(32)=4.83, *p*< .001*	19.01 (22.10)		t(30)=3.75, *p*< . 005*	12.76 (18.42)	
ɔː	t(32)=3.13, *p*< .005*	12.11 (20.26)		t(30)=2.01, *p*= . 053*	7.03 (20.51)	
ʊ	t(32)=3.89, *p*< .001*	11.20 (16.85)		t(30)=2.14, *p*< . 05*	6.51 (16.22)	
uː	t(32)=6.38, *p*< .001*	22.40 (20.13)		t(30)=4.99, *p*< . 001*	17.32 (19.20)	

Table 4.5. Results of paired-samples t-tests on the pre-test and post-test identification scores of the 11 vowels and accuracy gains obtained by the ID and ART training groups.

Subsequent ANOVAS were conducted separately on pre-test and post-test identification scores, with *vowel type* (11) as within-subjects factor and *type of training* (ID vs. ART) as between-subjects factor, to check (1) whether vowels were perceived at different accuracy rates at pre-test and (2) whether pre-test differences as a function of vowel were maintained at post-test. The results yielded a significant effect of *vowel type* both at pre-test (F(10, 620)=42.26, p<.001, η^2= .433) and post-test (F(10, 620)=70.83, p<.001, η^2= .533) and the interaction between *vowel type* and *type of training* did not reach statistical significance. Posthoc tests with pair-wise comparisons revealed that, both at pre-test and post-test, the mean percent correct identification of mid vowels /e/ and /ɜː/, which reached ceiling performance, was significantly different from the rest of vowels at the p< .001 level. The high back vowels /ʊ/ and /uː/ were identified at chance level and differed significantly from the rest of vowels at pre-test, but at post-tes /ʊ/ received the lowest accuracy in perception whereas /uː/ was identified at better than chance levels, although both were significantly different from the rest at the p> .001 level.

Next, independent-samples t-tests were conducted on pre-test and post-test scores for each vowel to assess whether the lack of significant differences as a function of type of training (generally observed in the previous ANOVA) was consistent across all target vowels. The tests revealed only significant between-group differences at post-test for /e/ ($t(62)=2.20$, $p< .05$), suggesting that ID and ART training groups ($T2_{ID}$= 97.7, $gains_{ID}$: 17.32; $T2_{ART}$= 94.66, $gains_{ART}$: 10.16) only differed significantly in their post-test performance for /e/. In summary, ID and ART training groups were similar with respect to their identification gains and only differed with respect to the accuracy with which /e/was perceived after training (see Table 4.6).

Vowel Identification	Between-group differences: ID **(N=32)** vs. ART training **(N=32)**	
	T1	**T2**
iː	t(62)=0.24, p= .814	t(62)=0.31, p= .757
ɪ	t(62)=1.33, p= .187	t(62)=0.38p=.703
e	t(62)=0.60, p= .549	t(62)=2.20, **p< . 05***
ɜː	t(62)=0.86, p= .394	t(62)=0.25, p= .802
æ	t(62)=1.53, p= .131	t(62)=0.65, p= .518
ʌ	t(62)=1.30, p= .197	t(62)=0.28, p=.776
ɑː	t(62)=0.44, p= .660	t(62)=0.36, p=.718
ɒ	t(62)=0.32, p= .747	t(62)=1.63, p=.107
ɔː	t(62)=0.02, p= .983	t(62)=1.35, p= .181
ʊ	t(62)=0.63, p= .533	t(62)=0.88, p=.383
uː	t(62)=0.6, p= .496	t(62)=0.46, p=.644

Table 4.6. Results of independent-samples t-tests on the identification scores of the 11 vowels at pre-test and post-test.

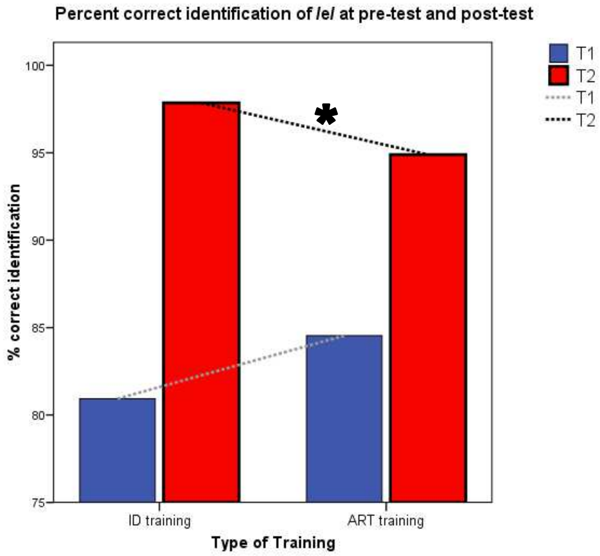


Figure 4.7. Pre-test and post-test identification scores of /e/ obtained by ID and ART training groups and lines indicating between-group differences at pre-test and post-test (p< .05*).

Summary

The findings suggest that training improved overall English learners' identification performance about 14.87 % (M_{T1}= 64.80 vs. M_{T2}= 79.67; $p<.05$) and learning generalized to all vowels. The effect of training was significant for all the English vowels, but a main effect of vowel was found on post-test identification performance. That is, some English vowels were more accurately perceived overall than others after training. Post-test identification performance revealed that the highest percentage of identification gains was obtained for the high-back vowel /uː/ (gains: 14.87%), although high back vowels were the most confused vowels at post-test (/ʊ/: 56.25; /uː/: 66.60). The best post-test performance was found for the mid vowels/e/ (T2: 96.42 %) and /ɜː/ (T2: 94.20%), a notable improvement was seen for /ɑː/ (T2: 83.59%; gains: 13.35%) and /ɪ/ (T2. 80.21 %; gains: 17.45), and only moderate improvement was reported for /ɔː/ (gains: 9.24%) and /ʊ/ (gains: 8.84%). Trainees improved considerably for /uː/ (gains: 19.86), /ɜː/ (gains: 17.90), /ɪ/ (gains: 17.45)) (average 18 percentage points) and improved about 12 % for the rest (/æ/= 12.37), /ʌ/=13.87, /ɑː/= 13.35).

When the effect of type of training on the post-test identification of each vowel was explored, no evidence of significant advantage for any training group was shown except for the fact that the ID training group identified /e/ significantly better than ART training groups at post-test (t(62)= 2.02, p= .048).

4.1.5. Effects of training, type of training and stimulus presentation on accuracy

Table 4.7 shows mean correct identification of vowels in audiovisual (AV) and auditory (A) stimulus presentation conditions (RQ1.1.1c). At post-test, trainees identified vowels in both AV and A conditions at higher accuracy rates than at pre-test, although visual inspection of the group data revealed they performed better in AV than in A conditions. Figure 4.8 displays the pre-test and post-test vowel identification accuracy scores in different stimulus conditions for the ID and ART training groups (RQ2.2c).

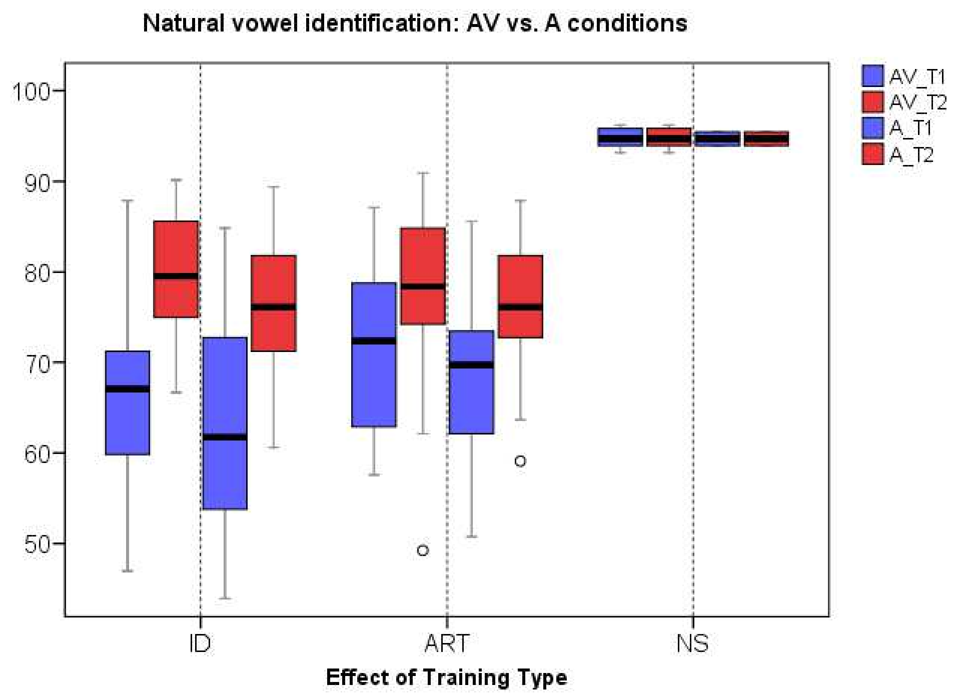


Figure 4.8. Percent correct identification of vowels in audiovisual (AV) and auditory (AV) stimuli presentation conditions by ID training, ART training groups and NS group.

Subject groups	Stimulus presentation					
	AV			A		
	T1	T2	gains	T1	T2	gains
ID training (*N*=64)	64.69 (19.53)	82.21 (6.04)	17.52 (18.51)	60.90 (18.79)	**78.60** **(7.22)**	**17.70** **(18.52)**
ART training (*N*=32)	68.11 (20.06)	79.52 (9.49)	11.42 (20.79)	65.76 (19.6)	**78.27** **(7.10)**	**12.51** **(19.68)**
Control (*N*=20)	65.23 (17.90)	70.07 (5.85)	4.85 (16.87)	61.10 (16.84)	67.39 (6.68)	6.29 (15.78)

Table 4.7. Percent correct identification of vowels, gains and standard deviations in audiovisual and auditory stimuli presentation conditions obtained by ID training and ART training groups.

Data were analyzed using a three-way ANOVA with *training* (T1 vs. T2) and *stimulus presentation* (AV vs. A) as within-subjects factors, and *type of training* (ID vs. ART) as between-subjects factors, to investigate the effects of stimulus presentation and training on vowel identification accuracy (RQ1.1.1c) and whether improvement in identification differed as a result of type of training (RQ2.2c).

This ANOVA revealed a main effect of *training* ($F(1, 62)=38.94$, $p<.001$, $\eta^2= .386$) and *stimulus presentation* ($F(10, 620)=16.88$, $p<.001$, $\eta^2= .214$), but the effect of *type of training*($F(1, 62)=0.235$, $p=.630$, $\eta^2= .004$) did not reach statistical significance. All the interactions were non-significant: *training* x *stimulus presentation* ($F(1, 62)=16.88$, $p=.528$, $\eta^2= .006$), *stimulus presentation* x *type of training* ($F(1, 62)=2.02$, $p=.160$, $\eta^2= .032$), *training* x *type of training* ($F(1, 62)=1.42$, $p=.238$, $\eta^2= .022$), *training* x *stimulus presentation* x *type of training* ($F(1, 62)=3.35$, $p=.205$, $\eta^2= .652$). The fact that the effect of type of training was not significant despite showing that the ART group displayed larger gains than the ID group (Table 7) suggests that the overall improvement is driven only by some subjects but is not general across all participants from the ART group.

These results indicate that training was effective in improving vowel identification in both AV and A conditions. The fact that trainees improved vowel identification in both AV/A conditions suggests that trainees were able to successfully integrate information contained in the (trained) auditory and visual modalities. Importantly, trainees were significantly better in AV condition than A condition after training, as higher accuracy was obtained for the AV stimuli than for their A counterparts at post-test. Considering that both ID and ART training were audiovisual types of high-variability training, it could be argued that both types of training are likely to be effective in making learners successfully more sensitive to visual cues. The results revealed that the type of training did not have a significant effect on the amount of improvement in AV and A stimulus presentation conditions, and the training groups were not significantly different in any of these two conditions.

Summary

HVPT significantly improved English learners' identification performance. Higher accuracy in natural vowel identification was obtained in AV condition (M_{T1}=66.36, M_{T2}=80.89; gains= 14.53) than in A condition (M_{T1}=62.48, M_{T2}=78.44; gains= 15.96). Both ID and ART training groups peformed similarly better when presented with auditory and audiovisual stimuli at post-test.

4.1.6. Effects of training, type of training and word type on accuracy

This section explores the effects of training (RQ1.1.1d) and type of training (RQ2.2d) on the learner's ability to correctly identify vowels and the ability to generalize learning to new words. Table 4.8 presents the group correct identification scores across trained and new words for the three subject groups. Both the ID and ART training groups showed improvement in correct vowel identification for trained words and new (untrained) words, whereas the NNS control group remained almost the same. Figure 4.9 shows that mean post-test performance was always greater than the mean performance prior to training.

From the results shown on Table 4.8 and Figure 4.9, it is apparent that trainees showed the most improvement when identifying vowels from trained words and only slightly poorer performance when those same vowels were presented in new words. At post-test, learners identified trained words at significantly higher accuracy rates than new words. One issue that deserves further investigation is the fact that the ID group outperformed the ART one for trained words, that is, learners generally identified vowels embedded in stimuli that had frequently appeared during the training period better than untrained words after ID training. This result suggested that the words used in the training appeared to be more easily identified by the ID training group than the „new" or unfamiliar words.

Subject groups	Word type					
	trained			new		
	T1	T2	gains	T1	T2	gains
ID training (*N*=64)	63.98 (19.00)	82.14 (5.50)	18.85 (3.28)	61.62 (19.46)	**78.67 (8.42)**	**17.06 (18.58)**
ART training (*N*=32)	69.77 (20.19)	79.18 (8.79)	9.41 (21.78)	64.10 (19.55)	**78.62 (9.16)**	**14.52 (17.78)**
Control (*N*=20)	67.01 (18.05)	*72.42 (6.91)*	5.42 (16.24)	59.31 (16.70)	65.04 (6.11)	5.73 (15.44)

Table 4.8. Pre-test, post-test and gain scores for trained and new words obtained by ID training, ART training and control groups.

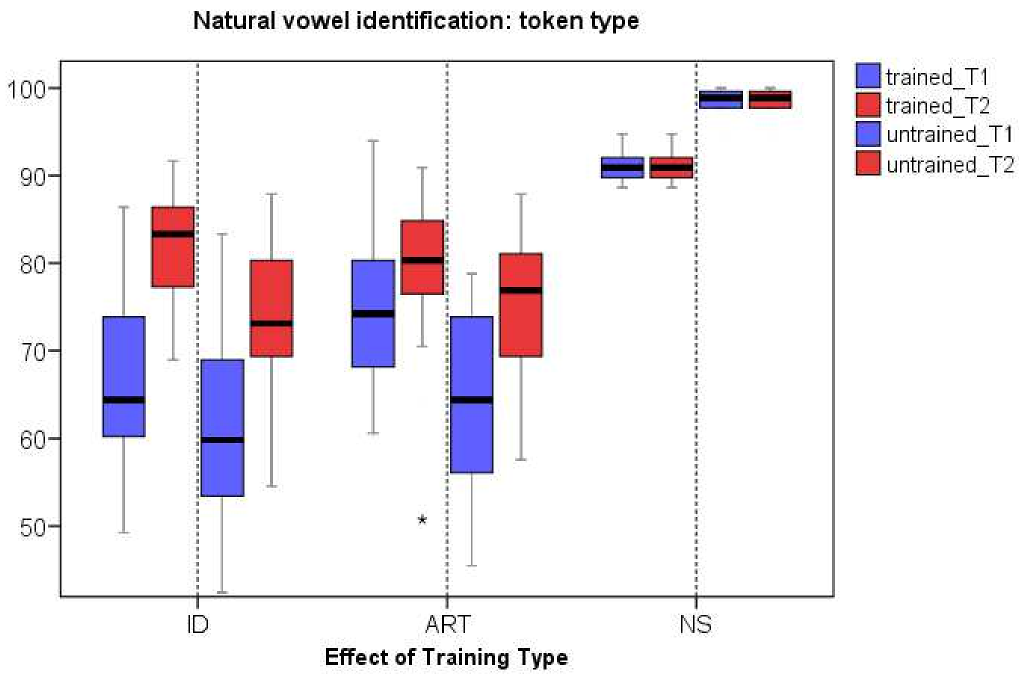


Figure 4.9. Pre-test and post-test identification scores for trained and new words obtained by ID training, ART training and NS groups.

To further explore this pattern of results, a three-way ANOVA with *training* (T1 vs. T2) and *word type* (trained vs. new) as within-subjects factors, and *type of training* (ID vs. ART) as between-subjects factor, investigated the effects of word type and training on vowel identification accuracy (RQ1.1.1d) and explored whether effects differed as a result of type of training (RQ2.2d). This ANOVA yielded a significant main effect of *training* ($F(1, 62)=17.00$, $p<.001$, $\eta^2= .215$) as well as a significant effect of *word type* ($F(1, 62)=38.94$, $p<.001$, $\eta^2= .386$), but the effect of *type of training* was not significant ($F(1, 62)=0.24$, $p= .630$, $\eta^2= .004$). The interactions *word type* x *training*, and *word type* x *type of training* were not significant. However, the significant interaction *training xword type x type of training* ($F(1, 62)=154.08$, $p<.05$, $\eta^2= .102$) appeared to indicate that there was a two-way interaction (training x word type) varying across the levels of a third variable, *type of training* (ID vs. ART).

Separate repeated-measures ANOVAS run for the ID and ART training groups revealed a significant effect of *training* and *word type* in each of the training groups: ID training (effect of *training*: $F(1, 32)= 22.64$, $p< .005$, $\eta^2= .267$), effect of *word type* ($F(1, 62)=30.75$, $p< .001$, $\eta^2= .490$), and ART training (effect of *training*: $F(1, 30)=6.69$, $p< .05$, $\eta^2= .182$), effect of *word type* ($F(1, 62)=11.46$, $p< .005$, $\eta^2= .276$). The results indicated that training was overall effective for improving vowel identification of trained words and improvement, interestingly, generalized to new (untrained) words. The significant three-way interaction obtained suggests that the effect of training differed across the two training groups: the ID training group outperformed the ART group in the identification of trained words at post-test. ID (improvement of 18%) and ART (improvement of 9%) training groups differed in the amount of improvement they made with respect to trained words. After ID training, learners identified trained words more accurately than new words, whereas ART trainees performed similarly in the two word type conditions. This could be interpreted as trainees being more likely to remember words previously included in identification tasks used for training (ID training) than words that they had been previously asked to imitate during the articulatory training sessions (ART training).

Summary

The findings point to a successful transfer of learning to the perception of the same vowel in new contexts (new words) and trained words. HVPT significantly improved English learners' identification performance with trained and new words. Trainees performed better with trained words (M_{T1}=66.25, M_{T2}=80.71; gains= 14.46) than novel words (M_{T1}= 62.58, M_{T2}= 78.63; gains= 16.05). However, generalization of improvement to untrained words (new contexts) is a valuable finding of ID and ART types of training.

4.1.7. Effects of training, type of training and talker conditions on accuracy

This section attempts to add to the results previously presented concerning the effects of training by providing data on the magnitude of the overall training effects when variables such as the *talker condition* (trained vs. untrained) is examined (RQ1.1.1d). It explores the learner's ability to generalize learning to new (untrained or unfamiliar) talkers and compares ID and ART training methods in terms of generalization to *untrained talkers* (RQ2.2d).

Descriptive statistics for perception accuracy in the trained and untrained talker conditions are shown in Table 4.9. At post-test, trainees performed significantly better in vowel identification regardless of the talker condition, which indicates that training generalized to new or unfamiliar talkers. The NNS control group showed almost no improvement from pre-test to post-test for words presented by the trained or untrained talkers, and there was very little difference in post-test performance for the vowels uttered by trained talkers (M=5.86 SD=16.47; $t(19)$=-1.59, p= .128), relative to those of the untrained talkers (M=7.46 SD=16.67; $t(19)$=-2.00, p= .060) (see Table 4.1.9).

Figure 4.10 examines generalization to talkers and shows, for the two training groups, the average percent correct identification at pre-test and post-test, in trained and untrained talker conditions. Compared to the ART training group, the ID training group performed better in the trained talker than in the untrained talker condition at post-test. For the ART training group, post-test vowel identification was similar across the two talker conditions.

Subject groups	Talker condition					
	trained			untrained		
	T1	T2	gains	T1	T2	gains
ID training (*N*=64)	64.91 (20.58)	84.71 (5.67)	19.79 (20.03)	63.07 (20.80)	**81.13 (8.68)**	**18.06 (19.97)**
ART training (*N*=32)	71.95 (20.82)	81.65 (9.07)	9.70 (22.45)	66.10 (20.16)	**81.05 (8.42)**	**14.94 (19.37)**
Control (*N*=20)	68.75 (18.22)	74.61 (7.03)	5.86 (16.47)	56.95 (17.14)	67.11 (6.35)	7.46 (16.67)

Table 4.9. Pre-test, post-test and gain scores in trained talker and untrained talker conditions obtained by ID training, ART training and control groups.

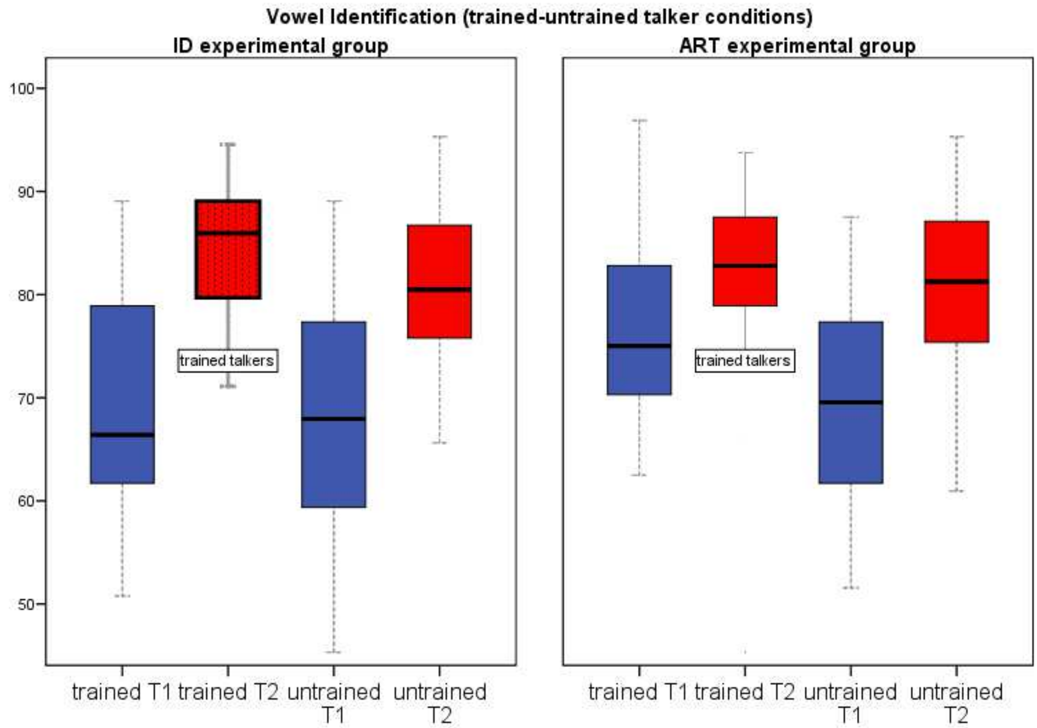


Figure 4.10. Pre-test, post-test and gain scores in trained talker and untrained talker conditions obtained by ID training, ART training and control groups.

A three-way ANOVA with *training* (T1 vs. T2) and *talker condition* (trained vs. untrained) as within-subjects factors, and *type of training* (ID vs. ART) as between-subjects factor, was conducted on identification scores to explore the effect of training (RQ1.1.1d) and type of training (RQ2.2d) on vowel identification and explore the effect of talker condition on identification accuracy. Results revealed a significant effect of *training* ($F(1, 62)=15.03$, $p< .001$, $\eta^2= .195$) and a significant effect of *talker condition* ($F(1, 62)=39.45$, $p< .001$, $\eta^2= .389$) on identification accuracy. The effect of *type of training (*$F(1, 62)= 0.37$, $p= .546$, $\eta^2= .006$) did not reach statistical significance but there was a significant three-way interaction of effect of *training, talker condition*and *type of training* ($F(1, 62)=49.34$, $p< .05$, $\eta^2= .033$). The other interactions were not significant either.

To examine the factors driving the three-way interaction, additional tests were conducted. Two-way repeated-measures ANOVAs were conducted for each type of training (ID vs. ART) separately, with effect of *training* and *talker condition* as within-subjects factors. For the ID training group, there was a significant effect of *training* ($F(1, 32)=8.75$, $p< .05$) and *talker condition* ($F(1, 32)=31.04$, $p< .001$) and the interaction of *training* and *talker condition* did not reach significance. These results indicated that ID training was significantly effective in improving vowel identification in both conditions and promoting generalization to untrained talkers. They also revealed that the degree of familiarity of the talker played a role in the accuracy with which vowels were correctly perceived at post-test. After ID training, learners performed significantly better when they heard the trained talkers than when they heard the unfamiliar talkers, $t(32)=2.96$, $p<.05$. For the ART group, the two-way ANOVA yielded a significant effect of *training* ($F(1, 30)=6.74$, $p< .05$) and *talker condition* ($F(1, 30)=11.44$, $p< .005$, $\eta^2= .195$) as well as a significant *training x talker condition* interaction ($F(1, 30)=7.52$, $p< .05$). This two-way interaction suggested that the effect of training on vowel perception varied as a function of talker condition.

Next, paired-samples t-tests were conducted for each training group separately to compare identification scores across the two talker conditions at each testing time. The tests yielded significant differences between the post-test identification scores obtained across talker conditions for the ID training group, $t(32)=2.96$, $p< .05$ (ART T2: $M_{trained}$= 84.71, $M_{untrained}$= 81.65). However, differences between trained and untrained talker conditions at post-test did not turn out significant for the ART training group, $t(30)=0.324$, $p= .748$ (ART T2: $M_{trained}$= 81.13, $M_{untrained}$= 81.05) (see Figure 4.10).

This section explored the generalization of phonetic learning to new talkers. The results showed significant improvement in vowel identification for both training groups in the trained and untrained talker conditions. It can be concluded that both ID and ART training generalized to untrained talkers. Importantly, this study shows the training is effective in generalizing learning to untrained talker conditions, probably because a greater number of vowels uttered by a multiplicity of talkers were used in this training study. It can be concluded that audiovisual high-variability phonetic training (with context, word and talker variability) promotes successful generalization across talkers in vowel identification, which suggests that trainees were able to abstract acoustic-phonetic features from trained voices and materials and, therefore, generalization of phonetic learning to new voices occurred. Previous literature (Lively, Logan, &Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively,&Pisoni, 1991; Clopper&Pisoni, 1994; Burk et al., 2006) has focused on the role of input in talker generalization of phonetic learning, and these results suggest that this process appears to be facilitated by the memory encoding processes that follow ID and ART training. We note, however, that the two training groups differed in the *trained talker* condition: although both ID and ART training groups improved across different talker conditions, the ID training group showed greater improvement than the ART training group in the *trained talker* condition. Moreover, the ID group performed significantly better overall on the words presented by *trained talkers* (easy-familiar) than *untrained* (hard-unfamiliar) *talkers*, despite overall improvement in vowel identification.

These results suggest, on the one hand, that in order to improve English vowel perception for non-native listeners, a great number of vowels, or the entire vowel inventory, as well as talker variability are needed. The between-group differences with regards to performance on trained words (i.e. the greater generalization to unfamiliar talkers shown by the ID group) suggests that the type of feedback administered through perceptual training might have played a role in the degree of success in the trained words. This involves the type of auditory-orthographic feedback used during ID training.

Summary

The two training methods were successful at inducing listeners to significantly improve their ability to identify English vowels uttered by trained and untrained talkers. Besides, the findings point to a successful transfer of learning to the perception of the same vowel produced by new talkers. It must be noticed, though, that the ID training group performed significantly better with trained than with untrained talkers, despite overall improvement in vowel identification. Generalization of improvement to new talkers is a valuable finding of these two types of HVPT.

4.1.8. Effects of training and type of training on error dispersion

This study examined whether training increased accuracy in vowel identification while it decreased vowel dispersion. The main goal of this section was to assess the effectiveness of training (RQ1.1.2) and the type of training (RQ2.2f) on mean error dispersion in vowel identification (ID1-task1), that is, the number of error categories and the distribution of responses for vowel stimuli, with a view to exploring what factors might affect error dispersion.

To examine the confusions in vowel identification in detail, Tables 4.10 and 4.11 show the percentage of correct vowel identification as well as the mean confusion matrices (Miller & Nicely, 1955) for vowels at pre-test and post-test for the ID and ART training groups. The following observations were made from the confusion matrices: (1) post-test correct identification of vowels (values along the diagonal) increased substantially, (2) the distribution of incorrect responses suggested that listeners made "actual" confusions (2-3 confusions) rather than random guessing (i.e. vowels that had similar formants were confused with one another, e.g. /iː/ and /ɪ/, /æ/ and /ʌ/, /ɔː/ and /ʊ/), (3) the frequency with which each vowel was mistakenly identified as one of the vowels of a closed set decreased substantially at post-test (e.g. misidentification of /iː/ as /ɪ/ decreased from 18.4% at pre-test to 8.7% at post-test) and, (4) apparently, ID and ART training groups had mostly the same pattern of confusions and revealed similar changes after training.

Stimulus		Multiple Response (%) – ID training group										
		/iː/	/ɪ/	/e/	/ɜː/	/æ/	/ʌ/	/ɑː/	/ɒ/	/ɔː/	/ʊ/	/uː/
/iː/	T1	**60.5**	18.4	3.0	3.6							
	T2	**76.2**	8.7	0.1	0.3							
/ɪ/	T1	26.9	**58.7**	2.1	3.0							
	T2	20.4	**80.1**	0.1	0.3							
/e/	T1	1.6	10.5	**80.3**	10.4							
	T2	0.9	7.8	**97.8**	4.0							
/ɜː/	T1	1.6	3.5	6.0	**73.0**							
	T2	0.3	1.8	0.9	**93.9**							
/æ/	T1					**59.1**	16.0	5.5	6.5			
	T2					**75.2**	11.3	4.0	3.1			
/ʌ/	T1					16.8	**54.5**	13.4	14.4			
	T2					16.7	**78.6**	9.2	12.6			
/ɑː/	T1					13.8	13.7	**68.0**	11.1			
	T2					4.4	5.3	**83.7**	5.7			
/ɒ/	T1					1.6	6.3	4.8	**57.0**			
	T2					-	1.9	1.7	**76.0**			
/ɔː/	T1									**59.4**	25.4	5.3
	T2									**71.5**	27.6	2.2
/ʊ/	T1									14.3	**46.0**	41.1
	T2									7.3	**57.2**	28.1
/uː/	T1									16.7	19.0	**44.7**
	T2									18.2	12.9	**67.1**

Stimulus		Multiple Response (%) – ART training group										
		/iː/	/ɪ/	/e/	/ɜː/	/æ/	/ʌ/	/ɑː/	/ɒ/	/ɔː/	/ʊ/	/uː/
/iː/	T1	**62.9**	14.3	0.5	1.2							
	T2	**75.0**	9.8	0.9	0.5							
/ɪ/	T1	28.2	**66.9**	4.2	2.9							
	T2	20.8	**80.1**	1.4	1.0							
/e/	T1	0.5	7.3	**84.6**	9.4							
	T2	1.8	6.8	**94.7**	3.9							
/ɜː/	T1	0.4	2.3	2.9	**79.2**							
	T2	0.9	2.5	2.5	**94.1**							
/æ/	T1					**71.2**	15.7	6.1	7.7			
	T2					**78.5**	13.4	4.3	4.9			
/ʌ/	T1					13.0	**62.5**	12.1	11.8			
	T2					13.1	**80.5**	10.5	11.6			
/ɑː/	T1					6.9	8.3	**72.1**	12.1			
	T2					7.8	3.6	**83.1**	9.0			
/ɒ/	T1					0.9	4.3	2.3	**59.8**			
	T2					0.1	1.9	1.7	**72.5**			
/ɔː/	T1									**60.8**	23.3	4.2
	T2									**67.8**	23.7	4.0
/ʊ/	T1									11.7	**48.8**	39.1
	T2									11.1	**55.3**	26.9
/uː/	T1									19.1	18.0	**48.8**
	T2									19.0	20.2	**66.1**

Tables 4.10-4.11. Vowel confusion patterns observed at pre-test (T1) and post-test (T2) for ID and ART training groups. Stimulus vowels are listed vertically; response categories are listed horizontally. Grey shadowed numbers in boldface (values along a diagonal) indicate average percent correct identification across vowels. Grey numbers indicate infrequent incorrect responses for a particular stimulus-response combination, that is, stimuli which are rarely chosen as a response (<4% accuracy).

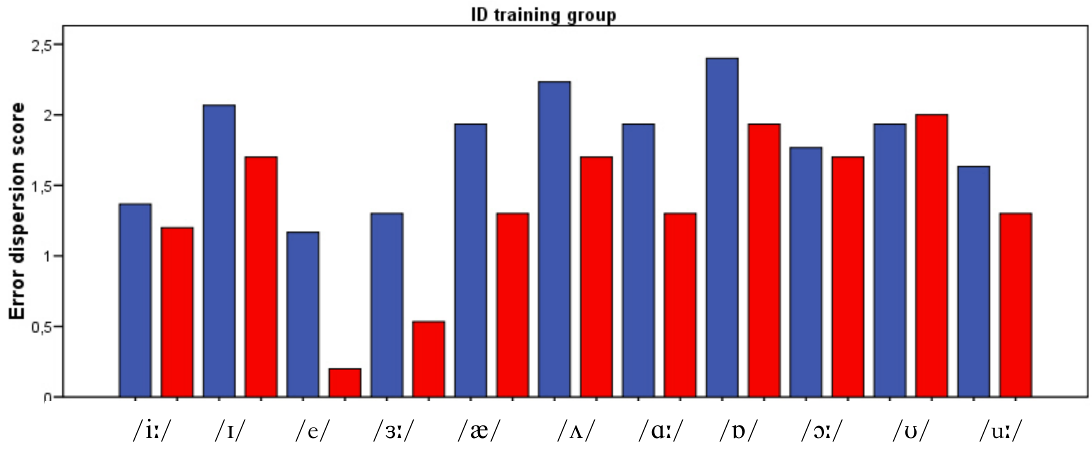
This section analyses the changes in error dispersion as a measure of the effectiveness of phonetic training in "the number of error classes per stimulus" (RQ1.1.2), in order to know whether there is a real difference between the distribution of responses at each testing time and between ID and ART training groups (RQ2.2f).

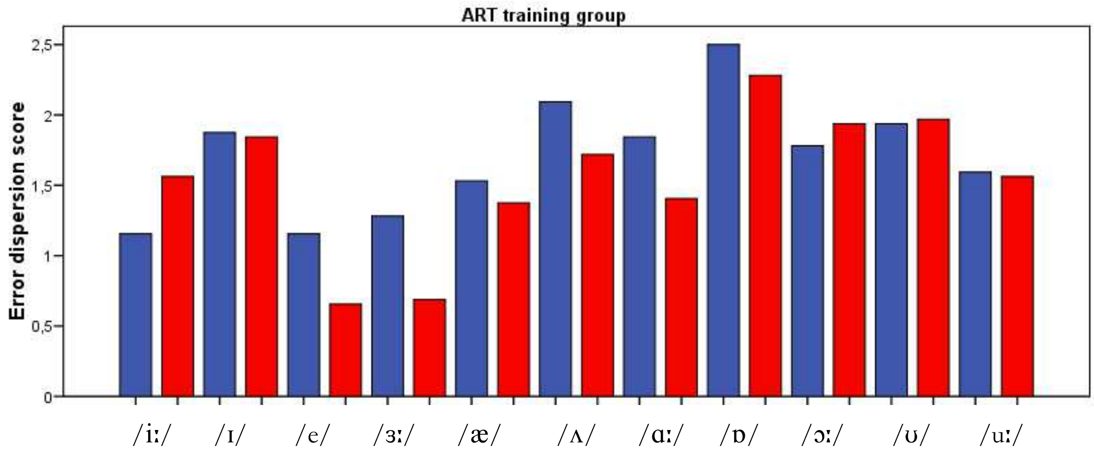
Table 4.12 shows descriptive statistics for mean error dispersion at pre-test and post-test as ameasure of distribution of incorrect responses in vowel identification, as well as a measure of the spreading of the errors over individual vowel stimulus and response categories (Son, 1994).

The descriptive data shows that, at post-test, error dispersion scores were lower and standard deviations decreased for both ID and ART training groups. However, diverging patterns of vowel (mis-)identification were found: the ID training group obtained lesser degree of dispersion of wrong responses (T1: M= 1.79; T2: M= 1.35) than the ART group (T1:M= 1.70; T2: M= 1.54) after training (see Table 4.1.12). To facilitate comparison of the error dispersion patterns at pre-test and post-test, the error dispersion values (averaged across the talker means for each vowel and training group) are shown in Figures 4.1.11-4.1.12. It will be observed that the ID training group showed more visible changes or differences in error dispersion scores than the ART group from pre-test to post-test, especially for /iː/, /e/, /ɒ/, /ɔː/ and /uː/.

Training groups		Error dispersion scores (vowel identification- ID1)											
		/iː/	/ɪ/	/e/	/ɜː/	/æ/	/ʌ/	/ɑː/	/ɒ/	/ɔː/	/ʊ/	/uː/	Mean ErrDisp
ID training	T1	1.38 (0.75)	2.13 (0.91)	1.16 (0.95)	1.38 (0.98)	1.94 (0.84)	2.25 (0.88)	1.94 (0.95)	2.41 (0.87)	1.78 (0.55)	1.94 (0.36)	1.65 (0.61)	1.79 (0.58)
	T2	1.19 (0.47)	1.69 (0.86)	0.19 (0.40)	0.53 (0.72)	1.31 (0.69)	1.69 (0.78)	1.31 (0.93)	1.97 (0.74)	1.72 (0.52)	2.00 (0.00)	1.28 (0.46)	1.35 (0.32)
	diff.	0.18 (0.64)	0.44 (1.10)	0.97 (1.00)	0.84 (0.95)	0.62 (1.04)	0.56 (1.01)	0.62 (1.04)	0.44 (1.04)	0.06 (0.62)	0.06 (0.36)	0.35 (0.61)	0.44 (0.51)
ART training	T1	1.16 (0.51)	1.88 (0.91)	1.16 (0.88)	1.28 (0.96)	1.53 (0.84)	2.09 (0.89)	1.84 (0.99)	2.50 (0.80)	1.78 (0.55)	1.94 (0.35)	1.59 (0.61)	1.70 (0.52)
	T2	1.56 (0.56)	1.84 (0.99)	0.66 (0.79)	0.69 (0.90)	1.38 (0.79)	1.72 (0.81)	1.41 (0.95)	2.28 (0.58)	1.94 (0.25)	1.97 (0.17)	1.56 (0.50)	1.54 (0.39)
	diff.	0.41 (0.76)	0.03 (1.23)	0.50 (0.88)	0.59 (1.01)	0.16 (0.88)	0.37 (1.16)	0.44 (1.16)	0.22 (0.97)	0.16 (0.57)	0.03 (0.40)	0.03 (0.59)	0.16 (0.55)

Table 4.12. Pre-test and post-test error dispersion in vowel identification, pre-test/post-test differences (diff.) and standard deviations for individual vowels obtained by ID and ART training groups.





Figures 4.11-4-12. Mean error dispersion scores for individual vowels at pre-test (blue) and post-test (red) obtained by ID and ART training groups.

A three-way ANOVA on dispersion scores, with *training* (T1 vs. T2) and *vowel* (11) as within-subjects factors, and *type of training (ID vs. ART)* as between-subjects factor, revealed a significant effect of *training* ($F(1, 60)=19.65$, $p< .001$, $\eta^2= .247$) and *vowel* ($F(10, 600)=52.65$, $p< .001$, $\eta^2= .467$). The interactions between the effect of *training* and *type of training* ($F(1, 60)=4.36$, $p< .05$, $\eta^2= .068$), and between *training* and *vowel* ($F(10, 600)=9.32$, $p< .001$, $\eta^2= .134$) were statistically significant. The effect of *type of training* was not statistically significant ($F(1, 60)=0.29$, $p= .591$, $\eta^2= .005$). The rest of the interactions did not reach significance. These results suggest, on the one hand, that pre-test/post-test differences in error dispersion differed among training groups and, on the other hand, they indicated that these changes in error dispersion differed across vowels.

To further explore these two-way interactions, two separate ANOVAS with *training* (T1 vs. T2) and *vowel type* (11) as within-subjects factors were conducted for each of the training groups. For the ID training group, the ANOVA yielded a significant effect of *training* ($F(1, 29)=22.17$, $p< .001$, $\eta^2= .433$) and *vowel* ($F(10, 290)=27.77$, $p< .001$, $\eta^2= .489$), and a significant *training* x *vowel* interaction ($F(10, 290)=5.06$, $p< .001$, $\eta^2= .149$). For the ART training group, the test also revealed a significant effect of *vowel* ($F(10, 310)=25.86$, $p< .001$, $\eta^2= .455$) and a significant *training* x *vowel* interaction ($F(10, 310)=5.07$, $p< .001$, $\eta^2= .000$). The two-way interactions indicated that the effects of training varied as a function of vowel for both ID and ART training groups.

When two-way ANOVAS, with *training* as within-subjects factor and *type of training* as between-subjects factors, were conducted on error dispersion scores separately for each of the vowels, a significant effect of *training* was found on error dispersion for most of the vowels (see Table 4.13). This meant that the training was effective in reducing the effective number of different error classes in the identification of /iːeɜːæʌɑːɒʊ/, which were identified with significantly lower dispersion at post-test. The degree of confusion was considerably reduced for 8 out of 11 vowels, a sign of acquisition. The effect of *training* did not reach significance for /ɪɔːuː/, indicating that the spreading of errors over these three vowels did not change significantly as a result of training. The results yielded a significant *training x type of training* interaction for /eæuː/, suggesting that the *type of training* played a role on the extent to which training diminished the degree of error dispersion for these three vowels.

Error dispersion scores (vowel identification)

Vowel	Effect of *training*				Effect of *type of training*				Two-way interaction *Training x type of training*			
	df	F	*p*	η^2	*df*	F	*p*	η^2	*df*	F	*p*	η^2
iː	1	1.55	.218	.024	1	0.45	.507	.007	1	11.43	**<.001***	.156
ɪ	1	1.76	.114	.040	1	0.07	.791	.001	1	1.93	.170	.030
e	1	38.93	**<.001***	.386	1	2.22	.141	.035	1	3.96	**.05***	.060
ɜː	1	34.20	**<.001***	.356	1	0.03	.867	.000	1	1.03	.313	.016
æ	1	16.69	**<.001***	.170	1	1.08	.303	.017	1	4.57	**<.05***	.069
ʌ	1	11.88	**<.001***	.161	1	0.15	.699	.002	1	0.47	.493	.008
ɑː	1	14.86	**<.001***	.193	1	0.00	1.00	.000	1	0.46	.499	.007
ɒ	1	6.74	**<.05***	.098	1	2.07	.155	.032	1	0.75	.390	.012
ɔː	1	0.39	.532	.006	1	1.30	.258	.021	1	2.15	.148	.033
ʊ	1	0.99	.322	.016	1	0.10	.758	.002	1	0.12	.730	.002
uː	1	6.49	**<.05***	.096	1	0.90	.347	.015	1	4.56	**<.05***	.070

Table 4.13. Results of two-way ANOVAS run on error dispersion scores for individual vowels, with *training* as within-subjects factor and *type of training* as between-subjects factors.

In fact, follow-up independent-samples t-tests revealed significant between-group differences in the identification of /eæuː/ at the p< .05 level, and suggested that the extent to which errors were "dispersed" at post-test varied as a function of the *type of training* (see Table 4.14). First, the ID training obtained significantly lesser degree of error dispersion for /æuː/ at post-test, whereas the error dispersion scores for these vowels did not change after ART training. Second, the ID group obtained significantly lower error dispersion at post-test (T2: 0.19; pre-test/post-test differences: 0.95) than ART group (T2: 0.66; pre-test/post-test differences: 0.50) for /e/. Figure 4.13 shows two different patterns of responses in the identification of /e/for ID and ART training groups.

Vowel Identification	Between-group differences: ID **(N=32)** vs. ART training **(N=32)**	
	T1	**T2**
iː	t(62)=1.36, p= .179	t(62)=-2.89, p< .005*
ɪ	t(62)=1.10, p= .274	t(62)=-0.67, p= .502
e	t(62)=0.00, p= 1.00	t(62)=-3.01, p< . 005*
ɜː	t(62)=0.39 p= .699	t(62)=-0.77, p= .444
æ	t(62)=1.93, p= .058	t(62)=-0.34, p= .738
ʌ	t(62)=0.70, p= .483	t(62)=-0.16, p= .876
ɑː	t(62)=0.39, p= .700	t(62)=-0.40, p= .691
ɒ	t(62)=-0.45, p= .657	t(62)=-1.88, p= .065
ɔː	t(62)=000, p= 1.00	t(62)=-2.14, p< .05*
ʊ	t(61)=-0.02, p= .982	t(62)=1.00, p= .321
uː	t(61)=0.33, p= .740	t(62)=-2.34, p< .05*

Table 4.14. Results of independent-samples t-tests on the pre-test and post-test error dispersion scores of individual vowels.

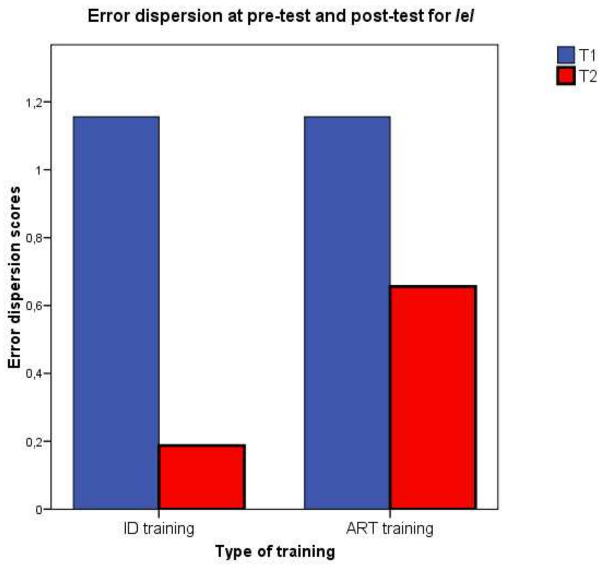


Figure 4.13. Pre-test and post-test error dispersion scores obtained for /e/ at pre-test and post-test by ID and ART training groups.

Next, separate paired-samples t-tests were conducted for ID and ART training groups to explore pre-test/post-test differences in error dispersion for each of the vowels (see Table 4.15). The results revealed a significant decrease in error dispersion for 8 vowels (/ɪeɜːæʌɑːɒuː) after ID training, and significantly lower error dispersion for 4 vowels (/iːeɜːɑː/) after ART training at the p< .05 level.

Vowel	Paired comparisons					
	ID training (*N*=32)			ART training (*N*=32)		
	df t p-value	*M*	*SD*	df t p-value	*M*	*SD*
iː	t(31)=1.64, p= .110	0.18	(0.64)	t(31)=-3.24, **p< . 005***	0.41	(0.76)
ɪ	t(31)=2.24, **p< .05***	0.44	(1.10)	t(31)=-2.72, p= . 887	0.03	(1.23)
e	t(31)=5.48, **p< .001***	0.97	(1.00)	t(31)=-2.27, **p< . 005***	0.50	(0.88)
ɜː	t(31)=5.00, **p< .001***	0.84	(0.95)	t(30)=-3.38, **p< . 005***	0.59	(1.01)
æ	t(31)=4.06, **p< .001***	0.62	(1.04)	t(30)=-1.55, p= . 325	0.16	(0.88)
ʌ	t(31)=3.14, **p< .005***	0.56	(1.01)	t(30)=-3.99, p= . 076	0.37	(1.16)
ɑː	t(31)=3.40, **p< .005***	0.62	(1.04)	t(30)=-2.79, **p< . 05***	0.44	(1.16)
ɒ	t(31)=2.37, **p< .05***	0.44	(1.04)	t(30)=-3.75, p= .214	0.22	(0.97)
ɔː	t(31)=0.57, p= .572	0.06	(0.62)	t(30)=-2.01, p= .134	0.16	(0.57)
ʊ	t(30)=-1.00, p= .325	0.06	(0.36)	t(30)=-2.14, p= .662	0.03	(0.40)
uː	t(30)=3-25, **p< .005***	0.35	(0.61)	t(30)=-4.99, p= .768	0.03	(0.59)

Table 4.15. Results of paired-samples t-tests on the error dispersion scores of the 11 vowel monophthongs obtained by ID and ART training.

In addition to the analysis of error dispersion scores, further ANOVAS were conducted on *error dispersion change* (pre-test/post-test differences in in the number of error categories per vowel stimuli), as a measure of the amount of changes (i.e.increase/decrease) in the dispersion of errors per vowel category and stimulus response. Figure 4.14 provides a visual illustration of error dispersion change in natural vowel identification.. As it can be seen, the ID group displayed a larger amount of error dispersion change than the ART group.

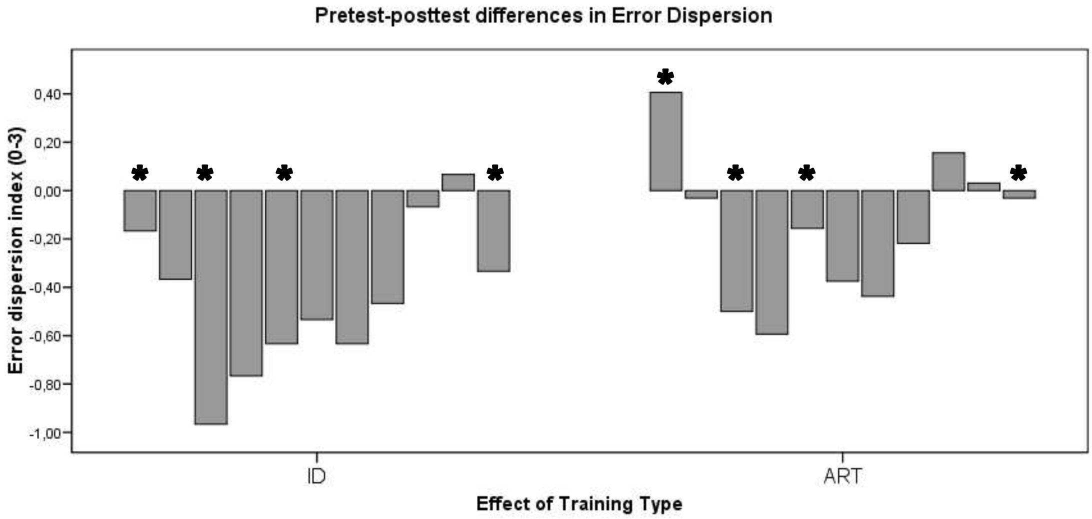


Figure 4.14. Pre-test/Post-test differences in error dispersion in vowel identification obtained by ID and ART training groups, with negative scores (bars upside down) indicating decrease in error dispersion after training and asterisks indicating significant between-group differences (*p*< .05).

A two-way ANOVA performed on *error dispersion change* with *vowel* (11) as within-subjects factor and *type of training* (ID vs. ART) as between-subjects factor yielded a significant effect of *vowel type* ($F(1,60)=6.00$, $p< .05$, $\eta^2= .091$) and a significant effect of *type of training* ($F(1, 60)=4.36$, $p< .05$, $\eta^2= .068$). There were no significant interactions at the $p< .05$. A follow-up independent-samples t-test on gains in error dispersion revealed significant between-group differences for /iːeæuː/. The ID group showed overall greater error dispersion changes than the ART group, but between-group differences turned out to be statistically significant for these four vowels.

This subsection has presented the effects of training on the distribution of responses over vowel stimuli after each type of training. After ID and ART training, learners showed higher accuracy in vowel identification as well as reduced error dispersion per vowel category. Both types of training led to a significant decrease in the number of incorrect response categories used in vowel identification. This can be interpreted as a positive sign of acquisition, indicating that, after 10 training sessions, learners are beginning to disregard some of the variation and be more consistent in their choices (Iverson &

Evans, 2009; (Bundgaard-Nielsen et al., 2011). Importantly, training has proved to be beneficial for gradually leading to a strengthening of the correct vowel category and reduction of the number and weight of the other categories, until "competing" categories get gradually abandoned in their identification choice (Gong et al., 2017).

However, this section has outlined a number of differences between ID and ART training groups with respect to the degree of error dispersion in vowel identification. ID and ART groups showed a significantly different pattern of confusions when they identified /iːeæuː/ at post-test. Whereas the ID group obtained lower error dispersion for /iː/, the ART group showed higher error dispersion at post-test for the same vowel. For /eæuː/, the ID group showed significantly greater decrease in error dispersion than the ART group did. To sum up, the ID training group obtained lower error dispersion than the ART group despite overall changes in the distribution of error responses reported for both groups.

In summary, the ID training group experienced a significantly higher decrease in the number of mistaken "similar" categories assigned to certain vowel stimuli at post-test than the ART group did. This result raises a question about the possible benefits of identification training for better reducing the kind of confusions appearing in vowel identification. The results also suggest that perceptual training may be highly effective and help learners distinguish ambiguous vowel sets that will trigger less possible response categories in vowel identification experiments and therefore better select the response.

Summary

Both types of training led to a significant decrease in error dispersion, the number of incorrect response categories used, in vowel identification. However, ID training (based on a multiple-choice categorization task with immediate AV feedback) was found to be a more effective approach than the ART training in *decreasing the amount of error dispersion in vowel identification. Whereas the ID group obtained lower error dispersion for /iː/, the ART group showed higher error dispersion at post-test for the same vowel. For /eæuː/, the ID group showed significantly greater decrease in error dispersion than the ART group.*

4.1.9. Summary of effects of training and type of training on perception accuracy

This section has presented the effects of training on *correct vowel identification* scores as well as the effects on the distribution of responses over vowel stimuli after each type of training, the degree of *error dispersion*. The results from the pre- and post-test vowel identification task (ID1-task1) have demonstrated that the two training methods were successful in improving listeners' ability to identify English vowels. In addition, the section has highlighted some remarkable differences between ID and ART training methods, showing a higher advantage for the ID training group in the identification of *trained* words and the distribution of error responses over vowel stimuli.

The results confirmed that vowel identification scores at post-test were significantly higher than at pre-test, suggesting that the two training methods were generally effective in improving vowel identification accuracy. Trainees showed significant improvements in vowel identification of 17.63% (ID training) versus 12.88% (ART training). ID and ART types of training were similarly effective in improving vowel identification accuracy.

There was a significant improvement in vowel identification of 10-18% for the three vowel sets (high-front, low and back vowels). However, these results indicate, on the other hand, that the *vowel set* seems to play a role in the learners' ability to perceive English vowels correctly and in the overall training effect. Both ID and ART types of training seemed to have a more visible impact on the high-front and low vowel sets, which were identified more accurately than back vowels at post-test. After each type of training, the pattern of identification remained similar to pre-test despite overall improvement: at post-test vowels in a more retracted mouth position (back vowels: M_{T2}=61-65%) were somehow more resistant to training than high-front (M_{T2}=85-86%) and low vowels (M_{T2}=78%).

Overall training improved learners' identification performance about 14.87% (T1: M=64.80 vs. T2: M=79.67; $p<.05$) and learning generalized to all vowels. However, some English vowels were thus more accurately perceived than others after training. Post-test identification performance revealed that the highest percentage of identification gains was obtained for the high-back vowel /uː/ (14.87%), although high back vowels the most confused vowels at post-test (/ʊ/: 56.25; /uː/: 66.60). The best post-test performance was found for the mid vowels /e/ (T2: 96.42 %) and /ɜː/ (T2: 94.20%), a notable improvement was seen for /ɑː/ (T2: 83.59%; gains: 13.35) and /ɪ/ (T2. 80.21 %; gains: 17.45), and only moderate improvement was reported for /ɔː/ (9.24)and /ʊ/ (8.84). Trainees improved considerably for /uː/ (19.86), /ɜː/ (17.90), /ɪ/ (17.45) (average 18 percentage points) and improved about 12% points for the rest (/æ/= 12.37), /ʌ/=13.87, /ɑː/: 13.35). No evidence of a significant advantage for one of the training groups was shown in vowel identification at post-test, except for the fact that the ID training group identified /e/ significantly better than the ART training group.

Results indicated that training was effective in improving vowel identification in both AV and A conditions, which suggests that trainees were able to successfully integrate information contained in the (trained) auditory and visual modalities, and the training groups were not significantly different in any of these two test conditions. Importantly, trainees were significantly better in the AV condition than in the A condition after training. Considering that both ID and ART training were audiovisual types of high-variability training, it could be argued that both types of training are likely to be effective in making learners more sensitive to audiovisual cues.

When we explored the generalization of phonetic learning across words and talkers, we found that, on the one hand, ID and ART training were overall effective in improving vowel identification of trained words as well as new (untrained) words and, on the other hand, both types of training generalized to untrained talkers. Despite the successful transfer of training to the perception of the same vowel in new words and talkers, some differences between the two groups can be outlined with respect to the generalization results:

(1) The ID training group (improvement of 18%) outperformed the ART group (improvement of 9%) in the identification of trained words at post-test. After ID training, learners identified trained words more accurately than new words, whereas ART trainees performed similarly in the two word type conditions. This could be interpreted as trainees being more likely to remember words previously included in identification tasks used for training (ID training) than words that they had been previously asked to imitate during the articulatory training sessions (ART training).

(2) The ID training group performed significantly better overall on the words presented by trained talkers (familiar) than on untrained (unfamiliar) talkers, despite the large improvement obtained in vowel identification regardless of the talker condition.

This section has also presented the effects of training on error dispersion, the distribution of error responses over vowel stimuli after each type of training. After ID and ART training, learners reduced error dispersion per vowel category besides higher accuracy in vowel identification. Both types of training led to a significant decrease in the number of incorrect response categories used in vowel identification, which is a sign of acquisition, indicating that, after 10 training sessions, learners are beginning to disregard some of the variation and be more consistent in their choices (Iverson & Evans, 2009; (Bundgaard-

Nielsen et al., 2011). Some differences between ID and ART training groups can be outlined with respect to the degree of error dispersion in post-test vowel identification:

(1) Whereas the ID group obtained lower error dispersion for /iː/, the ART group showed higher error dispersion at post-test for the same vowel.

(2) For /eæuː/, the ID group showed significantly greater decrease in error dispersion than the ART group.

(3) To sum up, the ID training group experienced a significantly higher decrease than the ART group in the number of mistaken "similar" categories assigned to certain vowel stimuli at post-test.

These results raise a question about the higher effectiveness of ID training for reducing the kind of confusions appearing in vowel identification, by helping learners distinguish ambiguous vowel sets and making them consider less possible response categories in vowel identification experiments.

4.2. Identification of synthesized vowels (ID2 – task2)

The **Sub-RQ1.2** examines to what extent learners will improve in the identification of synthesized L2 vowels from pre-test to post-test. There were two measures of accuracy in the perception of synthesized vowels and over-reliance on duration: mean percent *correct identification* of synthesized vowels and *duration effect score* (DES). This section explores the effects of training on correct identification of synthesized vowels (**RQ1.2.1**) and duration effect scores (**RQ1.2.2**) and how these two measures changed as result of training and type of training from pre-test to post-test (**RQ2.2**).

4.2.1. Comparability of subject groups at pre-test

The ID training (*M*= 71.35, *SD*= 9.78), ART training (*M*= 69.40, *SD*= 18.61), and control groups (*M*= 65.73, *SD*= 9.37) obtained similar pre-test perceptual identification accuracy scores (see Table 4.16). A one-way ANOVA confirmed that the three groups did not significantly differ from one another at pre-test in vowel identification accuracy ($F(2, 81)=1.13$, $p= .327$) when they identified duration-manipulated (synthesized) vowels (mean *percent correct identification* averaged across the 7 steps of each duration continuum), suggesting that the three subject groups were indeed comparable in terms of their identification skills at pre-test (see Figure 4.15): that is, duration differences, and not spectral differences, generally determined the listeners' responses in the identification of vowels with manipulated duration (80, 116.7, 153.3, 190, 226.7, 263.3, 300 ms) at pre-test.

The ID training, ART training and control groups showed a high degree of reliance on duration in the perception of /iːɪæʌʌɑːʊuː/) presented in a two-alternative forced choice task (C1 /iː/-/ɪ/, C2 /æ/-/ʌ/, C3/ʌ/-/ɑː/ and C4 /ʊ/-/uː/). That explains the fact that the three groups performed at chance level, or near-chance level, at pre-test.

As expected, the NS control group (*M*= 92.90, *SD*= 2.50) performed at ceiling in the identification of synthesized vowels (Table 4.16), as reliance on spectral cues determined their responses in the identification of vowels with manipulated duration. A one-way ANOVA showed that the effect of *subject group* (ID, ART, NNS control, NS control) on pre-test vowel identification scores was significant ($F(3, 91)=9.27$, $p< .001$). A Tukey posthoc test confirmed that the differences between the NS group and each of the NNS groups (ID, ART, control) were statistically significant at $p< .001$ for duration-manipulated /iː/-/ɪ/ and /ʊ/-/uː/, and significant at $p< .05$ for /ʌ/-/ɑː/. However, differences between NS and NNS groups were non-significant for /æ/-/ʌ/ (see Table 4.17). At pre-test, lines representing the /æ/ and /ʌ/ responses showed a very flat response pattern, meaning that vowel quality, and not duration, determined the listeners' choice within a forced-choice block containing two lax vowels.

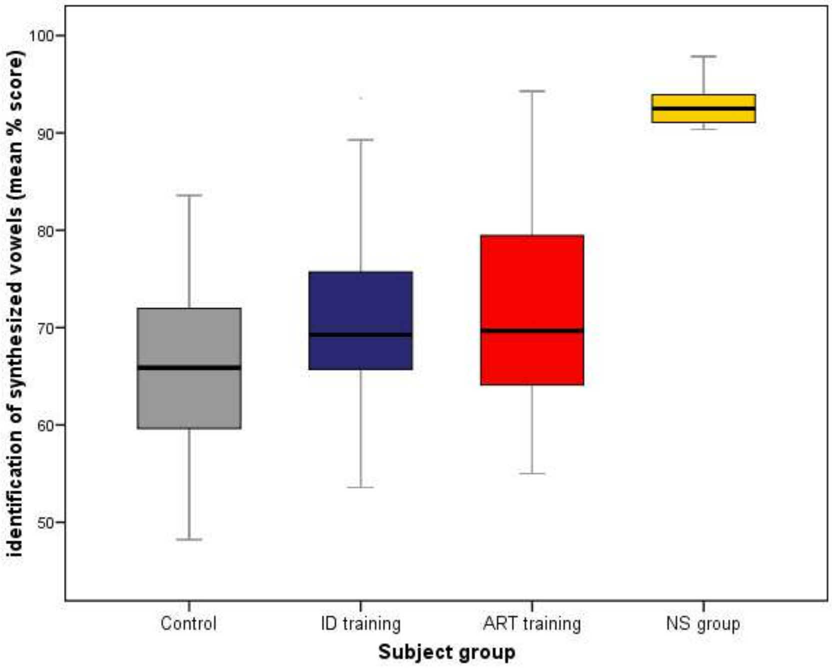


Figure 4.15. Average percent correct idenfication for synthesized English vowels obtained by ID, ART, NNS control and NS control groups at pre-test.

Identification of synthesized vowels at pre-test		Subject groups			
		ID training	**ART training**	**NNS control**	**NS**
Mean vowel identification		71.35 (9.78)	69.40 (18.61)	65.73 (9.37)	92.90 (2.45)
C1	/iː/	69.35 (19.49)	69.31 (24.30)	44.53 (28.47)	98.93 (2.12)
	/ɪ/	68.66 (20.30)	68.94 (26.76)	55.00 (28.08)	99.28 (2.02)
C2	/æ/	93.68 (11.42)	88.85 (20.51)	91.00 (22.46)	1.32 (0.47)
	/ʌ/	94.03 (11.46)	87.56 (20.65)	90.00 (21.71)	99.64 (1.01)
C3	/ʌ/	65.63 (12.71)	58.52 (20.02)	66.71 (11.77)	71.42 (12.31)
	/ɑː/	57.49 (13.68)	57.23 (20.40)	55.00 (14.50)	77.50 (15.38)
C4	/ʊ/	61.47 (19.19)	59.26 (27.42)	58.28 (21.42)	98.93 (1.48)
	/uː/	60.52 (23.04)	57.51 (26.51)	54.43 (22.08)	98.21 (2.12)

Table 4.16. Percent correct identification scores and standard deviations for the individual synthesized vowels obtained by ID training, ART training, NNS control and NS baseline groups at pre-test.

Identification of synthesized vowels at pre-test		Multiple-comparisons of subject groups (Posthoc)
Mean identification		$F(3, 91)=9.27$, $p<.001$*
C1	/iː/	$F(3, 91)=7.03$, $p< .001$*
	/ɪ/	$F(3, 91)=9.27$, $p< .001$*
C2	/æ/	$F(3, 91)=6.63$, $p< .001$*
	/ʌ/	$F(3, 91)=1.43$, $p= .425$
C3	/ʌ/	$F(3, 91)=2.25$, $p= .239$
	/ɑː/	$F(3, 91)=3.91$, $p< .05$*
C4	/ʊ/	$F(3, 91)=7.67$, $p< .001$*
	/uː/	$F(3, 91)=7.62$, $p< .001$*

Table 4.17. Results of multiple comparisons for pre-test percent correct identification of synthesized vowels obtained by ID training, ART training, control and NS groups.

4.2.2. Effects of training and type of training on mean identification of synthesized vowels

The research question RQ1.2.1examined whether audiovisual HVPT resulted in a significant improvement in the identification of English vowels with manipulated duration. At pre-test, the listeners relied exclusively on the duration cues to identify the vowels /iː ɪ æ ʌ ɑː ʊ uː/ presented in a two-alternative forced choice task (C1 /iː/-/ɪ/, C2 /æ/-/ʌ/, C3/ʌ/-/ɑː/ and C4 /ʊ/-/uː/). In terms of accuracy, the participants' choice was coded as either correct (one point) or incorrect (zero) for each vowel stimulus from the continua. Mean *percent correct identification* averaged across the 7 steps of each continuum would serve as a measure of cue weighing or the size of the effect of training on the manipulation of duration, and therefore results would indicate the degree of reliance on duration in the perception of the duration-manipulated /iː ɪ æ ʌ ɑː ʊ uː/. If the training was effective, the training groups should reveal a significant increase in the vowel identification scores at post-test, meaning lower effect of the manipulation of the duration, while the control group should not exhibit such improvement. RQ2.2 aimed to compare the effectiveness of the two training methods in improving accuracy in vowel identification and reducing the effect of the manipulation of duration.

It was hypothesized that if the listeners relied exclusively on the spectral cues, they should identify the tense and lax tokens at duration steps 1 (80 ms) and 7 (300ms) similarly. If they over-relied on duration cues, their % identification scores on lax vowels (the durationally ambiguous "hid", "had" and "hud") should decline along the continua and reach almost 0% at steps 6 and 7. Similarly, they would hardly identify the durationally ambiguous "he'd", "hard" and "who'd" at duration steps 1 and 2 as tense vowels. Compared with a NS group who responded exclusively to the spectral cues, the majority of trainees failed to show native-like perceptual patterns for the majority of the vowel contrasts at pre-test, except for /æ/and /ʌ/ when the two lax vowels were offered as responses (C2).

The descriptive statistics for pre-test and post-test identification accuracy and improvement scores (gains) for each group are shown in Table 4.18. As shown in Figure 4.16, the training groups increased in mean accuracy (improvement of 10%) and the standard deviations decreased, whereas the control group remained almost unchanged from pre-test to post-test.

A paired-samples t-test was run on the pre- and post-test identification scores of the control group to determine whether there was a statistically significant difference in the identification of synthesized vowels between the two testing times. Results indicated that post-test scores were not significantly higher (gains: M=4.55, SD= 8.72) for the group who had not received any treatment ($t(19)= -2.33, p> .05$).

A two-way ANOVA with mean percent correct vowel identification as the dependent variable, *effect of training* (T1 vs. T2) as the within-subjects factor, and *type of* *training* (ID vs. ART) as the between-subjects factor, was conducted to determine if the training had been effective in improving the identification of English vowels with manipulated duration (RQ1.2.1a) and if differences between testing times in vowel identification were due to the type of training the trainees had undergone (RQ2.2). There was a statistically significant effect of *training* on the identification of synthesized vowels ($F(1, 62)= 25.54$, $p< .001$, $\eta^2= .292$; see Figure 4.16). However, the effect of *type of training* did not reach statistical significance ($F(1, 62)= 0.20$, $p= .653$, $\eta^2= .003$). There was not a significant interaction between *training* and *type of training* ($F(1, 62)= 0.67$, $p= .415$, $\eta^2= .011$).

These results confirmed that ID and ART training groups identified synthesized vowels significantly better after training (synthesized $M_{T1}= 69.92$, $M_{T2}= 79.61$; gains= 10%). Up to this point, the type of training made no significant appreciable difference in the results relative to post-test identification of synthesized vowels. ID and ART types of training were similarly effective in improving the identification of English vowels even when the duration cue was ambiguous and could not be relied upon for L2 vowel recognition.

In the following subsections we examine how variables such as *vowel* and *manipulation of duration* might have affected the two dependent measures, *correct identification* and *duration effect score*, Subsequent ANOVAs included within-subject factors such as ***vowel*** (/iːɪæʌɑːʊuː/) and *manipulation of duration* (80, 116.7, 153.3, 190, 226.7, 263.3, 300 ms).

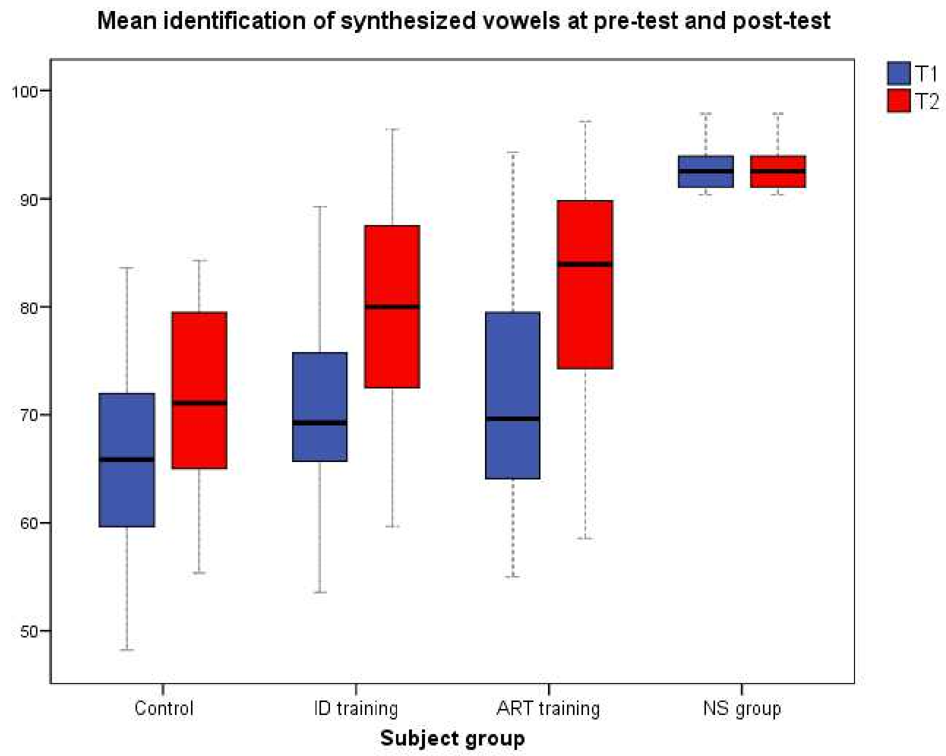


Figure 4.16. Average percent correct identification of synthesized vowels for ID training, ART training, NNS control and NS control groups at pre-test and post-test.

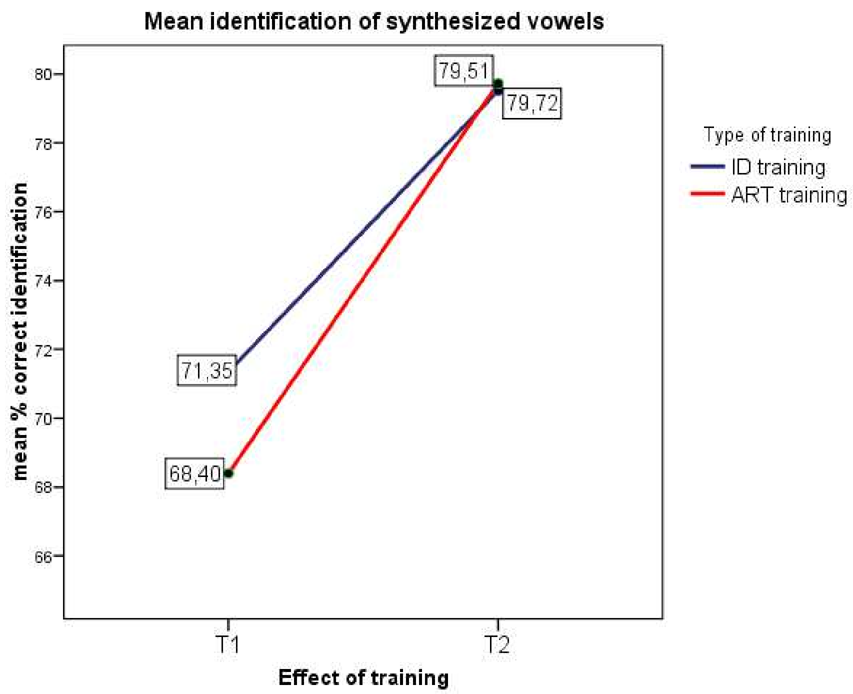


Figure 4.17. Average percent correct identification of synthesized vowels for ID training and ART training groups at pre-test and post-test.

	Vowel ID	Experimental group (*N*=64) ID training (*N*=32) T1	T2	gains	ART training (*N*=32) T1	T2	gains	Control group (*N*=20) T1	T2	gains	NS (*N*=10)
	synthesized vowels	70.94 (8.59)	79.18 (9.60)	8.24	68.91 (18.53)	80.04 (17.91)	12.88	65.73 (9.37)	70.28 (10.40)	4.55 (8.83)	92.90 (2.45)
	Tense vowels	62.45 (12.95)	72.81 (13.33)	10.36 (12.94)	61.35 (19.84)	74.83 (19.33)	13.49 (20.69)	59.95 (12.81)	61.05 (14.73)	6.09 (11.42)	91.55 (5.61)
	Lax vowels	76.69 (8.69)	83.53 (8.75)	6.84 (9.49)	72.63 (18.56)	82.65 (17.82)	10.03 (19.62)	72.20 (11.38)	75.83 (9.60)	3.63 (10.12)	93.71 (2.70)
C1	/iː/	68.39 (18.99)	70.80 (2.,36)	2.41 (30.64)	70.27 (24.51)	80.45 (27.97)	10.18 (35.04)	55.43 (28.47)	65.86 (30.09)	10.42 (34.64)	98.93 (2.12)
	/ɪ/	67.86 (20.08)	72.59 (29.45)	4.73 (33.91)	69.73 (26.70)	79.02 (28.47)	9.28 (33.18)	55.00 (28.08)	63.71 (28.36)	8.71 (31.00)	99.28 (2.02)
C2	/æ/	93.48 (11.55)	93.36 (9.40)	1.87 (9.78)	89.20 (29.28)	94.37 (18.37)	5.18 (21.90)	91.00 (22.46)	96.00 (9.96)	5.00 (20.66)	99.28 (1.32)
	/ʌ/	93.84 (11.59)	96.43 (8.80)	2.59 (7.45)	87.95 (29.43)	93.39 (18.28)	5.45 (22.06)	90.00 (21.71)	94.57 (9.94)	4.57 (19.41)	99.64 (1.01)
C3	/ʌ/	65.89 (12.82)	72.41 (16.06)	6.52 (16.46)	58.48 (19.70)	69.28 (19.87)	10.80 (20.75)	66.71 (11.77)	↓60.86 (12.27)	-5.86 (10.55)	71.43 (12.31)
	/ɑː/	56.96 (13.56)	67.68 (18.79)	10.71 (16.97)	57.77 (20.29)	65.53 (23.23)	7.77 (22.75)	55.00 (14.50)	55.71 (14.28)	7.77 (22.75)	77.50 (15.38)
C4	/ʊ/	60.80 (19.05)	79.91 (19.87)	19.11 (25.68)	59.64 (27.06)	78.30 (25.72)	18.39 (29.28)	58.28 (21.43)	64.00 (25.21)	5.71 (28.56)	98.93 (2.12)
	/uː/	59.91 (23.14)	78.30 (18.34)	18.66 (32.03)	58.21 (26.38)	80.00 (24.05)	21.78 (28.50)	54.43 (22.08)	61.57 (23.11)	7.14 (25.78)	98.21 (2.12)

Table 4.18. Pre-test, post-test and gain scores for the individual synthesized vowels obtained by ID training, ART training, NNS control and NS baseline groups.

Summary

As the results of the identification of synthesized vowels (vowels with manipulated duration) (IDII-task2) confirmed, HVPT significantly improved trainees' consistency in labeling the English synthesized vowels (T1: 69.92, T2: 79.61; p<.05). There were no significant differences between ID and ART training groups concerning gains in mean percent identification averages across all duration-manipulated vowels.

4.2.3. Effects of training and type of training on the identification of different vowels

Results for accuracy in the identification of synthesized vowels (/iː-ɪ; æ-ʌ; ʌ-ɑː; ʊ-uː/) at pre-test and post-test show that trainees improved on the identification of most of the vowels (see Table 4.19). Figures 4.18 and 4.19 show the mean % correct identifiction for /iː-ɪ; ʌ-ɑː; ʊ-uː/ produced after ID and ART training, respectively. /æ-ʌ/ were excluded from these figures as identification rates did not vary significantly from pre-test to post-test.

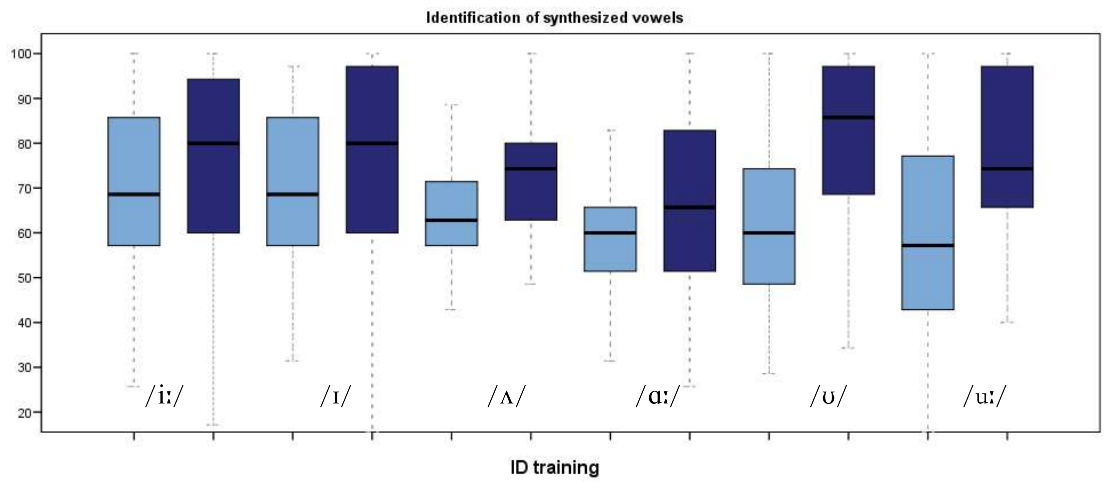


Figure 4.18 Boxplots of identification accuracy across all subjects in the ID training group, presented separately for each synthesized vowel at pre- (light blue) and post-test (dark blue).

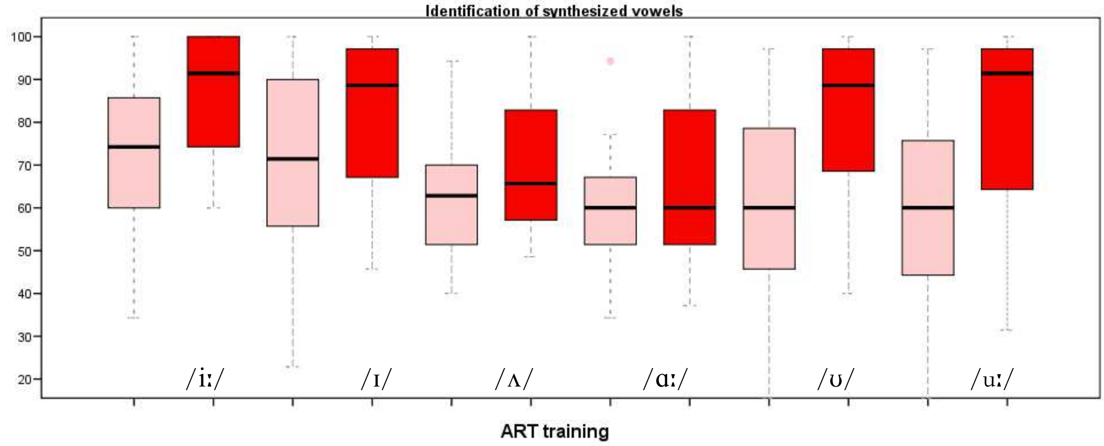


Figure 4.19. Boxplots of identification accuracy across all subjects in the ART training group, presented separately for each synthesized vowel at pre- (light red) and post-test (dark red).

The mean identification scores of the 8 synthesized vowels were submitted to a three-way repeated-measures ANOVA with *effect of training* (T1 vs. T2) and *vowel (*8) as within-subjects factors, and *type of training* (ID vs. ART) as between-subjects factor, to determine if the training had been equally effective in improving the identification accuracy of different synthesized vowels (RQ1.2.1) and whether differences could be due to type of training (RQ2.2).

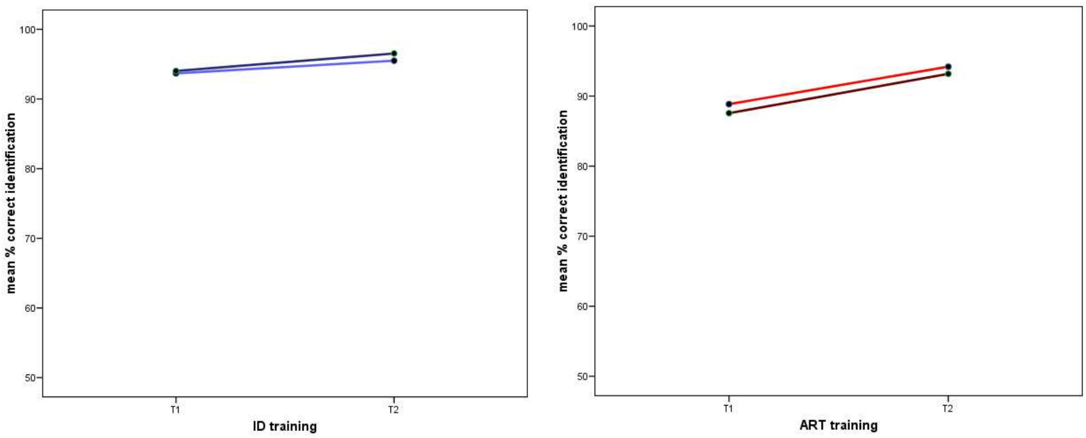
The results revealed that there was a significant *effect of training* ($F(1, 62)=25.54$, $p<.001$, $\eta^2= .292$), a significant effect of *vowel* ($F(7, 434)=56.35$, $p<.001$, $\eta^2= .476$), and a significant interaction between *training* and *vowel* ($F(7, 434)=5.41$, $p<.001$, $\eta^2= .080$). The main effect of *type of training* ($F(1, 62)=0.20$, $p=.635$, $\eta^2= .003$) and the rest of interactions did not reach statistical significance: *training x type of training* ($F(1, 62)=0.67$, $p=.415$, $\eta^2= .011$)*, vowel x type of training* ($F(7, 434)=2.25$, $p=.327$, $\eta^2= .018$)*, training x vowel x type of training* ($F(7, 434)=0.41$, $p=.897$, $\eta^2= .007$)*, vowel x type of training* ($F(7, 434)=2.25$, $p=.327$, $\eta^2= .018$). The significant two-way interaction suggests that post-test identification scores were not significantly higher than pre-test scores for all vowels. No significant differences between ID and ART training groups concerning vowel identification and reliance on duration were observed at this point.

In order to explore this significant interaction between *training* and *vowel*, subsequent paired-samples t-tests were conducted on pre-test and post-test identification scores of each vowel (Table 4.19). These revealed that identification scores at post-test significantly differed from those obtained at pre-test for 4 out of 8 synthesized vowels /ʌɑːʊuː/ ($p<$ .05). This, in turn, confirmed that identification accuracy gains were significant only for low and high-back vowels, for which trainees had started to respond to the spectral cues rather than duration cues after training. On the other hand, trainees apparently still over-relied at post-test on duration for vowels /iːɪæʌ/.

Identification of synthesized vowels		Paired comparisons (T1 vs. T2) ID and ART group (N= 64)	Gains and SD
C1	/iː/	t(63)=-1.53, p= .131	6.39 (33.88)
	/ɪ/	t(63)=-1.68, p= .098	7.01 (33.36)
C2	/æ/	t(63)=-1.67, p= .100	3.53 (16.90)
	/ʌ/	t(63)=-1.96, p= .054	4.02 (16.40)
C3	/ʌ/	t(63)=-3.70, , p< .001*	8.66 (18.70)
	/ɑː/	t(63)=-3.70, , p< .001*	9.24 (19.97)
C4	/ʊ/	t(63)=-5.18, , p< .001*	18.70 (28.88)
	/uː/	t(63)=-5.60, p< .001*	20.09 (28.72)

Table 4.19. Results of paired-samples t-tests on the identification scores of individual synthesized vowels, accuracy gains and standard deviations.

For lax vowels /æ/ and /ʌ/ within C2, identification rates did not vary significantly from pre-test to post-test and lines showed a very flat response pattern, not affected by vowel duration to the extent that the other vowel pairs were (see Figures 4.20-4.21). Being both lax vowels, and therefore generally perceived as relatively "short" by Catalan-Spanish learners, a closer inspection of the results results indicated that even the longest stimuli elicited a "lax" response (76-90%), meaning that vowel quality, and not duration, determined the listeners' choice within a forced-choice block containing two lax vowels (/æ/ and /ʌ/). The interest in the low front-central lax pair lies in the fact that this vowel contrast is considered problematic for these learners as there is no front-back distinction in Spanish or Catalan, both languages having one low vowel /a/. The forced-choice category ratings for /æ/- /ʌ/ are summarized in Figures 4.20.-4.21.



Figures 4.20.-4.21. Pre-test and post-test identification scores for /æ/ (blue) and /ʌ/ (red) at pre-test (light colour) and post-test (dark colour), presented separately for ID and ART training groups, with light colour representing pre-test and dark colours representing post-test.

Summary

HVPT had a positive effect on the trainees' percent correct ID of duration-modified /iː/-/ɪ/, /ʌ/-/ɑː/and /ʊ/-/uː/ continua. Identification rates obtained for /æ-ʌ/ did not vary significantly from pre-test to post-test. The effect of type of training on mean percent correct identification of vowels with manipulated duration did not reach significance. Both ID and ART training methods were similarly effective in increasing attention to spectral cues for accurate identification of L2 duration-manipulated vowels and identification accuract gains were significant only for low and high-back vowels /ʌɑːʊuː/ (p< .05). Trainees apparently still over-relied at post-test on duration for vowels /iːɪ/.

4.2.4. Effects of training and type of training on synthesized tense and lax vowels

The interest in tense-lax vowel pairs in our study (/iː-ɪ; ʌ-ɑː; ʊ-uː/) lies in the fact that non-native speakers tend to rely on temporal cues when identifying them (Flege, Bohn,

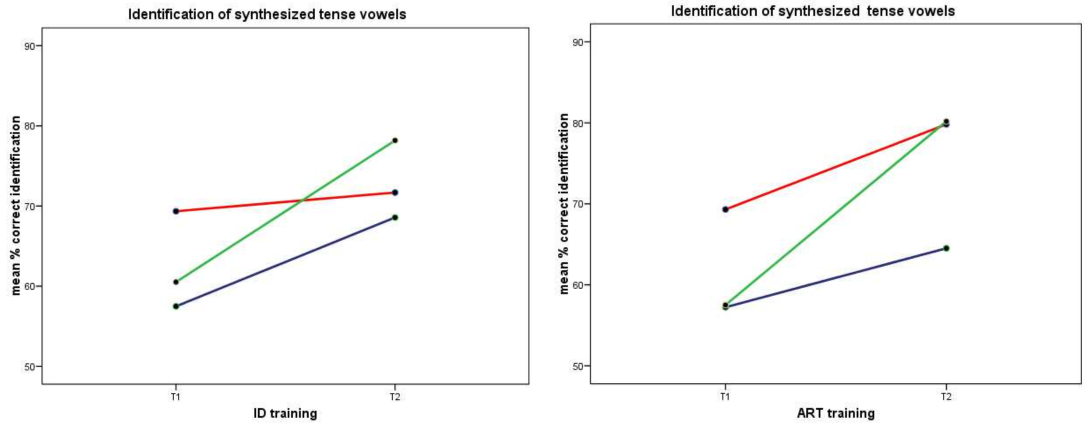
& Jang, 1997), specially Catalan learners of English, who have been found to rely mostly on duration for the /iː/-/ɪ/ contrast (Aliaga-Garcia & Mora, 2009; Cebrian, 2006). One of the research questions was (1) whether HVPT would be effective in increasing attention to spectral cues for accurate identification of duration-manipulated tense and lax vowels (RQ1.2.1). Another research question that we wanted to answer was (2) whether *vowel* was a factor explaining the effect of training on correct identification of duration-manipulated tense and lax vowels and gains (pre-test/post-test differences) (RQ1.2.1a). A third question concerning synthesized vowels was (3) whether ID and ART types of training were differently effective in increasing attention to spectral cues for accurate identification of "short" and "long" tense (RQ2.2a) and lax vowels (RQ2.2b). We hypothesized that training would make participants aware of other differences between tense and lax vowels such as vowel quality differences, resulting in a decrease in the reliance of duration.

This study suggests that HVPT led to significant improvements for both groups in the identification of both tense (/iː/,/ɑː/ and /uː/) and lax (/ɪ/, /æ/, /ʌ/ or /ʊ/) vowels (see Table 4.22), which confirms that training reduced English learners' reliance on vowel duration in English vowel perception. Although both ID and ART training groups improved the identification accuracy of tense (gains: 11.87%) and lax (gains: 8.38%) vowels at post-test, it should be noted that tense and lax vowels were identified at significantly different accuracy rates: lax vowels were identified at better accuracy rates than tense vowels both at pre-test ($t(63)=-.88$, $p< .001$) and post-test ($t(63)=-8.85$, $p< .001$). Importantly, ID and ART training groups showed significantly greater improvement for tense than for lax vowels ($t(63)=3.22$, $p= .002$).

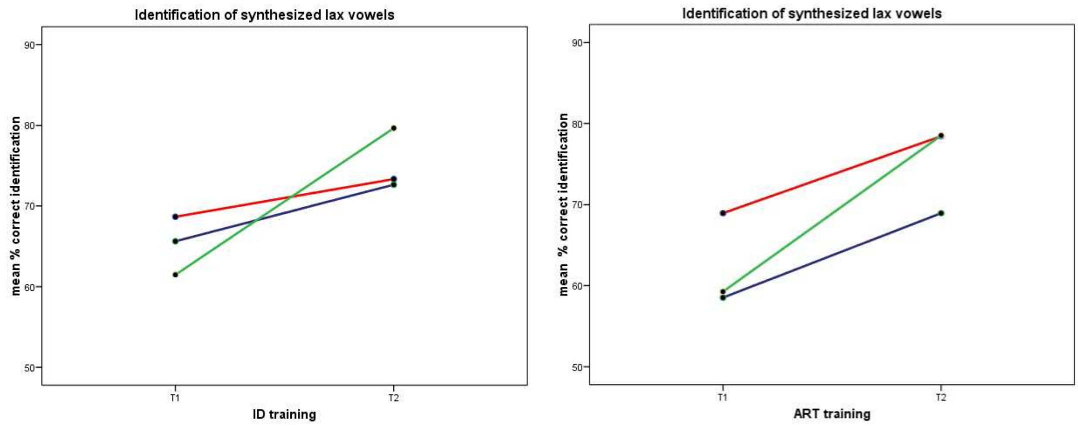
Figures 4.23-4.23 show group performance for duration-manipulated tense vowels. The descriptive data shows that synthesized tense vowels (T1: $M= 61.92$, $SD= 16.52$; T2: $M= 73.80$, $SD= 16.41$; gains of 12%) were identified at significantly different rates as a function of vowel quality (high vs. low) (see previous Table 4.17-4.18). At pre-test, the highest rate of accuracy was obtained for /iː/, then for /uː/. However, at post-test, the highest accuracy rate was obtained for /uː/, followed by /iː/. Both at pre-test and post-test, /ɑː/ was identified at a significantly lower accuracy rate than the other tense high vowels. Similarly, improvement was also observed for lax vowels (T1: M=63.79, SD= 16.12; T2: M=75.25, SD= 15.72) after training (Figures 4.24-4.25) and similar pattern of results (Figures 4.24-4.25). At pre-test, the highest rate of accuracy was obtained for /ɪ/, followed by /ʊ/ and then /ʌ/; at post-test, /ʊ/ received the highest accurscy rate, followed by /ɪ/. In both pre-test and post-test, /ʌ/ received the lowest accuracy rates despite improvement after training.

Repeated-measures ANOVAS were conducted separately on mean correct identification of tense (/iːɑːuː/) and lax (/ɪæʌʊ/) vowels, with effect of *training* (T1 vs. T2) and *vowel* (3/4) as within-subjects factors and *type of training* (ID vs. ART) as between-subjects factor (RQ1.21.; RQ2.2a and RQ2.2b; RQ1.2.2).

The ANOVAS conducted separately on mean correct identification of tense and lax vowels yielded a significant effect of *training* for both tense (F(1, 62)=30.97, p< .001, η^2= .333) and lax (F(1, 62)=25.91, p<.001, η^2= .295) vowels. The effect of *vowel* reached statistical significant for tense vowels (F(2, 124)=30.97, p<.001, η^2= .333) but was not significant for lax vowels (F(2, 124)=2.56, p=.082, η^2= .040). There was a significant *training x vowel* interaction for both tense (F(2, 124)=4.71, p<.05, η^2= .182) and lax (F(2, 124)=3.78, p<.05, η^2= .057) vowels. The *type of training* or other interactions were non-significant (p> .05). The significant two-way interactions suggested that the effect of training on the correct identification of duration-manipulated tense and lax vowels varied as a function of *vowel*. Pre-test/post-test differences in the identification of tense and lax vowels did not reach significance for all vowels. A closer inspection of the data revealed that vowel height (high vs. low) seem to play a crucial role in the effect that training had on the identification rates, without significant differences between ID and ART training groups. High synthesized vowels (/iːɪʊuː/) tend to be perceived at better accuracy rates than low vowels /ʌɑː/ regardless of the fact that the vowel is tense or lax. At post-test, low duration-manipulated vowels were identified at lower accuracy rates than high vowels (see Figures 4.22-4.25).



Figures 4.22.-4.23. Pre-test and post-test identification scores for high-front /iː/ (red), low /ɑː/ (blue) and high-back /uː/ (green) "tense" vowels, presented separately for ID and ART training groups.



Figures 4.24.-4.25. Pre-test and post-test identification scores for high-front /ɪ/ /(red), low /ʌ/(blue) and high-back /ʊ/ (green) "lax" vowels, presented separately for ID and ART training groups.

Summary

This study shows that HVPT led to significant improvements for both training groups in the identification of both tense (/iː/,/ɑː/ and /uː/) and lax (/ɪ/, /æ/, /ʌ/ or /ʊ/) vowels, confirming that training reduced English learners' reliance on vowel duration in English vowel perception. However, lax vowels were identified at better accuracy rates than tense vowels both at pre-test and post-test, and both training groups showed significantly greater improvement for tense than for lax vowels.

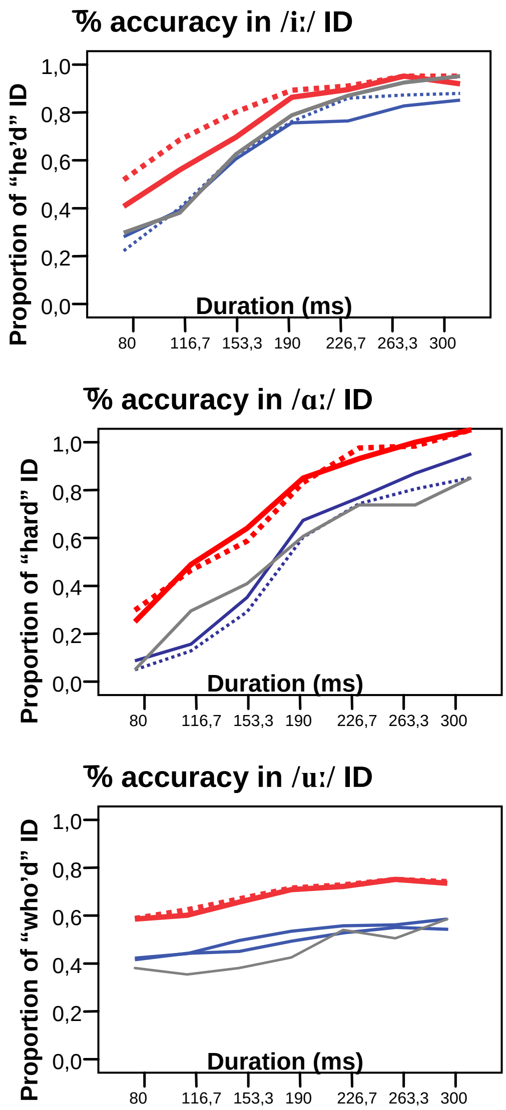
The two types of the high variability training had a significant effect on the identification of 6 out of the 8 vowels with manipulated durationThe highest improvement was observed for the high back vowels (ID: 18.58%; ART: 20.08%), followed by the high-front vowels (ID: 3.57%; ART: 9.73%), and then the low vowels (ID: 8.62%; ART: 9.28%). Both ID and ART training groups showed a substantial improvement (5.52%) in identifying the back tense vowel (Figures 4.21-4.22) and the back lax vowel (Figures 4.23-4.24), in comparison with the high-front and low vowels. After both ID and ART training, "vowel quality" shows a trend towards being incorporated as a phonetic cue in the identification of tense and lax vowels.

4.2.5. Effects of training and type of training on the manipulation of duration (or duration steps)

Figures 4.26-4.31 show the identification rates obtained for tense and lax vowels as a function of manipulation of duration (or duration steps: (S1 80, S2 116.7, S3 153.3, S4 190, S5 226.7, S6 263.3, S7 300 ms) (RQ1.2.3). These are the percentage that each of the vowels were identified correctly as a function of the duration step or manipulation of duration, when the duration cue was ambiguous for correct categorization.

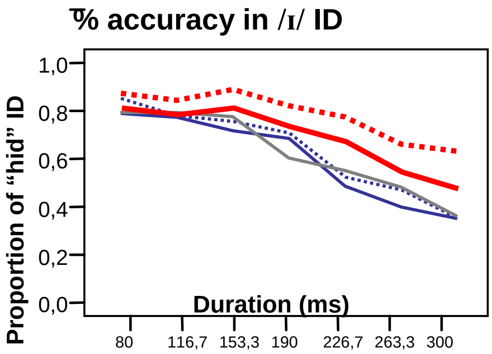
A close inspection of the data revealed that, for tense vowels, the listeners' perception changed from "lax" to "tense" response between stimuli S1-S3 at pre-test; however, at post-test they performed above chance level (>60%) after S3. For lax vowels, the listeners' choice changed from "lax" to "tense" response between S6-S7 at post-test, rather than after S4 as in the pre-test. ID and ART training groups did not differ in their post-test performance for tense and lax vowels.

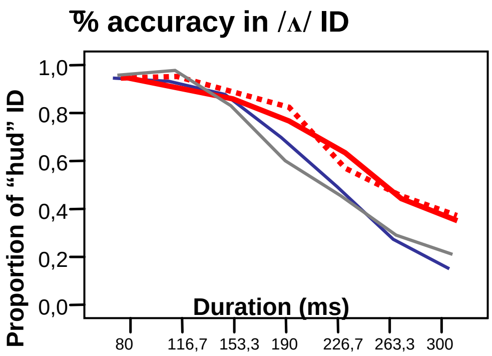
An examination of the performance with tense vowels (/iːɑːuː/) showed that training had a positive impact on the amount of "tense" responses to duration-manipulated tense vowels. Figures 4.26-4.28 show the percentage that tense vowels were identified as /iːɑːuː/ rather than /ɪʌʊ/ despite the manipulation of duration. At pre-test, the shortest stimuli (S1 80, S2 116.7 ms) failed to elicit a strong "tense" response (<50%) for /iːɑːuː/ for both ID and ART training groups. At post-test, accuracy rates were above chance level (>60%) along the seven steps of the duration continuafor the high tense vowels (/iu/) (Figure 4.26-4.28), whereas post-test accuracy rates remained below chance level (<50%) along the shorter stimuli (S1-S2) for the low tense /ɑː/, that is, when this was shorter than 153.3ms (S3) (Figure 4.28). ID and ART training methods were similarly effective in increasing attention to spectral cues for accurate identification of "short" and "long" tense vowels.

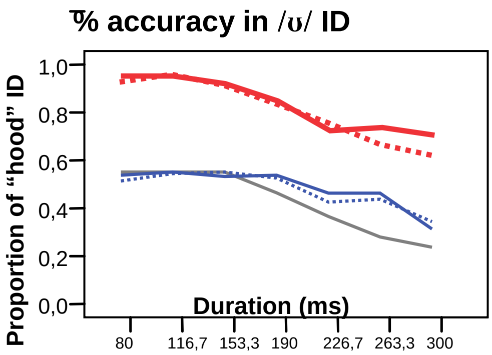


Figures 4.26.-4.28. Mean identification rates of /iː/, /ɑː/ and /uː/ plotted as a function of manipulation of duration for ID training (blue), ART (red) training and control (grey) groups, at pre-test (broken line) and post-test (solid line).

Similarly, the overall performance also improved for lax vowels (/ɪʊʌ/), suggesting an effect of training on modulating listeners' perception along the seven duration steps. Figures 4.29-4.31 display the mean identification rates of each lax vowel plotted as a function of duration steps. Similar to tense vowels, an examination of the performance with lax vowels (/ɪʌʊ/) showed that training had a positive impact on the amount of "lax" responses to duration-manipulated lax vowels. The longest stimuli (S7 300 ms) failed to elicit strong /ɪ/ or /ʊ/ responses (<50%) at pre-test; however, the majority of identification shifted to /ɪ/ or /ʊ/ even for the longest stimuli (S7) at post-test (Figures 4.30 and 4.31). By contrast, for /ʌ/ (within C3 /ʌ/-/ɑ/), pre-test accuracy rates were below chance level (<50%) after S4 (190 ms). At post-test, below-chance level accuracy (<50%) for /ʌ/ occured along S6-S7, that is, when the lax vowel was shorter than 226.7ms. ID and ART training methods were similarly effective in increasing attention to spectral cues for accurate identification of "short" and "long" lax vowels.

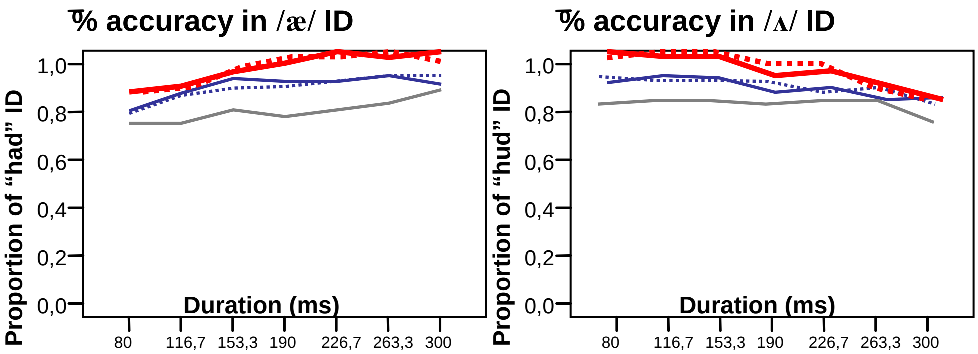






Figures 4.29.-4.31. Mean identification rates of /ɪ/, /ʌ/ and /ʊ/ plotted as a function of manipulation of duration for ID training (blue), ART (red) training and control (grey) groups, at pre-test (broken line) and post-test (solid line).

Lax vowels /æ/ and /ʌ/ within C2 (Figures 4.32-4.33) had been excluded from the analysis of variance as identification rates at each of end of the continuum (S1 and S7) did not vary significantly. Identification rates did not change significantly as a function of manipulation of duration and even the longest stimuli elicited a strong "lax" response (76-90%), meaning that vowel quality, and not duration, determined the Catalan-Spanish listeners' choice within a forced-choice block containing two lax vowels (C2: /æ/-/ʌ/). The forced-choice category ratings are summarized in Figures 4.32-4.33, where lines representing the /æ/ and /ʌ/ responses show a very flat response pattern, not affected by vowel duration.



Figures 4.32.-4.33. Mean identification rates of /æ/ and /ʌ/ plotted as a function of manipulation of duration for ID training (blue), ART (red) training and control (grey) groups, at pre-test (broken line) and post-test (solid line).

In order to the further analize the effect of training (RQ1.2.1b) and type of training (RQ2.2c) on vowel perception by manipulation of duration at pre- and post-test, a series of three-way ANOVAs were conducted separately on the identification accuracy scores (mean percent correct identification) of each vowel, with *effect of training* (T1 vs. T2) and *manipulation of duration (7 steps)* as within-subjects factors, and *type of training* (ID vs. ART) as between-subjects factor (see Table 4.20).

The ANOVAS yielded a significant effect of *training* for /iːʌɑː/ and a significant effect of *manipulation of duration* for /ɪʌɑː/. Moreover, there was a significant interaction between *training* and *manipulation of duration* for /iːɪʌɑː/, suggesting that the manipulation of duration (duration steps) affect the effect of training on the identification of high-front and low vowels (Table 4.20). In other words, the effect of training was not the same at all duration steps of the continua for /iːɪʌɑː/. The effect of type of training and other interactions did not reach significance for any of these vowels, which confirmed that the effect of training on manipulation of duration did not differ as a function of the type of training received.

There was a significant effect of *training* and a significant effect of *manipulation of duration* for high-back vowels /ʊuː/ (Figure 4.25). The effect of *type of training* did not reach significance. The *training x manipulation of duration* was not significant either. These results indicate, on the one hand, that the training was significantly effective in improving the identification rates of the tense and the lax back vowels when the duration cue was ambiguous and learners had to rely mainly on "quality" (rather than "duration") for correct identification. On the other hand, results confirmed that the manipulation of duration generally affected the performance of two training groups when they heard different instances of "long" and "short" /ʊuː/ at pre-test and post-test. In other words, this means that identification rates changed significantly as a function of *manipulation of duration*.

Vowel	Effect of *training*	Effect of *manipulation of duration*	Two-way interaction *training x manipulation of duration*
iː	$F(6, 372)=69.62$, $p<.001$, $\eta^2= .529$	-	$F(6, 372)=4.63$, $p<.001$*, $\eta^2= .070$
ɪ	-	$F(6, 372)=41.49$, $p<.001$, $\eta^2= .401$	$F(6, 372)=5.72$, $p<.001$*, $\eta^2= .092$
ʌ	$F(1, 62)=41.49$, $p<.001$, $\eta^2= .182$	$F(6, 372)=166.64$, $p<.001$, $\eta^2= .729$	$F(6, 372)=3.78$, $p=.001$*, $\eta2=.057$
ɑː	$F(1, 62)=13.43$, $p= .001$, $\eta^2= .178$	$F(6, 372)=184.82$, $p<.001$, $\eta^2= .749$	$F(6, 372)=5.72$, $p<.001$*, $\eta^2= .084$
ʊ	$F(1, 62)=26.45$, $p<.001$*, $\eta^2= .299$	$F(6, 372)=19.40$, $p<.001$*, $\eta^2= .238$	-
uː	$F(1, 62)=21.28$, $p<.001$*, $\eta^2= .335$	$F(6, 372)=26.82$, $p<.001$*, $\eta^2= .302$	-

Table 4.20. Results of three-way ANOVAS conducted on mean percent correct identification of each synthesized vowel., with *training* and *manipulation of duration* as within-subjects factor and *type of training* as between-subjects factors.

A series of paired-samples t-tests were run on the pre-test and post-test mean correct identification scores of each vowel at each duration step in order to find out where significant gains were after training (Tables 4.21-4.22) when the duration cue was ambiguous and learners had to rely mainly on "quality" for correct identification.

For duration-manipulated/iː/, ID and ART training groups were found to perform poorly (<50% accuracy) at pre-test at the shortest steps of the continuum, S1 and S2. After training, the t-tests revealed significant gains in correct identification for S1 and S2 (p< .05), meaning that after training participants performed significantly better at the shortest duration steps besides overall improvement from S1 to S7 (Table 4.21). Similarly, trainees over-relied on duration for /ɑː/ and performed rather poorly identifying this vowel at the shortest steps of the duration continua (S1-S3) at pre-test, which meant that when the vowel was shorter than 153.3 ms, the majority of identification shifted to /ʌ/ rather than /ɑː/. At post-test, the t-tests yielded significant gains at duration steps S1-S3 (p< .05), which meant that training had a significant impact at the "short" steps and made identification /ɑː/ rise above 50% by helping learners attend to quality cues rather than duration when the tense was was shortened (Table 4.21). For /uː/, poor performance was observed at pre-test for S1 and S2, and pre-test identification rates varied as a function of the manipulation of duration. At post-test, t-tests revealed significant gains for all duration steps (p< .05), and even the shortest stimuli (S1) elicited a strong /uː/ response (>60% accuracy). As regards tense vowels, the statistical results showed significant pre-test/post-test improvement in correct identification of /iːɑːuː/ when the vowels were shortened (80-153.3ms) (Table 4.21). There were no salient differences between ID and ART with regard to gains in the identification of tense vowels at the shortest duration steps or overall improvement. Significant improvement obtained at S1, S2 and S3 confirmed that training was effective in improving the performance of both training groups when they heard different instances of "long" and "short" tense vowels. In other words, even though the manipulation of duration affected the identification of "tense" vowels, a trend was observed towards a major sensitivity to vowel quality differences among all "tense" vowels.

For duration-manipualted /ɪ/, ID and ART training performed rather poorly (<50% accuracy) at pre-test at the longest step of the continuum (S7).After training, both groups performed above 50% at all steps of the continuum and t-tests revealed significant gains for S5-S7 ($p < .05$; Table 4.22), that is, when /ɪ/ was longer between 226.7 and 300ms. For /ʌ/ (within C3), the training groups showed very poor performance starting after S5 but training had a positive effect at S5, making identification rate rise above 50% from S5 to S7. Even though post-test identification remained poor at the longest steps S6-S7 of the vowel continua, t-tests revealed significant gains for S4-S7 ($p < .05$) (Table 4.22), when /ʌ/ was longer than 190ms. The statistical results yielded for /ʊ/ significant gains at almost all duration steps (S1, S2, S3, S4, S5 and S7) ($p < .05$; Table 4.22). Training had an overall significant impact on the identification rates of duration-manipulated /ʊ/ and a clear visible impact on S7. At pre-test, the lengthened back lax vowel /ʊ/ failed to be perceived as /ʊ/ at S7 (<50% accuracy); at post-test both ID and ART groups reached above 60% correct identification at all steps of the continua. Subsequent ANOVAs included within-subject factors such as ***vowel*** (/iːɪæʌɑːʊuː/) and *manipulation of duration* (80, 116.7, 153.3, 190, 226.7, 263.3, 300 ms).

Manipulation of duration	Paired comparisons (T1 vs. T2) ID and ART group (*N*= 64)					
	Tense					
	/iː/		/ɑː/		/uː/	
	T1 vs. T2	gains and SD	T1 vs. T2	gains and SD	T1 vs. T2	gains and SD
S1 80ms	t(63)=-2.93, *p*= < .05*	0.72 (1.96)	t(63)= -2.77, *p*< .05*	0.58 (1.67)	t(63)= -2.19, *p*< .05*	0.69 (2.51)
S2 116.7ms	t(63)=-2.86, *p*= < .05*	0.77 (2.14)	t(63)= -4.86, *p*< .001*	1.11 (1.83)	t(63)= -2.75, *p*< .05*	0.77 (2.23)
S3 153.3ms	t(63)=-1.29, *p*= .202	0.30 (1.84)	t(63)=-3.21, *p*< .05*	0.80 (1.99)	t(63)=-3.77, < .001*	1.02 (2.16)
S4 190ms	t(63)=-0.77 *p*= .442	0.20 (2.10)	t(63)=-1.61, *p*= .112	0.28 (1.40)	t(63)=-4-89, < .001*	1.14 (1.87)
S5 226.7ms	t(63)=-0.31, , *p*=.756	0.08 (2.00)	t(63)=-1.62, *p*=.110	0.27 (1.31)	t(63)=-4.63, < .001*	1.08 (1.72)
S6 263.3ms	t(63)=-3.70, , *p*= .532	0.14 (1.79)	t(63)=-0.83, *p*=.410	0.14 (1.35)	t(63)=-5.82, < .001*	1.25 (1.91)
S7 300ms	t(63)=-5.18, *p*= 1.000	0.00 (1.79)	t(63)=-0.50, *p*= .621	0.06 (1.01)	t(63)=-5.27, < .001*	1.09 (1.66)

Table 4.21. Results of paired-samples t-tests on mean percent correct identification by manipulation of duration obtained for each synthesized tense vowel.

Manipulation of duration	Paired comparisons (T1 vs. T2) ID and ART group (N= 64)					
	Tense					
	/ɪ/		/ʌ/		/ʊ/	
	T1 vs. T2	gains and SD	T1 vs. T2	gains and SD	T1 vs. T2	gains and SD
S1 80ms	t(63)= 0.54, p= .592	0.12 (1.86)	t(63)= -1.2, p= .236	0.14 (0.94)	t(63)= -5.31, p< .001*	1.31 (2.04)
S2 116.7ms	t(63)= 0.00, p= 1.000	0.00 (1.93)	t(63)= -0.92, p= .363	0.12 (1.09)	t(63)= -5.28, p< .001*	1.25 (1.89)
S3 153.3ms	t(63)=-1.32, p= .191	0.31 (1.89)	t(63)=-0.63, p= .529	0.11 (1.38)	t(63)=-5.31, p< .001*	1.14 (1.72)
S4 190ms	t(63)=-0.64 p= .521	0.16 (1.94)	t(63)=-2.23 p< .05*	0.42 (1.51)	t(63)=-3.05, p< .05*	0.77 (2.01)
S5 226.7ms	t(63)=-3.10, p=.003*	0.80 (2.06)	t(63)=-2.56, p< .05*	0.59 (1.86)	t(63)=-2.76, p< .05*	0.75 (2.17)
S6 263.3ms	t(63)=-2.39, p< .05*	0.58 (1.93)	t(63)=-3.40, p< .05*	0.81 (1.91)	t(63)=-1.50, p=.138	0.47 (2.49)
S7 300ms	t(63)=-2.66, p< .05*	0.73 (2.21)	t(63)=-3.86, p< .001*	0.83 (1.71)	t(63)=-2.65, p< .05*	0.86 (2.59)

Table 4.22. Results of paired-samples t-tests on mean percent correct identification by manipulation of duration obtained for each synthesized lax vowel.

In other words, at post-test trainees responded significantly different from pre-test towards the identification of "lax" vowels /ɪʌʊ/ and the manipulation of duration. The post-test results indicated that trainees were more sensitive towards quality differences rather than duration differences in the perception of the lax vowels even at the longer duration steps.

The advantage of ID and ART training was clearly observed with respect to reducing the duration reliance on vowel perception. The results reported above show improved perception of tense and lax vowels when these are "shortened" and "lengthened", respectively. This suggests that listeners successfully reduced their over-reliance on vowel duration for English vowel perception as shown in a perception task in which no duration cue was available (manipulation of duration).

Improved perception at the "long" and "short" ends of the duration continuum seem to indicate that high-variability phonetic training helped reduce listeners' dependence on duration for English vowel perception, esp. at the shorter end (S1-S3) of the continuum for "tense" vowels and at the longer end (S5-S7) for "lax" vowels. However, it should be noticed that post-test identification accuracy remained low (<50%) at the shortest step(s) (S1) for /iː/ and S1-S2 for /ɑː/, and still remained poor at the longest step (S7) for /ɪ/ and S6-S7 for /ʊ/, which meant that subjects still seem to have some difficulty in detecting quality differences and are still somehow sensitive to the *manipulation of duration* after training.

Therefore, this study suggests that ID and ART training may both help shift non-native listeners' perceptual strategy from heavy reliance on duration cues to less heavy reliance, while placing greater reliance on spectral cues. This perpetual task with duration-manipulated vowels may have triggered the use of spectral cues for English vowel perception, and tense-lax contrast perception (/iː-ɪ; ʌ-ɑː; ʊ-uː/). However, short-term training (ten sessions) may not be able to fundamentally or permanently change listeners' perceptual weights of acoustic cues for correct vowel identification.

Summary

This subsection has presented the effects of training and the manipulation of duration on Catalan-Spanish speakers when a comprehensive vowel inventory (8 vowels) was examined. These results thus suggest that Catalan-Spanish speakers of English were over-dependent on duration cues for most of the vowel tested , except for /æ/-/ʌ/, before training. At pre-test, the amount of accuracy in the perception of synthesized vowels could be predicted by their sensitivity to the manipulation of duration, as vowel duration was their primary cue in the identification of English vowels. Interestingly, at post-test listeners improved their performance along the seven duration steps, especially those steps where they had performed poorly at pre-test (shorter steps for tense vowels, and longer steps for lax vowels), but there were no significant differences as a function of the type of training received. In summary, both ID and ART training seem to have played a crucial role in helping to shif the listeners' perceptual weight towards spectral cues (rather than duration cues) and thus modulating their dependence on vowel duration cues.

4.2.6. Effects of training and type of training on mean Duration Effect Score (DES)

RQ1.2.2 examined whether audiovisual HVPT resulted in a significant decrease in the degree of reliance on duration in the perception of the duration-manipulated vowels (/iː ɪ æ ʌ ɑː ʊ uː/) measured through *duration effect scores* (DESs). At pre-test, the listeners relied exclusively on the duration cues to identify the vowels /iː ɪ æ ʌ ɑː ʊ uː/ presented in a two-alternative forced choice task (C1 /iː/-/ɪ/, C2 /æ/-/ʌ/, C3/ʌ/-/ɑː/ and C4 /ʊ/-/uː/). RQ2.2d compared the effectiveness of the two training methods for reducing DESs at post-test.

Listeners' DES was calculated by subtracting the mean % identification scores of duration step S1 (80 ms) from S7 (300ms) in the case of tense vowels (/iː, ɑː, uː/), and those of step S7 from S1 in the case of lax vowels (/ɪ, æ, ʌ, ʊ/). For example, the DES would be 100 % correct identification if a listener identified all the S7 tokens as "he'd" (/iː/) none of the S1 tokens as "he'd" (/iː/). The larger the % difference scores between the two endpoints (i.e. the higher the DES), the more the listeners relied on the duration cue (rather than spectral cues) for contrasting the tense-lax vowel pairs. If the training was effective, the training groups should reveal a decrease of DESs at post-test, while the control should not exhibit such changes in cue-weighting for correct vowel categorization.

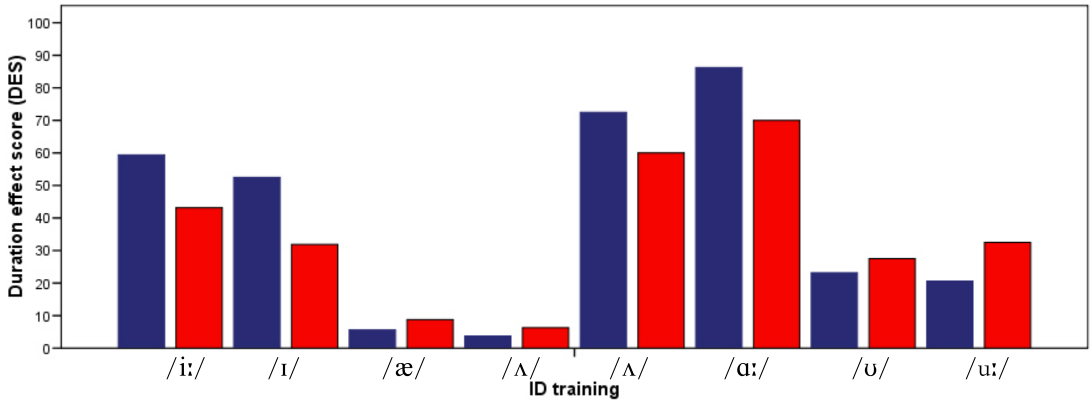
Table 4.23 shows DESs of ID and ART training groups at pre-test and post-test. A closer examination of data showed that DESs decreased after training for /iːɪʌɑː/ for the ID group, and decreased for /iːɪæʌɑː/ for the ART training (Figures 4.34.-4.35). However, DESs increased for both ID and ART groups for /ʊuː/contrast. Whereas the mean DES was 17.81% for the NS group, mean DES ranged 8%-86% as a function of vowel. Pre-test/Post-test differences in DESwere similar across training groups: DES decreased by 5.47% after ID training and 6.17% after ART training. The effect of training on DES varied as a function of the target vowel examined: DESs decreased significantly after both types training for /iːɪʌɑː/. meaning that listeners generally showed lower sensitivity to duration cues for categorizing these vowels. Differences between ID and ART were salient for /æʌ/, for which only the ART training group showed lower DES. Results seem to indicate that the ID training group still over-relied on the duration cue at post-test for categorizing/æ/-/ʌ/ and thatboth ID and ART groups over-relied on duration for categorizing the back vowels /ʊuː/.

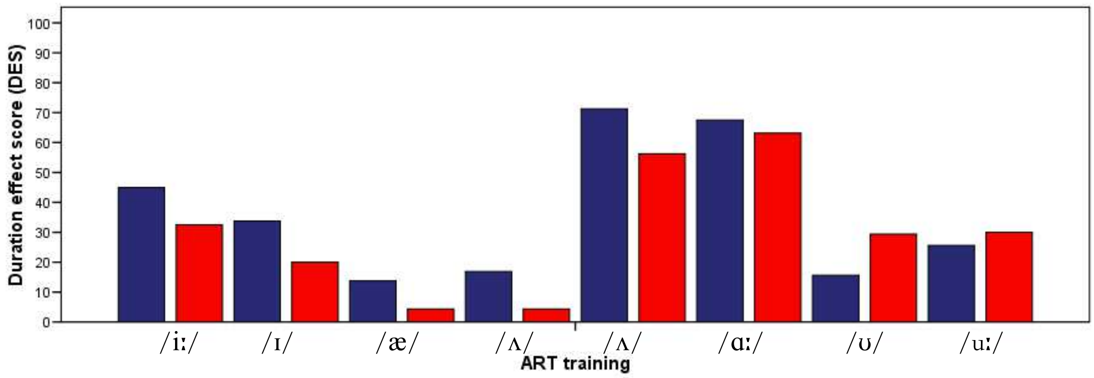
Duration Effect Score (DES)		ID training (N=32)			ART training (N=32)		
		T1	T2	gains	T1	T2	gains
C1	/iː/	59.38 (38.85)	43.13 (41.61)	16.25 (40.78)	45.00 (47.65)	32.50 (37.93)	12.50 (44.50)
	/ɪ/	52.50 (42.73)	31.88 (40.91)	20.62 (54.35)	33.75 (44.41)	20.00 (39.02)	13.75 (28.03)
C2	/æ/	5.62 (25.52)	8.75 (23.24)	-3.12 (23.34)	13.75 (21.81)	4.38 (12.83)	9.37 (18.30)
	/ʌ/	3.75 (24.59)	6.25 (20.60)	-2.5 (20.79)	16.87 (26.93)	4.38 (13.18)	12.50 (21.40)
C3	/ʌ/	72.50 (30.79)	60.00 (33.31)	12.50 (39.27)	71.25 (38.67)	56.25 (39.82)	15.00 (35.56)
	/ɑː/	86.25 (23.52)	70.00 (43.99)	16.25 (42.33)	67.50 (34.36)	63.12 (39.06)	4.37 (35.46)
C4	/ʊ/	23.12 (61.61)	27.50 (46.77)	-4.37 (84.05)	15.63 (58.31)	29.37 (39.01)	-13.75 (54.82)
	/uː/	20.63 (57.58)	32.50 (35.83)	-11.87 (69.07)	25.62 (57.30)	30.00 (38.35)	-4.37 (57.41)

Table 4.23. Pre-test and post-test duration effect scores and gains for individual synthesized vowels obtained by ID and ART training groups.

A paired-samples t-test was run on the pre- and post-test DESs of the control group to determine whether there was a statistically significant difference in the identification of synthesized vowels between the two testing times. Results indicated that for the group who had not received any treatment post-test DESs were not significantly different from pre-test for any of the synthesized vowels: /iː/, $t(19)= 0.68$, $p> .05$; /ɪ/, $t(19)= 0.55$, $p> .05$; /æ/, $t(19)= 0.82$, $p> .05$; C2/ʌ/, $t(19)= -1.16$, $p> .05$; C3/ʌ/, $t(19)= 1.02$, $p> .05$; /ɑː/, $t(19)= 2.31$, $p> .05$; /ʊ/, $t(19)= 1.86$, $p> .05$; /uː/, $t(19)= 2.64$, $p> .05$.

The mean DESs of the 8 synthesized vowels were submitted to a three-way repeated-measures ANOVA with *effect of training* (T1 vs. T2) and *vowel* (8) as within-subjects factors, and *type of training* (ID vs. ART) as between-subjects factor, to determine if the training had been equally effective in decreasing the DES of different duration-manipulated continua(/iː-ɪ; ʌ-ɑː; ʊ-uː/) (RQ1.2.2) and whether differences could be due to type of training (RQ2.2d).





Figures 4.34.-4.35. Mean duration effect scores in vowel identification at pre-test (blue) and post-test (red) obtained by ID and ART training groups.

The analysis revealed a significant main effect of *vowel* ($F(7, 434)= 56.45$, $p< .001$, $\eta^2=.477$) and the main effect of *type of training* ($F(1, 62)= 0.80$, $p= .374$, $\eta^2=.013$) was not statistically significant, which suggested that listeners from the ID and ART groups did not differ in their response to duration-manipulated vowels at post-test. There was a significant *training x vowel* interaction ($F(5, 315)= 4.10$, $p< .001$, $\eta^2=.062$). The effect of training did no reach significance and no other interactions turned out to be significant. The significant two-way interaction indicated that the effects of *training* on DES varied as a function of the vowel examined.

To further examine results, paired-samples t-tests were conducted on the pre-test and post-test DESs of each synthesized vowel for each of the training groups (Table 4.29). The results revealed a significant decrease in DES for synthesized /ɪ/ after both ID and

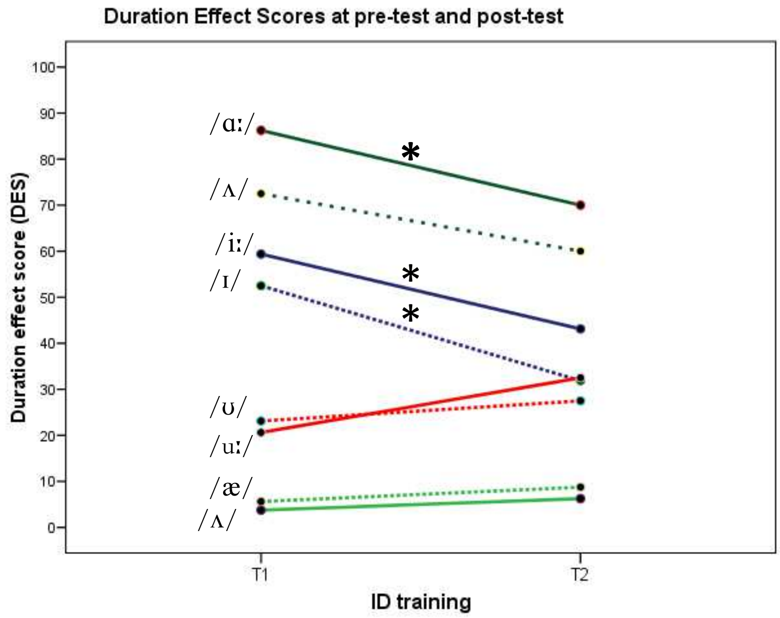
ART training. Pre-test/post-test differences in DES of /iː/ and /ɑː/ only reached significance after ID training, whereas the increase in DES of /ʌ/ was only significant after ART training.

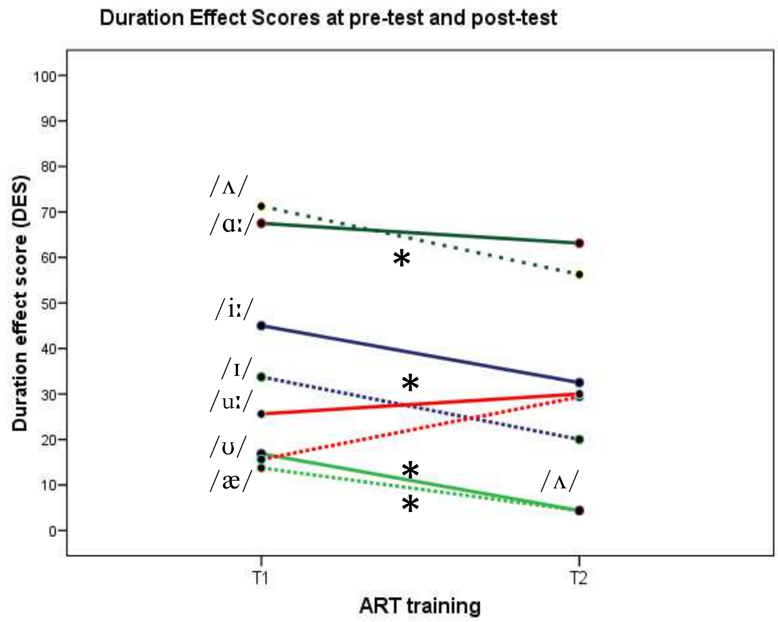
Duration effect score		Paired comparisons (T1 vs. T2) **ID training (*N*=32)**		**ART training (*N*=32)**	
		df t p-value	**diff**	**df t p-value**	**diff**
C1	/iː/	t(32)=2.25, p< .05*	16.25 (40.78)	30)=1.59, p= .122	12.50 (44.50)
	/ɪ/	t(32)=2.14, p< .05*	20.62 (54.35)	30)=2.78, p< .05*	13.75 (28.03)
C2	/æ/	t(32)=-0.76, p= .454	-3.12 (23.34)	30)=2.91, p< .05*	9.68 (12.53)
	/ʌ/	t(32)=-0.68, p= .501	-2.5 (20.79)	30)=3.32, p< .05*	12.90 (21.63)
C3	/ʌ/	t(32)=1.89, p= .068	12.50 (39.27)	30)=2.29, p< . 05*	14.84 (36.15)
	/ɑː/	t(32)=2.33, p< .05*	16.25 (42.33)	30)=0.51, p= .616	3.22 (35.44)
C4	/ʊ/	t(32)=0.34, p= .739	-4.37 (84.05)	30)=1.35, p= .186	13.55 (55.71)
	/uː/	t(32)=1.02, p= .313	-11.87 (69.07)	30)=0.37, p= .714	3.87 (58.29)

Table 4.24. Results of paired-samples t-tests on pre-test and post-test duration effect scores for individual synthesized vowels and pre-test/post-test differences.

Post-test DESs were generally lower for /iːɪʌɑː/ for both ID and ART groups (Figure 4.36), and were lower for /æʌ/ after ART training (Figure 4.36). These results suggest that training reduced the listeners' reliance onduration in English vowel categorization. Pre-test/post-test changes in DES for /iːɪæʌɑː/ are due to the impact of training (Figures 4.36-4.37).

At pre-test, the performance on "short" /iːɑː/ and "long" /ɪæʌʊ/ was rather poor, as duration-manipulated stimuli provided listeners with contradictory duration spectral cues. The fact that listeners showed lower DES for 6 of the 8 synthesized vowels (/iɑɪæʌʊ/) at post-ttest proves that, first, the two types of high-variability training facilitated the formation of robust vowel phonetic categories. Second, the results suggest that ID and ART training can shift language learners' attention towards less reliance on duration cues and more weigh on more relevant cues when the duration cue is ambiguous (e.g. "short" /iːɑː/ and "long" /ɪæʌʊ/).





Figures 4.36.-4.37. Pre-test and post-test duration effect scores for individual synthesized represented separately for ID and ART training groups. Line graphs represent increase/decrease in duration effect scores from pre-test to post-ttest, different line colours indicate vowel set (blue= high-front, green= low, red= high-back), and asterisks indicate significant pre-test/post-test differences in duration effect scores ($p < .05$).

Summary

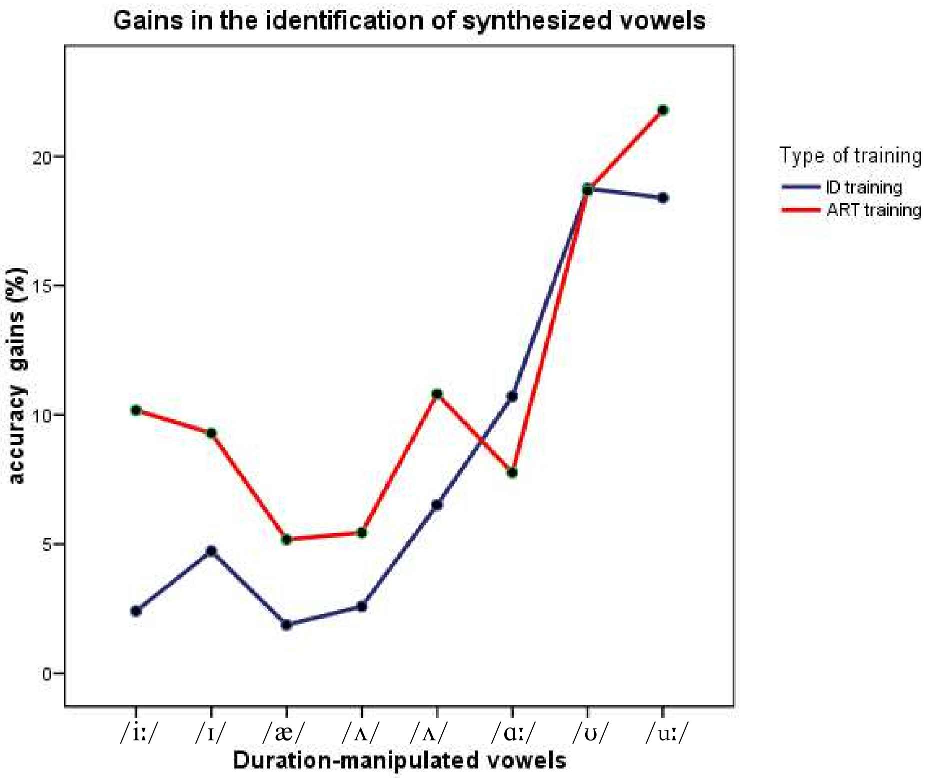
The pre-test mean duration effect scoress provide evidence for the fact that Catalan-Spanish listeners identified /iː/, /ɪ/, /æ/. /ʌ/. /ɑː/, /ʊ/ and /uː/ mainly based on duration cues, because the 5-vowel system of Spanish and the 8-vowel system of Catalan do not provide them with enough experience on small-scale spectral differences. After training, the perception task somehow forced them to resort to spectral cues for accurate vowel identification after the experience received with high-variability training. Even though ID and ART training groups still resorted to duration cues for categorizing /ʊ-uː/, learners generally started to favour spectral cues over temporal cues after training for all synthesized vowels and showed a trend towards identifying s /iːɪæʌɑː/ mainly based on spectral cues rather than only duration cues, or even a combination of spectral and duration cues.

4.2.7. Effects of training and type of training on changes in duration effect scores and gains in percent correct identification of synthesized vowels.

Training significantly improved English learners' identification performance on duration-manipulated vowels about 8.25% (*SD*= 20.07) (ID training) and 12.88% (*SD*= 29.35) (ART training) and learning generalized to vowel pairs involving low and high-back vowels (C3 /æ-ʌ/ and C4 /ʊ-uː/).

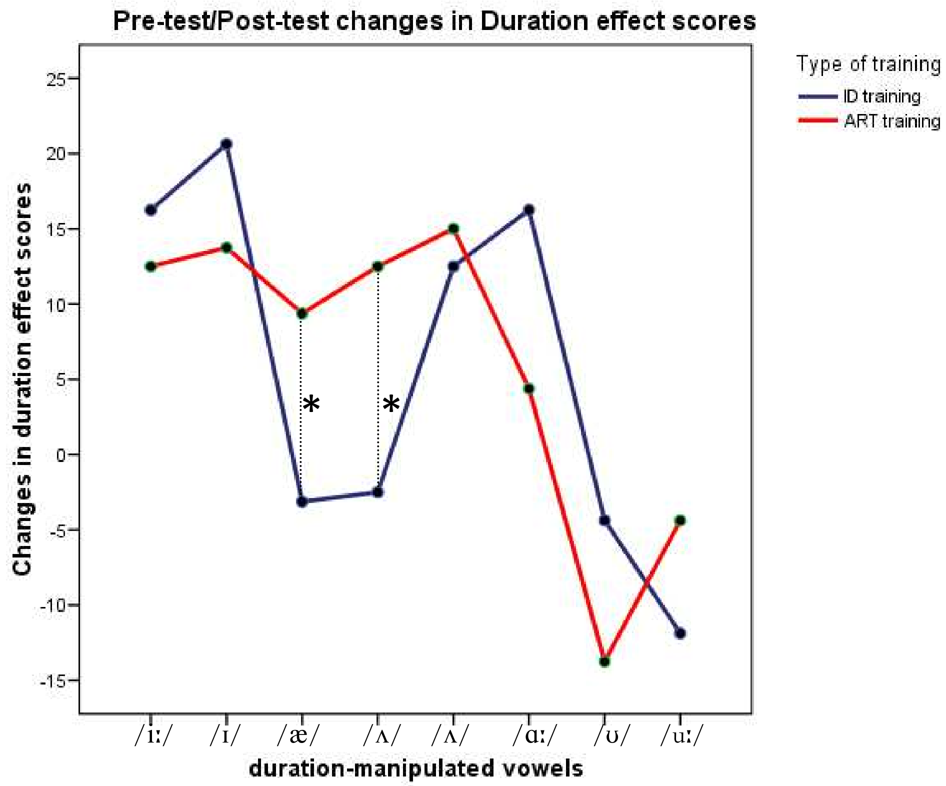
The highest percentage of gains in mean percent correct identification was obtained for /ʊ/ and (M_{ID}= 18.75, *SD*= 25.87; M_{ART}= 18.66, *SD*= 32.03) and **/uː/** (M_{ID} = 18.39, *SD*= 29.28; M_{ART} = 21.78, *SD*= 28.50), especially for the ART group (Figure 4.37). The lowest percentage of gains was found for **/æ/** (M_{ID} = 1.87, *SD*= 9.77; M_{ART} = 5.18, *SD* 21.90) and /ʌ/ (M_{ID}=2.59, *SD* 7.45; M_{ART} = 5.45, *SD*= 22.06), especially for the ID group. For /iːɪʌɑː/ gains ranged between 2.4 – 10.7%. Gains were generally greater for high-back vowels than for high-front and low vowels, for both ID and ART groups (see Figure 4.38).

The effects of *type of training* and *vowel* on correct identification of synthesized vowels were analyzed using a two-way ANOVA, with *vowel* (8) as the within-subjects factor and *type of training* (ID vs. ART) as the between-subjects factor. Results showed a significant effect of *vowel* ($F(7, 434)= 5.42$, $p< .001$, $η^2= .080$) on identification gains and the effect of *type of training* did not reach statistical significance ($F(1, 62)= 0.56$, $p> .05$, $η^2= .009$). There was a significant *vowel x type of training* interaction ($F(7, 434)= 00.35$, $p> .05$, $η^2=.006$). These results suggest that the amount of gains obtained for certain vowels varied as a function of type of training. The descriptive data shows that gains obtained were greater for the high-back vowels than for high-front front and low vowel contrasts; the lowest percentage of gains was found for /æʌ/, and the percentage of identification gains was higher for the ART group than for the ID group (4.63%).



Figures 4.38. Mean identification gain scores for individual synthesized obtained by ID and ART (red) (*N*=32) training groups.

Training changed English learners' duration effect scores (DESs) in the identification of duration-manipulated vowels about 5.82% (ID_{GAINS}: 5.47%, *SD*= 4.63; ART_{GAINS}: 6.17%, *SD*= 4.63) and *changes in DES from pre-test to post-test* generalized to all duration-manipulated vowels. Whereas the DESs of **/iːɪʌ/ decreased by 12-20% from pre-test to post-test, the DESs of/ʊ/-/uː/ increased by 4-13%. ID and ART training groups differed with respect to /æ/- /ʌ/ and /ɑː/ (see Figure 4.39).**



Figures 4.39. Pre-test/post-test changes in duration effect scores for individual synthesized obtained by ID and ART training groups.

Both ID and ART groups showed the highest amount of changes in DES for **/iːɪʌ/.** With respect to /ʊ/-/uː/, ID and ART groups went in opposite directions: at post-test, the ART group obtained higher changes in DES for /ʊ/ and lower DES for /uː/, whereas the ID group showed higher DES for /uː/ and lower DES for /ʊ/. With respect to /æʌ/, a couple of observations are worth noting. First, it appears that listeners did not use duration cues as the dominant perceptual cue to vowel categorization at pre-test. That explains that changes in DES of /æ/and /ʌ/ from pre-test to post-test are small, especially for the ID group. Second, although the ART group had a larger amount of changes in DES for /æ/and /ʌ/ than the ID group, the performance of these two groups was comparable at post-test.

A two-way ANOVA conducted on changes in DESs for each of the synthesized vowels, with *vowel* (8) as within-subjects factor and *type of training* (ID vs. ART) as between-subjects factor, revealed a significant effect of *vowel* ($F(7, 434)= 5.42$, $p< .001$, $\eta^2= .080$). The interaction between *type of training* and *vowel* and the effect of *type of training* were not statistically significant ($p> .05$). These results confirmed that pre-test/pos-test changes in DESs varied as a function of *vowel*.

4.2.8. Summary of effects of training and type of training on perception accuracy

This section has presented the effects of training on learners' ability to identify duration-manipulated tense and lax English vowels (80, 116.7, 153.3, 190, 226.7, 263.3, 300 ms) and has explored the effect of manipulation of duration on *mean percent correct identification* (averaged across the seven steps of the duration continua) and *duration effect scores* (DES).

The results above suggest that non-native listeners over-rely on duration cue for English vowel identification whereas native-speakers do not, for whom spectral cues are the primary cues. Catalan-Spanish listeners' over-reliance on duration cues for English vowel perception explained overall poor performance at pre-test and below chance performance (<50%). In the pre-test, accuracy rates of tense vowels were below chance level (<50%) when they were shorter than 116ms (/uː/), 153.3ms (/iː/) or 190ms (/ɑː/). That is, the shortest stimuli failed to elicit "tense" vowel responses. In the lax condition, the longest stimuli (S7 300 ms) failed to elicit "short" /ɪ/, /ʌ/ or /ʊ/ responses (<50%) at pre-test.

Training significantly improved learners' identification performance on duration-manipulated vowels about 8.25% (ID training) and 12.88% (ART training) and learning generalized to vowel pairs involving low and high-back vowels (æ, ʌ, ʊ,uː/). The highest percentage of gains in mean correct identification was obtained for /ʊ/ and **and /uː/ (**especially for the ART group. The lowest percentage of gains was found for **C2 /æ/ and /ʌ/, especially for the ID group. For /iɪʌɑ/, gains were between 2.4 – 10.7% (<11%). To sump up, g**ains were generally greater for high-back vowels than for high-front and low vowels, for both ID and ART groups. Both ID and ART types of training were thus effective in improving the identification of English vowels even when the duration cue was ambiguous and could not be relied upon for L2 vowel recognition.

Accuracy gains are explained by the impact of HVPT . Post-test identification tended to be more accurate at the short end (S1-S3) of the continuum for tense vowels (/iːɑːuː/), as well as at the longer end (S5-S7) for lax vowels (/ɪʌʊ/). This suggests that training reduced the listeners' over-reliance on duration. On the other hand, it should be noticed that post-test identification accuracy remained low (<50%) at the shortest step(s), S1 for /iː/ and at S1-S2 for /ɑː/, and still remained poor at the longest steps S7 for /ɪ/ and S6-7 for /ʊ/, which meant that subjects seemed to have some difficulty in detecting spectral differences and were still sensitive to the manipulation of duration when certain vowels were extremely lengthened or shortened despite overall improvement:

- For /ɑː/, ID and ART groups still performed poorly at post-test at S1 and S2, but starting from S3 (153,3 ms), the majority of identification shifted to /ɑː/.

- For /uː/, even the shortest stimuli S1 elicited a strong “tense” response (>60%) at post-test.
- With regard to /ɪ/, both groups performed above 50% at all steps of the continuum at post-test.
- For /ʌ/ (within C3), post-test identification remained poor at the longest steps S6 and S7 of the continua for both ID and ART training groups despite overall improvement.
- For /ʊ/, training had a clear visible impact on S7 for both ID and ART groups and, at post-test, both ID and ART groups reached above 60% performance at all steps of the continua.

A trend was observed towards a reduced reliance on duration among “tense” and “lax” vowels after both types of training.

This section has also presented the effects of training on duration effect scores (DES) (duration end point difference), a measure of the listeners’ degree of sensitivity to duration cues for vowel identification. Post-test DESs were generally lower for /iːɪʌɑː/for both ID and ART groups, and were lower for /æʌ/ only after ART training, meaning that listeners generally showed lower sensitivity to duration cues for categorizing those vowels. However, DESs increased for both ID and ART groups for back /ʊ/ and /uː/. Mean DES ranged 8%-86% as a function of vowel and pre-test/post-test changes in DES were similar across groups: DES decreased by 5.47% after ID training and 6.17% after ART training. Training significantly changed English learners’ DESs in the identification of duration-manipulated vowels about 5.82% and changes in DES generalized to all duration-manipulated vowels.

These results suggest that training reduced the listeners' sensitivity towards the manipulation of duration and made them access to spectral differences as an L2 cue in vowel categorization when the duration cue is ambiguous.

Finally, results seem to indicate that the ID training group still over-relied on the duration cue at post-test for categorizing /æ/-/ʌ/ and that, interestingly, both ID and ART groups over-relied on duration for categorizing the back /ʊ/-/uː/ contrast. At this point, differences between ID and ART groups emerged:

- ID and ART showed different response patterns in the identification of /æ/and /ʌ/ (C2, when /æ/-/ʌ/ was a forced-choice respomse). The DESs of the ID training group increased significantly at post-test, whereas the DESs of the ART group decreased significantly. The ART group became more sensitive to quality differences after training, whereas the ID group was still affected by the manipulated of duration in the identification of low vowels.
- ID and ART groups went in opposite directions with respect to back vowels /ʊ/ and /uː/. At post-test, the ART group obtained higher DES for /ʊ/ and lower DES for /uː/, whereas the ID showed higher DES for /uː/ and lower DES for /ʊ/.

In summary, the fact that ID and ART groups identified synthesized vowels significantly better after training and showed lower DES for 6 of the 8 synthesized vowels (/iɑɪæʌʊ/) is meaningful because the findings suggest that:

(1) The two types of high-variability training facilitate the formation of robust vowel phonetic categories, as the results demonstrate that learners were able to apply what they have learned to vowels that had been durationally manipulated. ID and ART training probably improved identification by making the

application of existing/learned categories to durationally manipulated vowels more automatic and efficient.

(2) Given the Catalan-Spanish learners' over-reliance on the duration cue, ID and ART training can shift language learners' attention to weight more relevant cues (spectral cues) when the duration cue is ambiguous (e.g. "short" /iːɑː/ and "long" /ɪæʌʊ/). Learners can benefit from high-variability training without cue-manipulated stimuli in order to obtain a more accurate perception of vowel pairs, based on more relevant cues.

(3) The findings suggest that high-variability training may be needed to focus learners on how efficiently use the spectral cues and diminish the listeners' over-sensitivity to duration differences. However, there was no clear advantage of one type of training over the other.

(4) The improved performance on duration-manipulated tokens at post-test suggests that training on natural and variable tokens (focusing on spectral differences), including the whole vowel category inventory, can be more effective or just as effective as training on both quality and duration differences.

(5) Improved perception on durationally manipulated vowels is probably crucial for L2 production as well, in order to achieve a native-like production of L2 vowels.

4.3. Discrimination of vowels (DIS – task3)

The **Sub-RQ1.3** examines to what extent learners will improve in the discrimination of natural L2 vowels from pre-test to post-test (DIS-task 3). This section presents the effects of training on mean *correct vowel discrimination* as well as the differential effects of identification (ID) and articulatory (ART) types training on this measure (**RQ2.3**).

4.3.1. Comparability of subject groups at pre-test

The ID, ART, and control groups obtained similar pre-test discrimination accuracy scores (see Table 4.25). A one-way ANOVA confirmed that the three groups did not significantly differ from one another at pre-test in vowel discrimination accuracy ($F(2, 75)=0.11$, $p= .898$), suggesting that the groups were indeed comparable in terms of their discrimination skills before training (see Figure 4.40).

Vowel Discrimination at pre-test	Subject groups			
	ID training	ART training	NNS control	NS
Mean vowel discrimination	66.14 (25.77)	64.23 (4.22)	62.22 (26.93)	98.11
C1 /iː/-/ɪ/	65.62 (27.41)	60.94 (30.12)	61.90 (28.81)	75.00 (8.91)
C2 /ɪ/-/e/	76.56 (32.48)	75.00 (33.87)	76.19 (35.03)	100.00 (.000)
C3 /e/-/ɜː/	79.17 (32.52)	73.96 (33.85)	71.43 (31.64)	100.00 (.000)
C4 /ɪ/-/ɜː/	83.84 (33.19)	80.73 (36.44)	80.95 (35.72)	100.00 (.000)
C5 /e/-/æ/	71.35 (32.30)	71.87 (34.76)	73.81 (34.41)	97.92 (5.89)
C6 /ɜː/-/æ/	84.89 (33.43)	81.77 (36.26)	83.33 (35.08)	100.00 (.000)
C7 /æ/-/ʌ/	72.92 (32.17)	70.83 (33.87)	64.28 (30.56)	100.00 (.000)
C8 /ʌ/-/ɑː/	80.00 (34.08)	75.62 (35.46)	71.43 (33.61)	100.00 (.000)
C9 /æ/-/ɑː/	84.37 (33.85)	80.21 (36.28)	80.95 (35.72)	100.00 (.000)
C10 /ɜː/-/ʌ/	77.60 (32.96)	77.60 (36.32)	67.86 (33.63)	100.00 (.000)
C11 /ɜː/-/ɑː/	52.23 (25.30)	52.23 (28.48)	53.06 (31.41)	91.07 (7.39)
C12 /ɒ/-/ɑː/	55.62 (26.75)	63.12 (32.96)	47.14 (28.94)	80.00 (26.19)
C13 /ʌ/-/ɒ/	67.71 (29.91)	67.71 (34.89)	60.71 (30.39)	100.00 (.000)
C14 /ɒ/-/ɔː/	72.39 (32.13)	70.83 (36.17)	72.62 (34.43)	89.58 (8.62)
C15 /ʊ/-/uː/	77.23 (32.43)	76.78 (36.24)	75.51 (33.82)	100.00 (.000)
C16 /ɔː/-/uː/	81.77 (33.16)	80.73 (35.95)	79.76 (35.31)	100.00 (.000)

Table 4.25. Percent correct discrimination scores for the individual vowel contrasts and standard deviations obtained by ID training, ART training, NNS control and NS baseline groups, at pre-test.

4.3.2. Effects of training and type of training on mean percent correct vowel discrimination

The first research question (RQ 1.3) examined whether audiovisual HVPT resulted in a significant improvement in the discrimination of natural English vowels from pre-test to post-test. If the training was effective, the training groups should reveal a significant increase in the discrimination accuracy scores at post-test, while the control should not exhibit such improvement. The second question (RQ 2.3) investigated which type of training was better able to increase vowel discrimination accuracy.

The descriptive statistics for pre-test and post-test discrimination accuracy and improvement scores (gains) for each subject group are shown in Table 4.26. As shown in Figure 4.39, ID and ART training groups increased their mean discrimination accuracy (gains: ID=12.72%, *ART*= 15.09%) at post-test , and the effect of training was generalized to all vowel sets (high-front, low, and back vowel contrasts), whereas the control group remained almost unchanged from pre-test to post-test. Trainees improved from 77.2% to 93.7% in vowel discrimination in *fixed context* and, interestingly, obtained worse post-test scores in *context variability* (from 52% pre-test to 63% post-test) than in the fixed context condition. The vowel contrasts /ɪ/-/ɜː/, /ɜː/-/æ/ and /æ/-/ɑː/ (81-86%) were by far the most accurately discriminated among the 16 vowel contrasts, whereas /ɜː/-/ɑː/and /ɒ/-/ɑː/ were discriminated the most poorly among all the vowel contrasts both at pre-test (52-59%) and post-test (65-70%).

Paired-samples t-tests were run on the pre- and post-test discriminationscores of the control group to determine whether there was a statistically significant difference in mean vowel discrimination accuracy between the two testing times. Results indicated that post-test mean discrimination scores were not significantly higher (gains: *M*= 5.57, *SD*= 15.53) than pre-test scores ($t(19)$= -1.60, p= .125).

A two-way ANOVA with mean percent correct vowel discrimination (discrimination accuracy scores averaged across all vowel contrasts) as the dependent variable , effect of *training* (T1 vs. T2) as the within-subjects factor, and *type of training* (ID vs. ART) as the between-subjects factor, was conducted to determine if the training had been effective, and if differences between testing times were due to the type of training the trainees had received. There was a statistically significant effect of *training* (F(1, 62)= 15.60, p< .001, η2= .201). However, the effect of *type of training* did not reach statistical significance (F(1, 62)= 0.05, p= .831, η^2= .001). The interaction between *training* and *type of training* did not reach significance (F(1, 62)= 0.11, p= .738, η^2= .002). These results confirmed that vowel discrimination scores were significantly higher at post-test than they were at pre-test, suggesting that the training was generally effective in increasing vowel discrimination accuracy. Both training groups showed significant improvement in vowel discrimination (ID training: 12.72%, ART training: 15.09%).

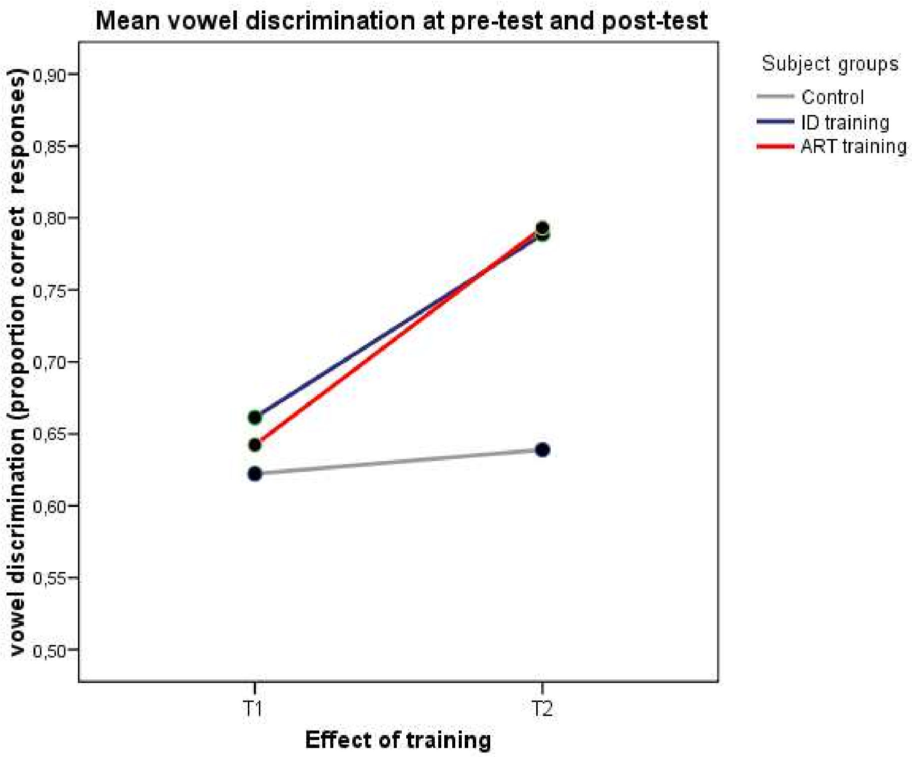


Figure 4.40. Average percent correct vowel discrimination for ID training, ART training and NNS control groups at pre-test and post-test.

Vowel discrimination	ID training (*N*=32) T1	ID training (*N*=32) T2	ID training (*N*=32) gains	ART training (*N*=32) T1	ART training (*N*=32) T2	ART training (*N*=32) gains	Control group (*N*=20) T1	Control group (*N*=20) T2	Control group (*N*=20) gains	NS (*N*=10)
Mean discrimination	66.14 (25.77)	78.86 (28.54)	*12.72 (26.26)*	64.23 (4.22)	79.32 (4.25)	*15.09 (29.94)*	62.22 (26.93)	63.90 (0.27)	1.67 (2.42)	98.11
fixed context	78.91 (30.68)	93.90 (4.35)	*14.99 (32.46)*	75.53 (33.73)	93.55 (3.76)	*18.02 (34.54)*	74.66 (32.28)	77.45 (33.18)	2.79 (3.48)	
Context variability	52.05 (20.62)	62.25 (5.59)	*10.20 (20.88)*	51.61 (23.28)	63.52 (6.34)	*11.91 (25.19)*	48.55 (21.66)	49.11 (21.95)	0.55 (1.69)	
C1 /iː/-/ɪ/	65.62 (27.41)	73.96 (12.65)	8.33 (29.93)	60.04 (30.12)	77.08 (9.23)	16.14 (32.82)	61.90 (28.81)	63.09 (28.63)	1.19 (4.45)	75.00 (8.91)
C2 /ɪ/-/e/	76.56 (32.48)	92.71 (8.40)	16.14 (31.82)	75.00 (33.87)	93.75 (8.20)	18.75 (35.10)	76.19 (35.03)	76.19 (35.03)	-	100.00 (.000)
C3 /e/-/ɜː/	79.17 (32.52)	92.71 (10.31)	13.54 (31.80)	73.96 (33.85)	92.19 (10.36)	18.23 (30.92)	71.43 (31.64)	77.38 (34.35)	5.95 (8.29)	100.00 (.000)
C4 /ɪ/-/ɜː/	83.85 (33.19)	97.92 (5.60)	14.06 (34.16)	80.73 (36.44)	98.96 (4.10)	18.23 (37.23)	80.95 (35.72)	83.33 (36.40)	2.38 (6.05)	100.00 (.000)
C5 /e/-/æ/	71.35 (32.30)	89.58 (13.88)	18.23 (29.74)	71.87 (34.76)	93.23 (10.25)	21.35 (38.39)	73.81 (34.41)	75.00 (35.01)	1.19 (4.45)	97.92 (5.89)
C6 /ɜː/-/æ/	84.89 (33.43)	99.48 (2.95)	14.58 (33.80)	81.77 (36.26)	99.48 (2.95)	17.71 (36.40)	83.33 (35.08)	84.52 (36.08)	1.19 (4.45)	100.00 (.000)
C7 /æ/-/ʌ/	72.92 (32.17)	90.62 (12.65)	17.71 (34.11)	70.83 (33.87)	96.35 (7.00)	25.52 (36.66)	64.28 (30.56)	71.43 (33.93)	7.14 (10.77)	100.00 (.000)
C8 /ʌ/-/ɑː/	80.00 (34.08)	96.87 (11.48)	16.87 (36.32)	75.62 (35.46)	95.00 (13.44)	19.37 (39.51)	71.43 (33.61)	75.71 (35.24)	4.28 (11.58)	100.00 (.000)
C9 /æ/-/ɑː/	84.37 (33.85)	97.39 (7.46)	13.02 (13.02)	80.21 (36.28)	96.35 (8.18)	16.14 (37.26)	80.95 (35.72)	82.14 (35.48)	1.19 (4.45)	100.00 (.000)
C10 /ɜː/-/ʌ/	77.60 (32.96)	98.44 (4.93)	20.83 (32.79)	77.60 (36.32)	94.79 (9.87)	17.19 (36.53)	67.86 (33.63)	72.62 (34.96)	4.76 (7.81)	100.00 (.000)
C11 /ɜː/-/ɑː/	52.23 (25.30)	67.41 (18.94)	15.18 (32.20)	52.23 (28.48)	62.50 (24.00)	10.27 (34.08)	53.06 (31.41)	53.06 (31.41)	.000	91.07 (7.39)
C12 /ɒ/-/ɑː/	55.62 (26.75)	67.50 (21.40)	11.87 (33.26)	63.12 (32.96)	72.50 (18.14)	9.37 (39.34)	47.14 (28.94)	48.57 (29.05)	1.43 (5.34)	80.00 (26.19)
C13 /ʌ/-/ɒ/	67.71 (29.91)	86.46 (16.63)	18.75 (32.72)	67.71 (34.89)	81.25 (17.83)	13.54 (39.81)	60.71 (30.39)	60.71 (30.39)	-	100.00 (.000)
C14 /ɒ/-/ɔː/	72.39 (32.13)	84.89 (13.62)	12.50 (32.79)	70.83 (36.17)	90.62 (13.34)	19.79 (39.58)	72.62 (34.43)	75.00 (33.17)	2.38 (6.05)	89.58 (8.62)
C15 /ʊ/-/uː/	77.23 (32.43)	93.30 (12.03)	16.07 (33.40)	76.78 (36.24)	97.32 (6.73)	20.53 (36.27)	75.51 (33.82)	76.53 (33.90)	1.02 (3.82)	100.00 (.000)
C16 /ɔː/-/uː/	81.77 (33.16)	96.35 (8.18)	14.58 (32.99)	80.73 (35.95)	96.87 (7.85)	16.14 (36.04)	79.76 (35.31)	79.76 (35.31)	-	100.00 (.000)

Table 4.26. Percent correct discrimination scores, gains and standard deviations obtained in the pre-test and post-test by ID ART training, control and NS baseline groups.

We also examined whether the general improvement in vowel discrimination was consistent across different vowel sets and test conditions. Subsequent ANOVAs included within-subject factors such as *vowel set* (**high-front, low-mid, high-back vowels), *vowel contrast* (12)** and *test condition* (fixed context vs. context variability) affecting discrimination accuracy.

Summary

The results of the discrimination task confirmed that HVPT significantly improved English vowel discrimination by Catalan-Spanish speakers as shown by the success with which trainees discriminated English vowel contrasts in the post-test (M_{T2}= 79.1) compared to the pre-test (M_{T1}= 65.2). The main effect of type of training on post-test discrimination was non-significant.

4.3.3. Effects of training and type of training on discrimination of distinct vowel sets

Results for accuracy in the discrimination of vowel sets (**high-front, low, high-back vowels**) at pre-test and post-test, and discrimination gains, are presented on Table 4.27 (RQ1.3a and RQ2.3a). Figure 4.41 shows the discrimination accuracy for each vowel set and each training group at pre-test and post-test. Generally, the post-test discrimination scores were higher than the pre-test scores for the three vowel sets, and accuracy in vowel discrimination varied as a function of vowel set. Trainees improved on the discrimination of high-front (gains 13-18%), low (gains 16-17%), and back (gains 14-18%) vowels regardless of the type of training. Both ID and ART training groups showed a similar pattern of discrimination results at pre-test: the back vowel contrasts (/ɒ/-/ɔː/, /ʊ/-/uː/, /ɔː/-/uː/) got the highest accuracy in discrimination, followed by high-front (/iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/, /ɪ/-/ɜː/), and low (/e/-/æ/,/ɜː/-/æ/, /æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɜː/-/ʌ/, /ɜː/-/ɑː/, /ɒ/-/ɑː/, /ʌ/-/ɒ/) contrasts (see Table 4.27). At post-test, back vowel contrasts (M_{ID}= 86.98, M_{ART}= 85.97) were still discriminated at higher accuracy rates than high-front (M_{ID}= 66.42 M_{ART}= 78.68) and low (M_{ID}= 65.23, M_{ART}= 63.11) vowels contrasts. The ID training got the highest percentage of gains on low vowels (16%) whereas the ART obtained the greatest gains with back vowel contrasts.

Vowel DIS	Experimental group (N=64)						NS (N=10)
	ID training (N=32)			ART training (N=32)			
	T1	T2	gains	T1	T2	gains	
Vowel set 1 (C1-4)	76.30 (30.22)	89.32 (5.18)	13.02 (30.10)	72.66 (32.36)	90.49 (4.91)	17.84 (32.26)	93.75 (2.23)
Vowel set 2 (C5-13)	71.86 (28.45)	88.20 (6.27)	16.34 (29.0)	71.22 (31.98)	87.94 (6.62)	16.72 (34.65)	96.55 (2.87)
Vowel set 3 (C14-16)	77.13 (30.79)	91.52 (7.88)	14.38 (31.25)	76.12 (34.12)	94.94 (5.43)	18.82 (35.34)	96.53 (2.87)

Table 4.27. Percent correct discrimination scores, gains and standard deviations for the three vowel sets (high-front, mid, low) obtained by ID training, ART training and NS groups.

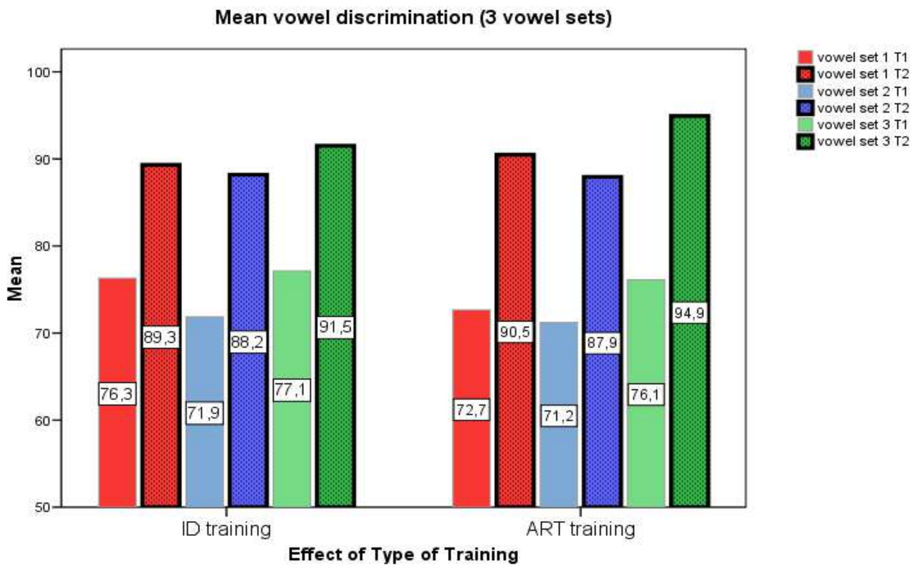


Figure 4.41. Mean percent correct discrimination at pre-test and pos-test for each vowel set (1= high-front, 2=low, 3=back) by ID and ART training groups.

A three-way ANOVA tested the effects of *training* (T1 vs. T2), *vowel set* (high-front, low, back) and *type of training* (ID vs. ART) on mean correct vowel discrimination. There was a significant effect of *training* (F(2, 124)= 32.14, p< .001, η^2= .201) and *vowel set* (F(2, 124)= 32.15, p< .001, η^2= .342), but the effect of *type of training* did not reach significance.

The *training x vowel set x type of training* interaction ($F(2, 124)= 2.93$, $p= .057$, $\eta^2= .045$) and other interactions did not reach significance. These results revealed that discrimination scores of the three vowel sets increased significantly at post-test and varied significantly as a function of *vowel set*, but not as a function of *type of training*. Bonferroni-adjusted pairwise comparisons revealed that mean discrimination scores from subset 3 (76-77%) were significantly higher than scores obtained for subset 1 (72-76%) and 2 (71%) before training at the $p< .001$ level. After training, discrimination scores obtained for subset 3 (91-94%) also continued to be significantly higher than scores obtained for subsets 2 (88-87%) and 1 (89-90%), regardless of the type of training received ($p< .001$). This indicates a significant difference in discrimination accuracy as a function of the quality of the vowel contrasts, high-front vowel pairs being discriminated at lower percent correct rates than mid and back vowel pairs. Back vowel pairs were discriminated more accurately than the rest, reaching ceiling scores, whereas low vowel pairs were the most poorly discriminated. The significant effect of training and the lack of significant differences between ID and ART groups confirm that ID and ART training methods were similarly effective in improving vowel discrimination and improvement generalized to all vowel sets.

To asses which type of training was better able to increase vowel discrimination accuracy, a series of t-tests were conducted on pre-test and post-test discrimination scores and gains. As shown in Table 4.28, the t-tests only yielded significant differences between ID and ART training groups for back vowel contrasts (vowel set 3) at post-test. These confirmed that the ART group discriminated vowel set 3 significantly better than the ID group. With respect to discrimination gains, we observed that the ART training group obtained slightly larger gains than ID group for all vowel sets, esp. for vowel set 3. The ID group obtained greater gains for vowel set 2 than the ART group, whereas the ART group obtained greater gains than the ID group in the discrimination of high-front and back vowel contrasts. However, the t-tests confirmed that ID and ART groups did not differ significantly with respect to the % of gains obtained for any of the vowel sets.

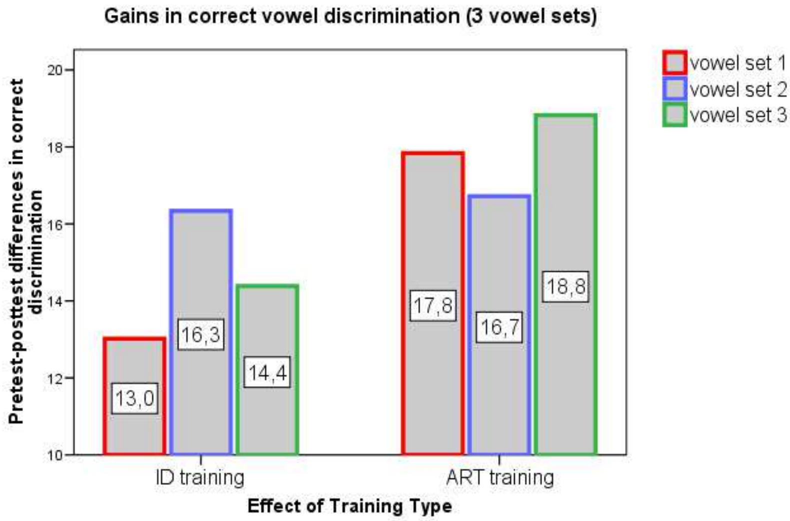


Figure 4.42. Mean percent discrimination gains across all subjects in the ID and ART training groups for vowel set 1 (red; high-front), vowel set 2 (blue; low) and vowel set 3 (green; back).

Vowel Discrimination	Between-group differences: ID (***N*=32**) vs. ART training (***N*=32**)		
	T1	**T2**	**gains**
Vowel set 1	t(62)=0.50, p= .618	t(62)=-1.30, p= .200	t(62)=-0.67, p= .502
Vowel set 2	t(62)=0.50, p= .617	t(62)=0.14, p= .889	t(62)=-0.45, p= .657
Vowel set 3	t(62)=0.37, p= .624	t(62)=-2.39, p< .05*	t(62)=-0.91, p= .365

Table 4.28. Results of independent-samples t-tests on the discrimination scores of 3 vowel sets (high-front, mid, low) at pre-test and post-test and accuracy gains obtained by training groups.

Summary

The effect of training was generalized to all vowel sets (high-front, low, and back vowel pairs). Discrimination scores showed that high-back vowels (vowel set 3: /ɒ/-/ɔː/, /ʊ/-/uː/, /ɔː/-/uː/) were discriminated more accurately than high-front (vowel set 1: /iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/, /ɪ/-/ɜː/) and low vowels (vowel set 2: /e/-/æ/, /ɜː/-/æ/, /æ/-/ʌ/, /ʌ/-/aː/, /æ/-/aː/, /ɜː/-/ʌ/, /ɒ/-/aː/, /ʌ/-/ɒ/), both at pre-test and post-test, and low vowel pairs (vowel set 2) were discriminated *worse than high pairs (vowel sets 1 and 3). When discrimination gains were compared across training groups, the ART group showed a higher percentage of gains for high-front (vowel set 1) and high-back (vowel set 3) vowel pairs (VS1: 18%, VS3: 19%) than the ID group (VS1: 13%, VS3: 14%). For the ID group, the highest amount of gains was obtained for low vowels, whereas the ART training group obtained the highest percentage of gains for back vowel pairs.*

4.3.4. Effects of training and type of training on discrimination of vowel contrasts

Results for accuracy in the discrimination of vowels at pre-test and post-test, previously presented on Table 4.31 show that trainees improved on the discrimination of the sixteen vowel contrasts (RQ1.3b and RQ2.3b).

The NNS control group showed almost no or little improvement from pre-test to post-test (Figures 4.32), and differences between pre-test and post-test performance were non-significant for most of the vowel contrasts (see Table 4.31, Figure 4.43).

Figures 4.44 and 4.45 show the discrimination for each vowel produced after ID and ART training, respectively. Overall, the descriptive data indicated that training was effective in improving the discrimination of the English vowels and no evidence of advantage for any training group was shown. Both ID and ART types of training increased the ability to discriminate natural English vowels correctly.

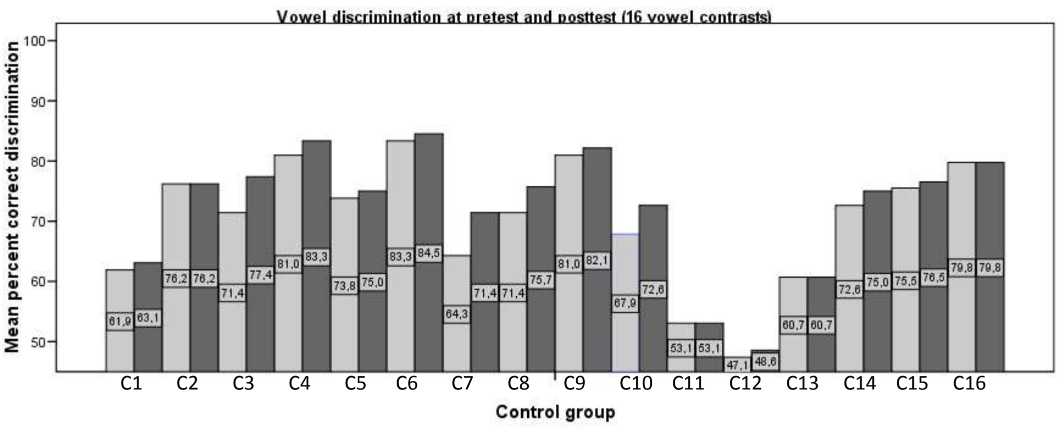


Figure 4.43. Mean discrimination accuracy across all subjects in the control group, presented separately for each vowel contrast at pre-test (light grey) and post-test (dark grey).

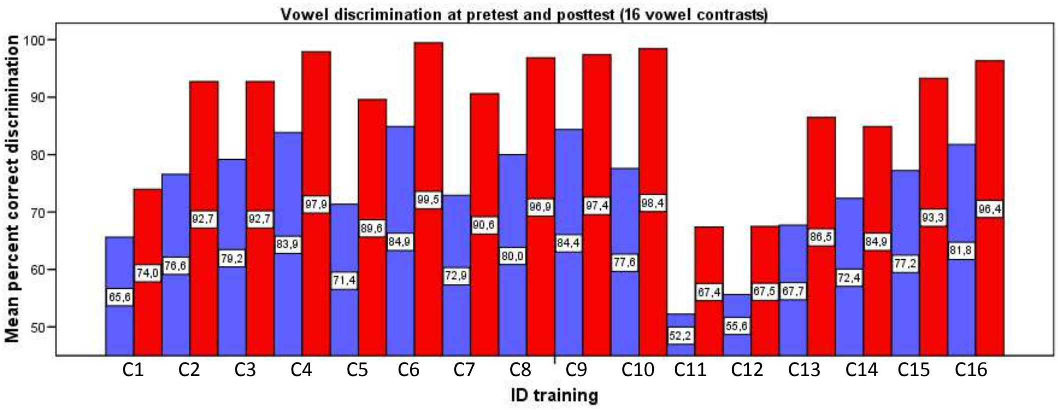


Figure 4.44. Mean discrimination accuracy across all subjects in the ID training group, presented separately for each vowel contrast at pre-test (blue) and post-test (red).

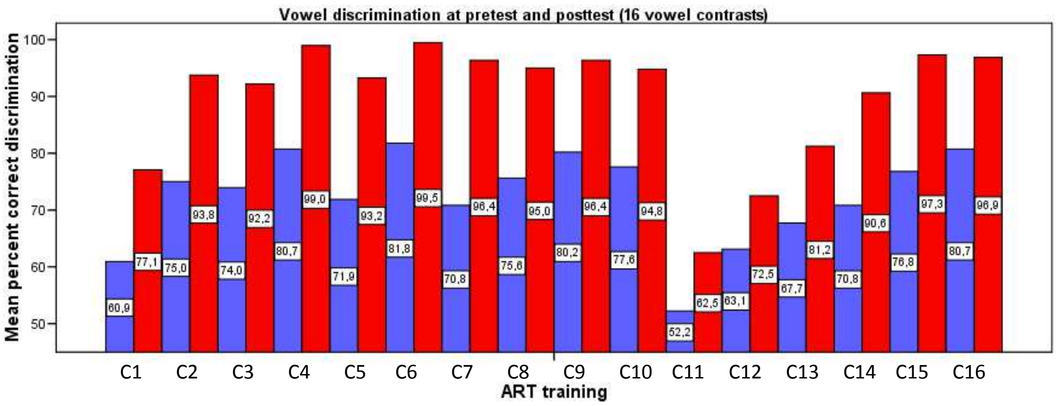


Figure 4.45. Mean discrimination accuracy across all subjects in the ART training group, presented separately for each vowel contrast at pre-test (blue) and post-test (red).

As shown in Figures 4.43-4.44, the vowel contrasts /ɪ/-/ɜː/, /ɜː/-/æ/ and /æ/-/ɑː/ (81-86%) were by far the most accurately discriminated among the 16 vowel contrasts, whereas /ɜː/-/ɑː/and /ɒ/-/ɑː/ were discriminated the most poorly among all the vowel contrasts before (52-59%) and after training (65-70%). Training was effective in improving the discrimination of the 16 English natural vowels and no evidence of advantage for any training group was observed. In general, the ART group obtained a similar percentage of gains as the ID group in all vowel contrasts, except for VC11 (/ɜː/-/ɑː/), VC12 (/ɒ/-/ɑː/), and VC13 (/ʌ/-/ɒ/), for which the ID group obtained a higher percentage of gains.

A mixed between-within ANOVA run on pre-test and post-test discrimination scores, with the effect of *training* (T1 vs. T2) and *vowel contrast* (16) as within-subjects factors and *type of training* (ID vs. ART) as between-subjects, yielded a significant effect of *training* ($F(1, 62)= 16.60$, $p<.001$, $\eta^2=.211$) and a significant effect of *vowel contrast* ($F(15, 930)= 62.72$, $p<.001$, $\eta^2=.503$) on discrimination. The effect of *type of training* did not reach significance ($F(1, 62)= 0.01$, $p=.930$, $\eta^2=.000$) but there was a significant *training x vowel contrast* interaction ($F(15, 930)= 2.55$, $p=.001$, $\eta^2=.040$). The other interactions were not statistically significant. The results suggested that the effect of training on discrimination accuracy varied as a function of vowel contrast, rather than as a function of the type of training received. Improvement in vowel discrimination varied as a function of vowel. In fact, following ID and ART training, listeners still had great difficulty at post-test with some of the contrasts involving the low tense vowel /ɑː/, such as C11 /ɜː/-/ɑː/ and C12 /ɒ/-/ɑː/ (<70%), and could only partially distinguish the /iː/-/ɪ/ contrast well, whereas they performed at ceiling (>95%) for the high-front /ɪ/-/ɜː/ contrast and four low contrasts (/ɜː/-/æ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɜː/-/ʌ/). On average, listeners distinguished the other contrasts relatively well (80-90%): /ɪ/-/e/, /e/-/ɜː/, /e/-/æ/, /æ/-/ʌ/, /ʌ/-/ɒ/, /ɒ/-/ɔː/, /ʊ/-/uː/.

Paired-samples t-tests conducted on the pre-test and post-test discrimination scores of ID and ART groups revealed significant gains for most of the contrasts (Table 4.29). ID and ART groups obtained significant gains for 15 and 12 of the vowel contrasts, respectively. However, between-group differences in mean percent gains turned out non-significant according to an independent-samples t-tests conducted on discrimination gains for each vowel contrast (p> .05). The findings indicated that were no significant effects of type of training on the listeners' discrimination of English vowels.

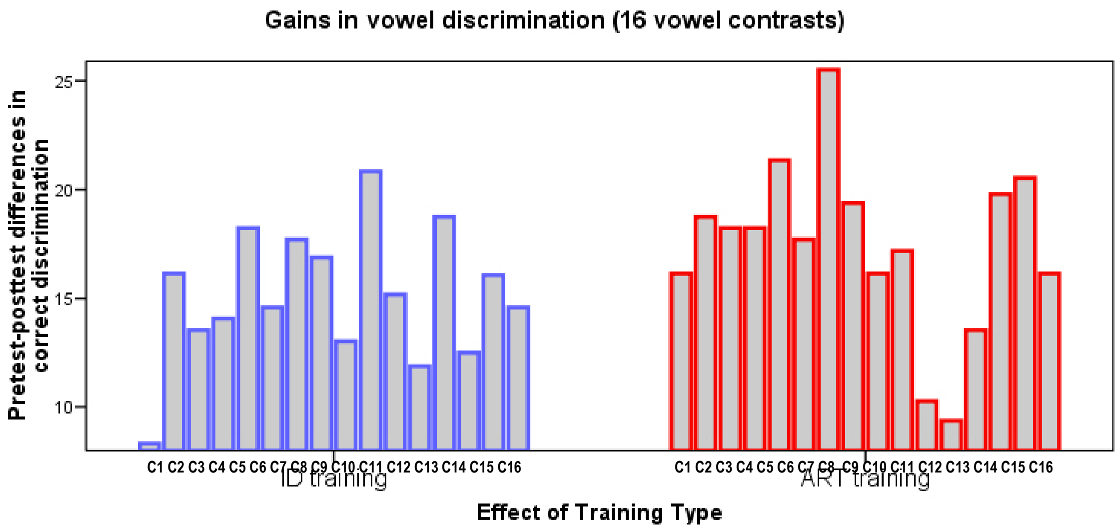


Figure 4.46. Mean discrimination gain scores for 16 vowels obtained by ID and ART training groups.

Summary

HVPT was effective in significantly improving the discrimination of the 16 English natural vowels and no evidence of advantage for any type of training was shown.

The vowel contrasts /ɪ/-/ɜː/, /ɜː/-/æ/ and /æ/-/ɑː/ (81-86%) were by far the most accurately discriminated among the 16 vowel contrasts, whereas /ɜː/-/ɑː/and /ɒ/-/ɑː/ were discriminated the most poorly among all the vowel contrasts before (M_{T1}= 52-59%) and after training (M_{T2}= 65-70%). In general, the ART group obtained a higher percentage of gains than the ID group in all vowel contrasts, except for /ɜː/-/ɑː/, /ɒ/-/ɑː/, and /ʌ/-/ɒ/, but between-group differences in mean percent gains turned out non-significant. The findings indicated that were no significant main effects of type of training on the learners' discrimination of English vowels.

Vowel contrasts	Paired comparisons (vowel discrimination)					
	Control (*N*=20)		**ID training (*N*=32)**		**ART training (*N*=32)**	
	df t p-valu	**Diff.**	**df t p-valu**	**Diff.**	**df t p-valu**	**Diff.**
C1 /iː/-/ɪ/	$t(20)= -1.00, p= .336$	1.19 (4.45)	$t(32)= 0.08, p= .125$	8.33 (29.93)	$t(32)= 0.08$, $\mathbf{p< .007}$*	16.14 (32.82)
C2 /ɪ/-/e/	$t(20)= -1.06, p= .376$	-	$t(32)= 0.08$, $\mathbf{p< .007}$*	16.14	$t(32)= 0.08$, $\mathbf{p= .005}$*	18.75
C3 /e/-/ɜː/	$t(20)= -2.69$, $\mathbf{p< .05}$*	5.95	$t(32)= 0.08$, $\mathbf{p< .05}$*	13.54	$t(32)= 0.08$, $\mathbf{p= .002}$*	18.23
C4 /ɪ/-/ɜː/	$t(20)= -1.47, p= .165$	2.38	$t(32)= 0.08$, $\mathbf{p< .05}$*	14.06	$t(32)= 0.08$, $\mathbf{p= .009}$*	18.23
C5 /e/-/æ/	$t(20)=-1.00, p= .336$	1.19	$t(32)= 0.08$, $\mathbf{p= .002}$*	18.23	$t(32)= 0.08$, $\mathbf{p= .004}$*	21.35
C6 /ɜː/-/æ/	$t(20)= -1.00, p= .336$	1.19	$t(32)= 0.08$, $\mathbf{p< .05}$*	14.58	$t(32)= 0.08$, $\mathbf{p< .05}$*	17.71
C7 /æ/-/ʌ/	$t(20)= -2.48$, $\mathbf{p< .05}$*	7.14	$t(32)= 0.08$, $\mathbf{p= .006}$*	17.71	$t(32)= 0.08$, $\mathbf{p< .001}$*	25.52
C8 /ʌ/-/ɑː/	$t(20)= -1.38, p= .189$	4.28	$t(32)= 0.08$, $\mathbf{p< .05}$*	16.87	$t(32)= 0.08$, $\mathbf{p< .09}$*	19.37
C9 /æ/-/ɑː/	$t(20)= -1.00, p= .336$	1.19	$t(32)= 0.08$, $\mathbf{p< .05}$*	13.02	$t(32)= 0.08$, $\mathbf{p< .05}$*	16.14
C10 /ɜː/-/ʌ/	$t(20)= -2.29$, $\mathbf{p< .05}$*	4.76	$t(32)= 0.08$, $\mathbf{p= .001}$*	20.83	$t(32)= 0.08$, $\mathbf{p< .05}$*	17.19
C11 /ɜː/-/ɑː/	$t(20)= -2.26, p= .386$	-	$t(32)= 0.08$, $\mathbf{p< .05}$*	15.18	$t(32)= 0.08, p= .098$	10.27
C12 /ɒ/-/ɑː/	$t(20)= -1.00, p= .336$	1.43	$t(32)= 0.08$, $\mathbf{p= .05}$*	11.87	$t(32)= 0.08, p= .187$	9.37
C13 /ʌ/-/ɒ/	$t(20)= -2.16, p= .346$	-	$t(32)= 0.08$, $\mathbf{p= .003}$*	18.75	$t(32)= 0.08, p= .064$	13.54
C14 /ɒ/-/ɔː/	$t(20)= -2.28, p= .165$	2.38	$t(32)= 0.08, ,$ $\mathbf{p= .05}$*	12.50	$t(32)= 0.08$, $\mathbf{p= .008}$*	19.79
C15 /ʊ/-/uː/	$t(20)= -1.00, p= .336$	1.02	$t(32)= 0.08$, $\mathbf{p< .05}$*	16.07	$t(32)= 0.08$, $\mathbf{p= .003}$*	20.53
C16 /ɔː/-/uː/	$t(20)= -0.24, p= .237$	-	$t(32)= 0.08$, $\mathbf{p< .05}$*	14.58	$t(32)= 0.08$, $\mathbf{p< .05}$*	16.14

Table 4.29. Results of paired-samples t-tests on pre-test and post-test vowel discrimination scores for 16 vowel contrasts and accuracy gains (diff.) obtained by ID training, ART training and control groups.

4.3.5. Effects of training and type of training on vowel discrimination in different test conditions (fixed context vs. context variability)

The discrimination data were also analyzed to explore the effects of context variability on discrimination accuracy and any interactions between the effects of training and the effects of context variability. In the following analyses we tested the effect of training on vowel discrimination in two different *test conditions* (fixed context vs. context variability) (RQ1.3c) and explored the effects of type of training across test condtions (RQ2.3c) As shown in the previous Table 4.31, listeners discriminate vowels in the fixed context condition at better rates than vowels in context variability condition, both at pre-test and post-test. Figure 4.46 displays mean correct discrimination in all contexts, fixed context and context variability, showing improvement from pre-test to post-test. Figure 4.48 shows trainees revealed greater discrimination gains in fixed context than in context variability conditions.

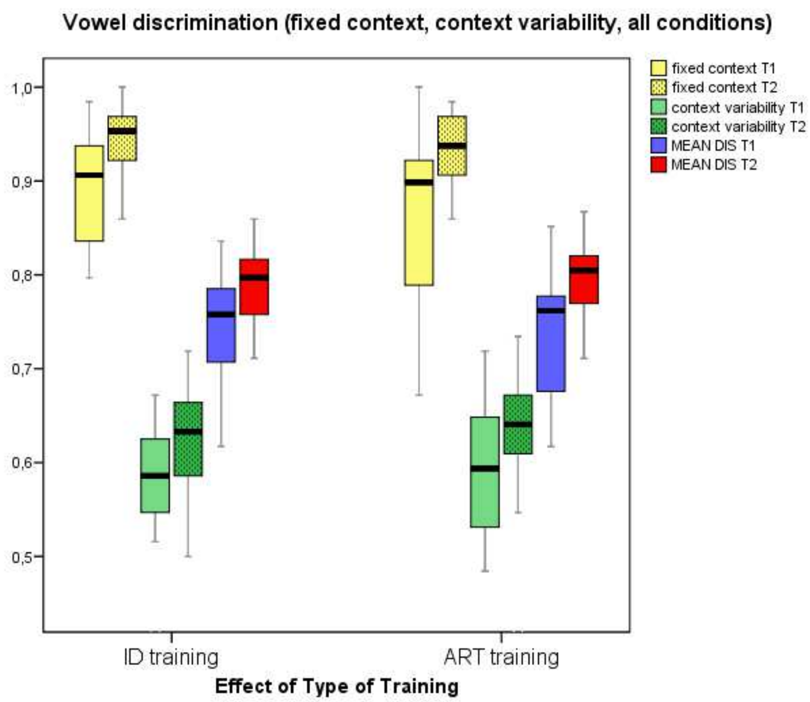


Figure 4.47. Percent correct discrimination of vowels in fixed context, context variability and all conditions obtained by ID and ART training groups.

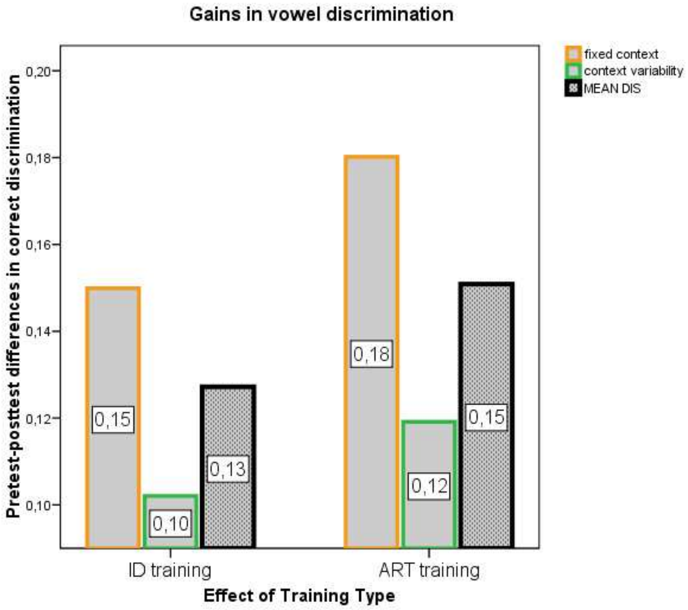


Figure 4.48. Mean percent discrimination gains in fixed context (yellow) and context variability (green) and all conditions (black) across all subjects in ID training and the ART training.

A three-way ANOVA run on mean discrimination scores (averaged across all vowel contrasts) with *test condition* (fixed vs. variability) and effect of *training* (T1 vs. T2) as within-subjects factors, and *type of training* (ID vs. ART) as between-subjects factor, yielded a significant main effect of *training* ($F(1, 62)= 15.62$, $p< .001$, $\eta^2= .201$) and *test condition* ($F(1, 62)= 927.81$, $p< .001$, $\eta^2= .937$), but the main effect of *type of training* was not statistically significant ($F(1, 62)= 0.05$, p= .829, $\eta^2= .001$). These indicated, on the one hand, that the training had a significant impact on discrimination scores and, on the other, that discrimination differed significantly across different test conditions. ID and ART training groups did not differ in any of the test conditions. There was a significant interaction between *training* and *condition* ($F(1, 62)= 13.47$, $p= .001$, $\eta^2= .178$), indicating that the effect of training on discrimination varied as a function of test condition, with listeners of both training groups performing more accurately in fixed than variable contexts both at pre-test and post-test (see Figure 4.47). Discrimination accuracy differed significantly across test conditions both at pre-test ($t(63)= -2.15$, $p= .035$) and post-test ($t(63)= 10.43$, $p< .001$)

despite overall improvement after training, with context variability imposing a greater difficulty in the learners' ability to distinguish different vowel sounds. The *type of training* did not lead to differences in discrimination in fixed context or context variability.

In order to further explore this two-way *training x test condition* interaction, paired-samples t-tests were run separately on each test condition for each vowel contrast. The results showed a significant effect of training for 9 out of the 16 vowel contrasts in each test condition (< .05) (Table 4.35). There was a significant improvement for vowel contrasts /iː/-/ɪ/, /ɪ/-/ɜː/, /ɜː/-/æ/, /æ/-/ɑː/ and /ɜː/-/ʌ/ in fixed context and context variability. However, t-tests revealed some differences across test conditions for certain contrasts:

(1) There was a significant impact of training on high-back vowel contrasts (/ʊ/-/uː/ and /ɔː/-/uː/) in *fixed context*.

(2) Trainees improved significantly on the discrimination of low back vowel contrasts (/ɜː/-/ɑː/, /ɒ/-/ɑː/, /ʌ/-/ɒ/) in *context variability*.

(3) Mid vowel contrasts /e/-/æ/, /e/-/ɜː/ were significantly better discriminated in *fixed context* after training.

Trainees obtained larger discrimination in fixed context (ID: 15%; ART: 18%) than in context variability (ID: 10%; ART: 11%) and, in general, the ART obtained larger gains than the ID discrimination. Paired-sample t-tests comparing discrimination scores in fixed context and context variability revealed signinficantly greater discrimination gains in the context condition for the majority of vowel contrasts: /iː/-/ɪ/ ($t(63)= -12.61$, $p< .001$), /ɪ/-/e/ ($t(63)= 5.97$, $p< .001$), /e/-/ɜː/ ($t(63)= -2.11$, $p= .039$), /e/-/æ/ ($t(63)= -4.06$, $p< .001$), /ʌ/-/ɑː/ ($t(63)= -2.25$, $p= .028$), /ɜː/-/ʌ/ ($t(63)= -2.94$, $p= .005$), /ɜː/-/ɑː/ ($t(63)= -9.13$, $p< .001$), /ɒ/-/ɑː/ ($t(63)= -5.34$, $p< .001$), /ʌ/-/ɒ/ ($t(63)= -4.37$, $p< .001$), /ɒ/-/ɔː/ ($t(63)= -1.99$, $p= .051$), and /ɔː/-/uː/ ($t(63)= -2.11$, $p= .039$).

Vowel Discriminatin	Paired comparisons (T1 vs. T2) ID and ART group (N= 64)			
	Test condition			
	Fixed context		Context variability	
	T1 vs. T2	gains and SD	T1 vs. T2	gains and SD
Mean discrimination	t(63)= -1.79, *p*= .079	7.18 (36.64)	t(63)= 0.21, *p*= .833	0.98 (36.96)
C1 /iː/-/ɪ/	t(63)= -1.79, *p*= .078*	8.98 (40.18)	t(63)= 4.14, *p*< .001*	20.70 (39.98)
C2 /ɪ/-/e/	t(63)= -0.91, *p*= .368	4.30 (37.94)	t(63)= -3.21, *p*= .002*	-14.06 (35.04)
C3 /e/-/ɜː/	t(63)= -2.37, *p*= .021*	10.94 (36.97)	t(63)= -1.51, *p*= .135	-6.25 (33.03)
C4 /ɪ/-/ɜː/	t(63)= -2.91, *p*= .005*	13.28 (36.45)	t(63)= -2.98, *p*= .004*	-13.28 (35.63)
C5 /e/-/æ/	t(63)= -2.43, *p*= .018*	11.33 (37.25)	t(63)= -0.31, *p*= .759	-1.56 (40.55)
C6 /ɜː/-/æ/	t(63)= -3.18, *p*= .002*	14.06 (35.32)	t(63)= -3.09, *p*= .003*	-13.67 (35.33)
C7 /æ/-/ʌ/	t(63)= -1.64, *p*= .107	8.20 (40.10)	t(63)= -1.52, *p*= .133	-7.81 (41.04)
C8 /ʌ/-/ɑː/	t(63)= -1.92, *p*= .060	9.76 (40.73)	t(63)= -0.49, *p*= .624	-2.73 (44.35)
C9 /æ/-/ɑː/	t(63)= -2.62, *p*= .011*	12.11 (37.00)	t(63)= -2.25, *p*= .028*	-9.76 (34.68)
C10 /ɜː/-/ʌ/	t(63)= -3.73, *p*< .001*	17.58 (37.71)	t(63)= -2.49, *p*= .015*	-10.55 (33.87)
C11 /ɜː/-/ɑː/	t(63)= 0.91, *p*= .367	4.69 (41.28)	t(63)= 6.60, *p*< .001*	46.87 (56.78)
C12 /ɒ/-/ɑː/	t(63)= 1.30, *p*= .199	7.81 (48.15)	t(63)= 4.99, *p*< .001*	30.08 (48.22)
C13 /ʌ/-/ɒ/	t(63)= -0.89, *p*= .378	4.69 (42.23)	t(63)= 2.24, *p*= .028*	14.45 (51.52)
C14 /ɒ/-/ɔː/	t(63)= -1.05, *p*= .298	5.47 (41.66)	t(63)= 0.00, *p*= 1.00	0.00 (39.34)
C15 /ʊ/-/uː/	t(63)= -2.17, *p*= .034*	10.16 (37.46)	t(63)= -1.82, *p*= .073	-8.98 (39.43)
C16 /ɔː/-/uː/	t(63)= -2.74, *p*= .008*	12.50 (36.46)	t(63)= -1.64, *p*= .105	7.81 (38.02)

Table 4.30. Results of paired-samples t-tests on pre-test and post-test vowel discrimination scores for individual vowel contrasts, and accuracy gains, in fixed context and context variability conditions.

Summary

Trainees improved from 77.2% to 93.7% in vowel discrimination in fixed context, but obtained worse post-test scores when they discriminated vowels in context variability (from 52% pretraining to 63% posttraining). The effect of training on discrimination varied as a function of the test condition: different contexts lowered discrimination performance. Although ID and ART groups performed significantly better in fixed context and context variability conditions after training, they showed more accurate performance in fixed context and more difficulties with context variability at both testing times.

4.3.6. Summary of the discrimination results

This section has presented the effects of training on vowel discrimination scores. The results from the pre- and post-test vowel AX discrimination task (DIS-task3) have demonstrated that the two training methods were successful in improving listeners' ability to distinguish contrasting English vowels. Vowel discrimination scores were significantly higher at post-test than at pre-test, for all vowel sets (3), vowel contrasts (16), and test conditions (fixed context and context variability). Trainees showed significant improvements in vowel discrimination (ID: 12.72%; ART : 15.09%). ID and ART types of training were similarly effective in improving vowel discrimination accuracy.

Four main concluding remarks can be made concerning improvement in vowel discrimination after training:

(1) This study revealed that short-term high-variability training (based on identification or imitation), not based on a discrimination task, significantly improved non-native listeners' vowel discrimination. The fact that the training included multiple talkers besides context variability in the stimuli might have a played a crucial role in discrimination improvement when discriminating vowels produced by untrained talkers in varied contexts.

(2) The effect of training on discrimination accuracy varied as a function of *vowel set*.

ID and ART training groups showed a similar pattern of discrimination results at pre-test and post-test: the back vowel contrasts (/ɒ/-/ɔː/, /ʊ/-/uː/, /ɔː/-/uː/) got the highest accuracy in discrimination, followed by high-front (/iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/, /ɪ/-/ɜː/), and low (/e/-/æ/,/ɜː/-/æ/, /æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɜː/-/ʌ/, /ɜː/-/ɑː/, /ɒ/-/ɑː/, /ʌ/-/ɒ/) contrasts (see Table 4.3.3) before and after training.

(3) The effect of training on discrimination accuracy varied as a function of *vowel contrast* Following ID and ART training, listeners still had great difficulty at post-test with contrasts such as /ɜː/-/ɑː/ and /ɒ/-/ɑː/ (<70%) and could only partially distinguish the /iː/-/ɪ/ contrast well. However, trainees performed at ceiling (>95%) for the high-front /ɪ/-/ɜː/ contrast and four low contrasts (/ɜː/-/æ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɜː/-/ʌ/). On average, listeners distinguished the other contrasts relatively well (80-90%): /ɪ/-/e/, /e/-/ɜː/, /e/-/æ/, /æ/-/ʌ/, /ʌ/-/ɒ/, /ɒ/-/ɔː/, /ʊ/-/uː/.

(4) The effect of training on discrimination varied as a function of the *test condition*: different contexts lowered discrimination performance.

Although ID and ART groups performed significantly better in fixed context and context variability conditions after training, they showed more accurate performance in fixed context and more difficulties with context variability at both testing times.

Overall, context variability makes it more difficult for some, but not all, listeners to resolve formant frequency as accurately as for vowels in isolation (Kewley-Port, 1995, 2001).

One explanation for the fact that listeners performed significantly better and obtained larger gains in fixed context than in context variation may be that vowel discrimination is affected by the nature of the surrounding phonetic context, and this could be due, in turn, to the fact that context interferes with perceptual memory.

As shown by results, listeners' ability to make reliable vowel quality distinctions is lowered when the vowels are placed in different consonant contexts rather than in a fixed /h*V*d/. Results demonstrate that after several days training experience with high-variability stimuli provided in training, there is significant improvement. However, there are still "context effects" in post-test discrimination accuracy and a closer analysis of the data shows that post-test performance decreased by 31.65% for ID and 30.03% for ART with context variance (as compared to fixed context). This happened because the ability to accurately discriminate vowels was degraded by the condition that provided learners with the vowels embedded in different consonantal contexts during the CVC testing blocks.

1.4. Summary of perception results

In order to test whether HVPT resulted in improvements in the perception of the L2 vowels, we compared pre-test and post-test performance on three perceptual tasks. Subjects were pre- and post-tested on the identification of natural AV and A stimuli (ID-I) and duration-manipulated A stimuli (ID-II), and on the discrimination of natural vowel contrasts (DIS). Stimuli included trained tokens as well as novel tokens produced by new talkers to test for generalization effects of training.

The participants' perceptual phonological competence was assessed by computing mean percent correct vowel identification scores (*percent correct*) for natural vowels (ID-I) for each subject, *type of training* (ID vs. ART), *word type* (trained vs. untrained), *stimulus presentation* (AV vs. A), *vowel set* (high-front, low and back) and *vowel*(/iː, ɪ, e, ɜː, æ, ʌ, ɑː, ɒ, ɔː, ʊ, uː/), at pre- and post-test (T1 vs. T2). The distribution of errors among response options was measured through the degree of *error dispersion* (score from 1 to 3 wrong responses).*Error dispersion* was used as an independent measure that quantifies the number of error categories per response.

In order to assess improvement in the identification of synthesized vowels(8 vowels with manipulated duration) (ID-II), and in particular L2 cue weighting after training, mean percent correct vowel identification scores (*percent correct*) were computed for each subject, *type of training* (ID vs. ART), *vowel contrast* (or vowel continua) and synthesized *vowel*(/i, ɪ, æ, ʌ, ɑ, ʊ, u/), and *manipulation of duration*(for every *vowel*) (step 1, 80; step 2, 116.7; step 3, 153.3; step 4, 190; step 5, 226.7; step 6, 263.3, step 7, 300 ms), at pre- and post-test (T1 vs. T2). The effect of training on learners' identification of synthesized English vowels (ID-II) was examined by assessing the degree of categoriality and sensitivity to changes in vowel duration along eight vowel continua. The robustness of vowel identification learning was assessed by testing the consistency with which listeners categorized vowels with duration ranging from 80 to 300ms. A*duration effect score* (DES) was computed understood as the difference in mean *percent correct* categorization at the two endpoints of the duration continua.

The participants' perceptual phonological competence was also assessed by computing *mean percent correct* of correctly discriminated vowel pairs or overall discrimination (*percent correct*) for each subject, *type of training* (ID vs. ART), *test condition* (fixedcontext vs. context variability),vowel set (Vset1: high & mid front vowels; Vset2: low vowels; Vset3: back vowels) and *vowel contrast* (/iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/, /ɪ/-/ɜː/; /e/-/æ/, /ɜː/-/æ/, /æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɜː/-/ʌ/, /ɒ/-/ɑː/, /ʌ/-/ɒ/; /ɒ/-/ɔː/, /ʊ/-/uː/, /ɔː/-/uː/), at pre- and post-test (T1 vs. T2). Mean percent *gains* were also obtained by subtracting pre-test from post-test scores.

Generally speaking, training effects on the perception of natural vowels showed that HVPT significantly improved English learners' identification performance. Results demonstrated that the two training methods, ID and ART training, significantly improved overall accuracy in the identification of natural vowels. The two training methods were successful in significantly improving learners' ability to identify English vowels. Besides, the findings point to a successful transfer of learning to the perception of the same vowel in new contexts (untrained words) by new talkers.

HVPT significantly improved English learners' identification performance about 14.87% (M_{T1}=64.80 vs. M_{T2}=79.67; $p<.05$) and learning generalized to all test conditions. Higher accuracy in natural vowel identification was obtained in AV condition (M_{T1}=66.36, M_{T2}=80.89; gains= 14.53) than in A condition (M_{T1}=62.48, M_{T2}=78.44; gains= 15.96). Trainees performed better with trained words(M_{T1}=66.25, M_{T2}=80.71; gains= 14.46) than novel words (M_{T1}=62.58, M_{T2}=78.63; gains= 16.05).

The effect of training was significant for all the English vowels, but a main effect of *vowel* was found on post-test identification performance. The highest percentage of identification gains was obtained for the high-back vowel /uː/ (gains: 14.87%), although high back vowels were the most confused vowels at post-test (/ʊ/: M_{T2}= 56.25; /uː/ M_{T2}=66.60). The best post-test identification performance was found for the mid vowels /e/ (M_{T2}=96.42) and /ɜː/ (M_{T2}=94.20), a notable improvement was seen for /ɑː/ (T2: M_{T2}=83.59; gains: 13.35%) and /ɪ/ (M_{T2}=80.21; gains: 17.45%), and only moderate improvement was reported for /ɔː/ (gains: 9.24%)and /ʊ/ (gains: 8.84%). Trainees improved considerably for /uː/ (gains: 19.86%), /ɜː/ (gains: 17.90%), /ɪ/ (gains: 17.45%) (average 18 percentage points) and improved about 12 percentage points for the rest (/æ/ gains: 12.37%), /ʌ/gains: 13.87%, /ɑː/gains: 13.35%).

Both ID and ART training methods significantly increased the ability to identify natural English vowels correctly. Interestingly, the findings confirmed that improvements in perception also resulted from the ART training, and that the "untrained" modality also benefitted significantly from ART training. This suggests that the transfer of production training to improved perception led a to a "tuning" of the corresponding perceptual representations. Similar results have been reported by Kartushinaet al. (2015). This finding does not conform to L2 training studies holding that the *type of training* only improves the "trained modality" whereas the "untrained modality" benefits little, if at all, from training (Akahane-Yamada et al. 1998; Bradlow et al. 1997; Lopez-Soto and Kewley-port, 2009).

When the effect of *type of training* on mean vowel identification was explored, no evidence of advantage for any training group was shown. In fact, no significant effect of *type of training* was found for overall natural vowel identification. However, there was a significant *type of training* x *word type* interaction, suggesting that the ID group outperformed the ART one for *trained* tokens. This meant that the ID group displayed a significantly higher decrease in the degree or dispersion of errors than the ART group per vowel examined. In contrast with the ART group, the ID group had significantly reduced the scope of mistaken "similar" categories (or errors) assigned to 6 of the 11 vowel stimuli as a response to training.

On the other hand, diverging patterns of vowel (mis-)identification were found for each training group with regards to *error dispersion*: the ID group obtained lower degree of dispersion of wrong responses (M_{T1}= 2.03; M_{T2}=1.35) than the ART group (M_{T1}=1.89; M_{T2}=1.55) after training. A significant effect of *type of training*was obtained at post-test for /iː/ (p= .005), /e/ (p= .004), /ɒ/ (p= .036), /ɔː/ (p= .036) and /uː/ (p= .023) error dispersion scores, indicating that the ID group showed more visible differences in error dispersion than the ART group after training. Apparently, the ID training was based on an identification task may have contributed to increasing the degree of perceived dissimilarity between the English vowels offered as possible responses better than the ART training. Another possible explanation for the fact that the ID training group showed less error dispersion than the ART group at post-test is the similarity between the categorization task used for training and testing.

As the results of theidentification of synthesized vowels(vowels with manipulated duration) (ID-II) confirmed,HVPT significantly improved trainees' consistency in labeling the English synthesized vowels (T1: 69.92, T2: 79.61; p<.05) and displayed lower DES after training. HVPT had a significant positive effect on the trainees' percent correct ID of duration-modified /i/-/ɪ/, /ʌ/-/ɑ/and /ʊ/-/u/ continua. However, the effect of *type of training*on mean percent correct identification of vowels with manipulated duration did not reach significance. This finding is especially interesting when taking into account that both ID and ART training methods were similarly effective in increasing attention to spectral cues for accurate identification of duration-manipulated vowels (ID: M_{T1}=70.94, SD= 8.59,

M_{T2}= 79.18, SD= 9.60; ART: M_{T1}= 68.91, SD= 18.53, M_{T2}= 80.04, SD= 17.91), especially duration-manipulated high-front vowels. A main effect of *vowel* was found on DESs, revealing that the effect of duration on vowel identification significantly differed according to the *vowel*.At post-test, the DESs obtained for /iː/ (ID: 41.8, ART: 33.5), /ɪ/ (ID: 30.3, ART: 21.3), /ʌ/ (ID: 59.4, ART: 56.8) and /ɑː/ (ID: 68.48, ART: 64.52) were significantly higher than those obtained for /æ/ (ID: 8.5; ART: 4.5) /ʌ/ (ID: 6.1; ART: 4.5) and /ʊ/-/uː/. The effect of *type of training* on DESs failed to reach significance at post-test, indicating that both ID and ART training groups underwent a similar significant decrease in DESs after 10 one-hour training sessions. These results indicate that HVPT may have facilitated the learners attending to the duration variability present in the stimuli. We fact that the type of training did not significantly determine the results paints a complex picture of L2 vowel perception. For instance, it is important to recognize that including variability in (auditory or articulatory) training is likely to contribute to more robust perceptual category development and the chances for generalization to a full range of contexts, talker characteristics and stimuli with varying duration.These findings suggest that L2 listeners may become better at using acoustic information through HVPT for L2 vowel categorization.

The results of the discrimination task confirmed that HVPT significantly improved English vowel discrimination (M_{T2}= 79.1) M_{T1}= 65.2), and the effect of training was generalized to all *vowel sets* (high-front, low, and back vowel pairs). Trainees improved from 77.2% to 93.7% in vowel discrimination in *fixed context*, but obtained worse post-test scores when they discriminated vowels in *context variability* (from 52% pretraining to 63% posttraining). Discrimination scores showed that high-back vowels (vowel set 3: /ɒ/-/ɔː/, /ʊ/-/uː/, /ɔː/-/uː/) were discriminated more accurately than high-front (vowel set 1: /iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/, /ɪ/-/ɜː/) and low vowels (vowel set 2: /e/-/æ/, /ɜː/-/æ/,

/æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɜː/-/ʌ/, /ɒ/-/ɑː/, /ʌ/-/ɒ/), both at pre-test and post-test, and low vowel pairs (vowel set 2) were discriminated worse than high pairs (vowel sets 1 and 3). The vowel contrasts /ɪ/-/ɜː/, /ɜː/-/æ/ and /æ/-/ɑː/ (81-86%) were by far the most accurately discriminated among the 16 vowel contrasts, whereas /ɜː/-/ɑː/and /ɒ/-/ɑː/ were discriminated the most poorly among all the vowel contrasts before (M_{T1}= 52-59%) and after training (M_{T2}= 65-70%). HVPT was effective in significantly improving the discrimination of the 16 English natural vowels and no evidence of advantage for any type of training was shown. When discrimination gains were compared across training groups, the ART group showed a higher percentage of gains for high-front (vowel set 1) and high-back (vowel set 3) vowel pairs (VS1: 18%, VS3: 19%) than the ID group (VS1: 13%, VS3: 14%). For the ID group, the highest amount of gains was obtained for low vowels, whereas the ART training group obtained the highest percentage of gains for back vowel pairs. In general, the ART group obtained a higher percentage of gains than the ID group in all vowel contrasts, except for VC11(/ɜː/-/ɑː/), VC12 (/ɒ/-/ɑː/), and VC13 (/ʌ/-/ɒ/), but between-group differences in mean percent gains turned out non-significant.

Contrary to other studies (Iverson & Evans 2009), both types of training had a significant effect on English vowel discrimination alongside an effect on vowel identification. Importantly, both types of training contributed to increase the degree of perceived dissimilarity among target vowel pairs and, consequently, to the formation of new L2 vowel categories.

Chapter 5

Results II: Production

The aim of this chapter is to present the results of the vowel production measurements at pre-test and post-test, with a view to estimating the size of changes in vowel articulation, particularly amount of spectral shift and duration changes after identification and articulatory HVPT. The results are presented in three sections corresponding to the analysis of three production measures at pre-test and post-test. Sections 5.1 presents the results for the research questions examining the effects of training (RQ 1.4.1) and the effects of type of training (RQ2.4a) on *spectral distances* (SD) (Flege, Bohn, & Jang, 1997) calculated for ten vowel pairs. Section 5.2 explores the effect of training (RQ1.4.2) and type of training (RQ2.4b) on *tense-lax duration ratios* (DR) estimated for six vowel pairs. Finally, section 5.3 examines the effect of training (RQ1.4.3) and type of training (RQ2.4c) on the following acoustic measures to better understand training-related changes in vowel production accuracy: *frontnes*s (Bark), *height* (Bark) and *duration* (ms).

Within each section we provide results on the effects of *training* (RQ1) and the effects of the *type of training* (RQ2) on the production measures:

Research Question 1 (RQ1): *Main effect of training* (T1 vs. T2)

Will AV HVPT lead to changes in the Catalan-Spanish speakers' production of English vowels? That is, is AV HVPT an effective method to improve L2 vowel production?

Research Question 2 (RQ2): *Effect of type of training* (ID vs. ART)

Will there be differential gains in L2 vowel production as a function of *type of training* (Identification vs. Articulatory training)?

Results are presented in the following order: (1) comparability of subject groups at pre-test in vowel production accuracy, (2) effects of *training* and *type of train*ing (ID vs. ART) on production accuracy, and (3) effect of training and stimulus variables on vowel production. For all statistical tests the alpha level was set at .05.

5.1. Spectral distance (SD) scores

The **Sub-RQ1.1** examines to what extent learners will improve in the production of 11 vowel monophthongs. Specifically, we investigated whether training altered the spectral distances (SDs) derived from 10 vowel contrasts (/iː/-/ɪ/,/ɪ/-/e/, /e/-/ɜː/, /æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɒ/-/ɑː/, /ʌ/-/ɒ/, /ɒ/-/ɔː/, /ʊ/-/uː/), by significantly raising or lowering the trainees' tongue position after training. This section presents the effects of training on mean *spectral distances* (RQ1.4.1) as well as the differential effects of identification (ID) and articulatory (ART) types of training on this measure (RQ2.4a).

The analysis was executed using a series of separate ANOVAs with SDs as the dependent variable, with effect of *training* (T1 vs. T2), *test condition* (fixed context vs. context variability), *vowel set* (high-front, low, high-back) and *vowel contrast* (10) as within-subjects factors each time, and *type of training* (ID vs. ART) as between-subjects factor.

5.1.1. Comparability of subject groups at pre-test

The ID training (*M*= 1.43, *SD*= 0.25), ART training (*M*= 1.39, *SD*= 0.20) and control (*M*= 1.33, *SD*= 0.21) groups had a similar range of SD scores at pre-test. A one-way ANOVA confirmed that the three groups did not significantly differ from one another at pre-test in SD scores ($F(2, 55)= 0.82$, $p= .447$), suggesting that the groups were indeed comparable in terms of mean SD scores before training (see Table 5.1).

Results confirmed that, prior to training, there were no significant differences between subject groups concerning vowel production accuracy (Figure 5.1).

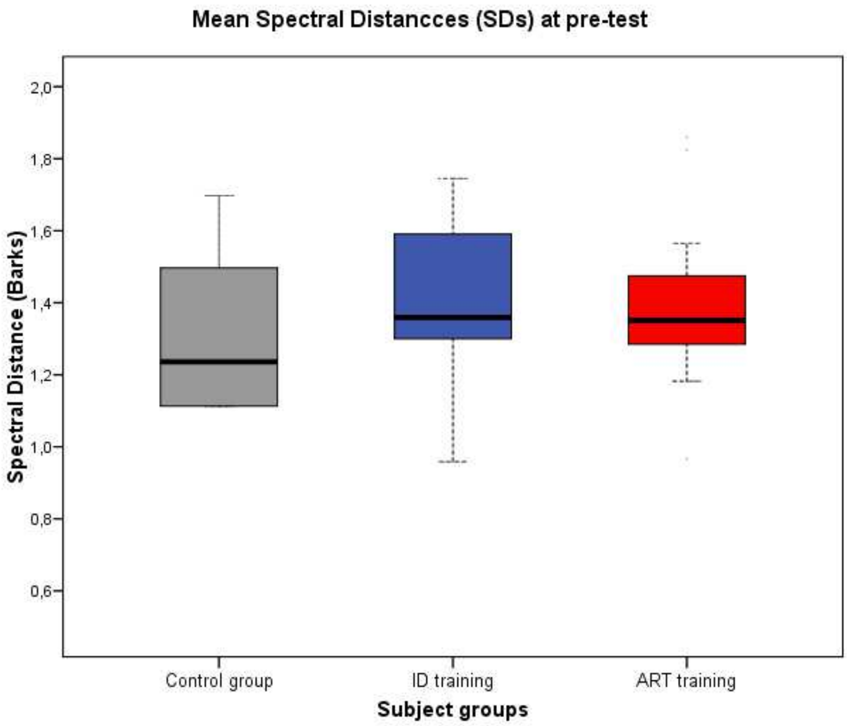


Figure 5.1. Boxplots of mean vowel spectral distances (SD) in vowel productions by ID training (blue), ART training (red), and NNS control (grey) groups at pre-test.

Spectral Distances (SD) at pre-test	Subject groups			
	ID training	ART training	NNS control	NS
Mean SDs	1.43 (0.25)	1.39 (0.20)	1.33 (0.21)	1.61 (0.45)
C1/iː/-/ɪ/	2.37 (0.49)	2.34 0.47	2.24 (0.30)	3.24 (0.53)
C2/ɪ/-/e/	2.10 (0.62)	2.30 (0.44)	2.24 (0.30)	2.89 (0.97)
C3/e/-/ɜː/	1.68 (0.64)	1.49 (0.39)	1.52 (0.84)	1.57 (0.93)
C4/æ/-/ʌ/	0.71 (0.46)	0.80 (0.44)	0.76 (0.23)	0.71 (0.35)
C5/æ/-/ɑː/	1.42 (0.69)	1.42 (0.58)	1.19 (0.37)	1.19 (0.49)
C6/ʌ/-/ɑː/	1.01 (0.50)	0.84 (0.29)	0.66 (0.25)	0.66 (0.47)
C7/ɒ/-/ɑː/	1.04 (0.61)	1.00 (0.59)	0.83 (0.48)	1.02 (0.52)
C8/ʌ/-/ɒ/	1.41 (0.37)	1.14 (0.41)	1.08 (0.36)	1.12 (0.27)
C9/ɒ/-/ɔː/	1.35 (0.70)	1.27 (0.43)	1.42 (0.35)	1.55 (0.98)
C10/ʊ/-/uː/	0.88 (0.38)	0.91 (0.40)	0.76 (0.49)	2.16 (0.58)

Table 5.1. Mean spectral distances (SD) (Bark) (and standard deviations) for the 10 vowel contrasts produced by ID training, ART training, NNS control and NS groups at pre-test.

We also tested for NNS-NS differences at pre-test (Figure 5.2). An ANOVA conducted on mean SD (averaged across all vowel distances) at pre-test with *subject group* (ID, ART, NS) as between-subjects factor did not show significant between-group differences ($F(2, 48)= 2.42$, $p=.100$) at pre-test.

However, as shown in Table 5.1, the descriptive data suggested that the SDs of high-front and high-back vowel contrasts were higher for the NS group than for the training groups, whereas the SDs of low vowel contrasts were lower for NSs than for the rest. To further explore differences between the training groups (ID and ART training) and the NS baseline group at pre-test, a mixed between-within ANOVA was run on spectral distances, with *subject group* (NNS vs. NS) as a between-subjects factor and *vowel contrast* (10) as within-subjects factor. The ANOVA revealed a significant main effect of *vowel contrast* ($F(9, 423)= 38.1$, $p< .001$, $\eta^2=.448$) and *subject group* ($F(1, 47)= 5.54$, $p=.038$, $\eta^2=.088$) on SDs. There was also a significant interaction between *vowel contrast* and *subject group* (F(9, 423)= 5.68, p<.001, η^2=.108), indicating that the differences in SD scores as a function of subject group were not consistent across all vowel contrasts. Pairwise comparisons showed that, at pre-test, NSs had significantly higher mean SD (averaged across the 10 vowel contrasts; M= 1.61) than the training groups (M= 1.37). Significant differences between training and NS groups were found for three high vowel contrasts: /iː/-/ɪ/, /ɪ/-/e/and /ʊ/-/uː/ (Table 5.2). Bonferroni-adjusted pairwise comparisons revealed that training groups produced significantly lower SDs than NS for /iː/-/ɪ/(p< .001), /ɪ/-/e/(p<. 05), and /ʊ/-/uː/ (p< .001) at pre-test, and higher SDs than NS for /æ/-/ɑː/ although this difference was not significant. The ID and ART groups showed slightly lower SD than NSs for the rest of the vowels, although differences did not reach significance. Overall, the amount of SD between vowels spoken by NS was greater than that of ID and ART groups.

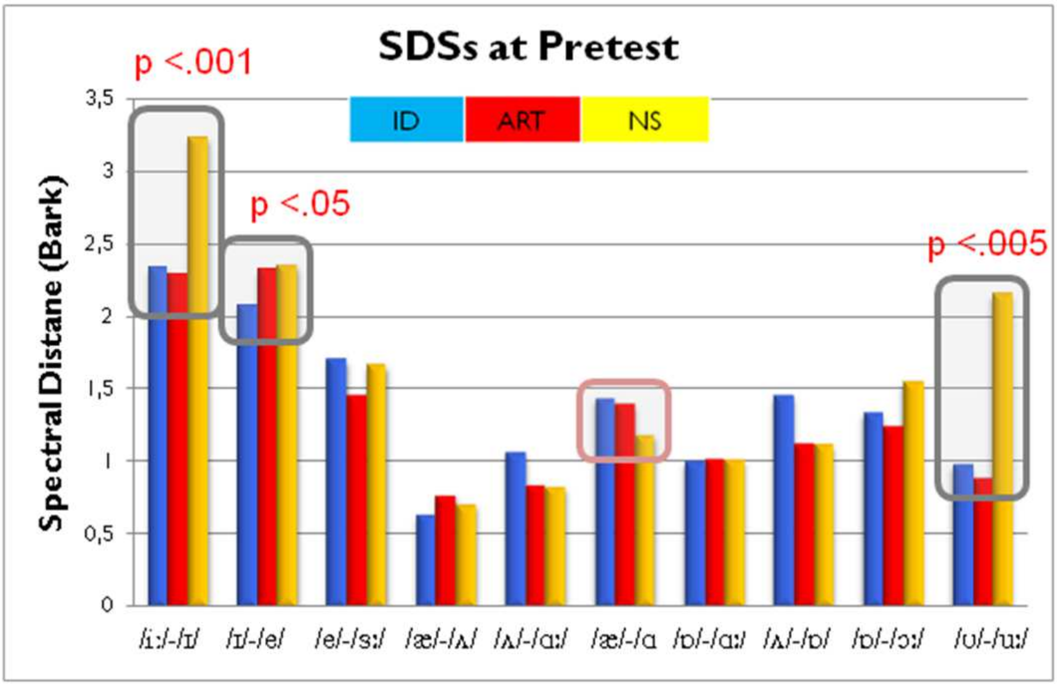


Figure 5.2. Mean spectral distances (SD) for the 10 vowel contrasts produced by ID training, ART training, NNS control and NS groups at pre-test, and significant NS-NNS differences (p< .05).

Vowel contrasts/spectral distances	One-way ANOVA between-group differences
/iː/-/ɪ/	$F(1, 64)= 18.91, p< .001$*
/iː/-/e/	$F(1, 64)= 8.02, p= .006$*
/e/-/ɜː/	$F(1, 64)= 0.28, p= .595$
/æ/-/ʌ/	$F(1, 64)= 0.00, p= .959$
/æ/-/ɑː/	$F(1, 64)= 1.09,p= .302$
/ʌ/-/ɑː/	$F(1, 64)= 3.55, p= .064$
/ɒ/-/ɑː/	$F(1, 64)= 0.03, p= .963$
/ʌ/-/ɒ/	$F(1, 64)= 0.30, p= .585$
/ɒ/-/ɔː/	$F(1, 64)= 1.09, p= .300$
/ʊ/-/uː/	$F(1, 64)= 21.42, p< .001$*

Table 5.2. ANOVA conducted on pre-test mean spectral distances (SD) computed for each vowel contrast (10) and each *subject group* (ID and ART training, andNS groups).

5.1.2. Effects of training and type of training on spectral distances in different test conditions (fixed context vs. context variability)

The production data were also analyzed to explore the effects of context variability on spectral distances and any interactions between the effects of training and the effects of context variability. In the following analyses we tested to what extent HVPT affected the production of English vowels in two different *test conditions* (fixed context vs. context variability), particularly the amount of spectral shift in each test condition (Table 5.3) (RQ1.4.1a). The study investigated which training method was better able to increase vowel production accuracy, by leading to a larger significant amount of spectral shift over the course of vowel production in each test condition. For the analyses of SDs, the within-subjects factors were: effect of *training* (T1 vs. T2), *test condition* (fixed context vs. context variability) and *vowel contrast* (10). The *type of training* (ID vs. ART) was the between-subjects factor.

As shown in Table 5.3 and Figures 5.3-5.4, mean SDs were higher in the fixed context (/h_d/) than in the context variability condition (CVC). In the fixed context, post-test SDs remained stable in the case of ID training but increased after ART training. In the context variability condition, post-test SDs slightly decreased with respect to pre-test for both ID and ART training.

Type of training	**Spectral Distances**								
	Test condition								
	Fixed context			Context variability			Mean SDs (all conditions)		
	T1	T2	gains	T1	T2	gains	T1	T2	gains
ID training (N=32)	1.65 (0.30)	2.02 (1.85)	0.37 (1.82)	1.50 (0.27)	**1.46 (0.18)**	**0.04 (0.20)**	1.44 (0.27)	1.41 (0.26)	0.03 (0.17)
ART training (N=32)	1.52 (0.22)	1.59 (0.25)	0.07 (0.26)	1.45 (0.16)	**1.45 (0.24)**	0.00 (0.16)	1.38 (0.20)	1.38 (0.21)	0.00 (0.20)

Table 5.3. Pre-test and post-test spectral distances (SDs and standard deviations of vowel productions in fixed context, context variability and all conditions, obtained by ID training and ART training groups.

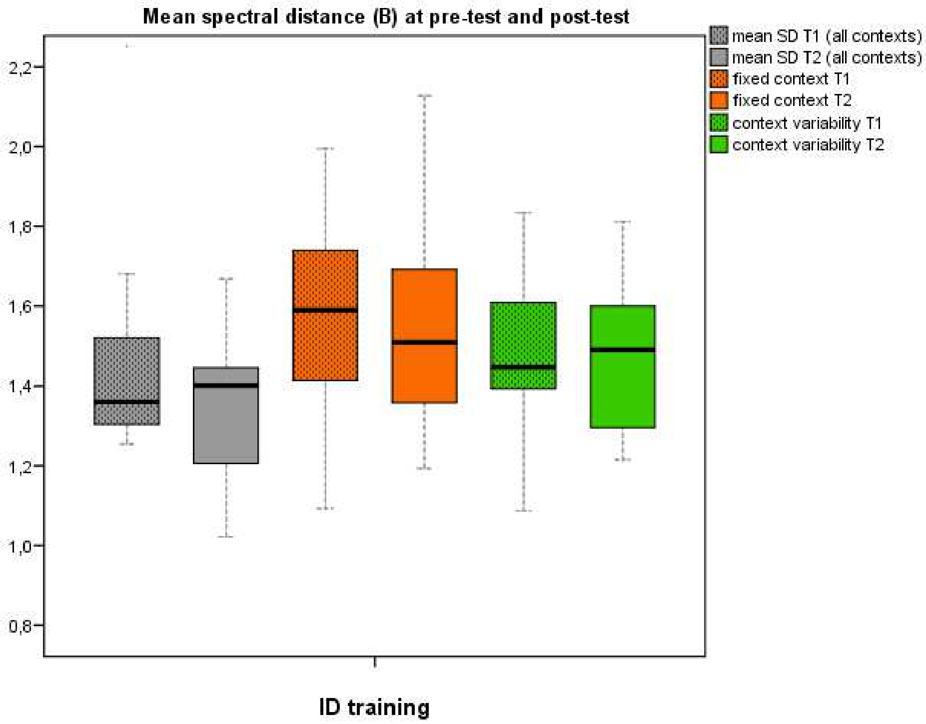


Figure 5.3. Pre-test and post-test spectral distances (SDs) of vowel productions in all conditions, fixed context, and context variability obtained by ID training group.

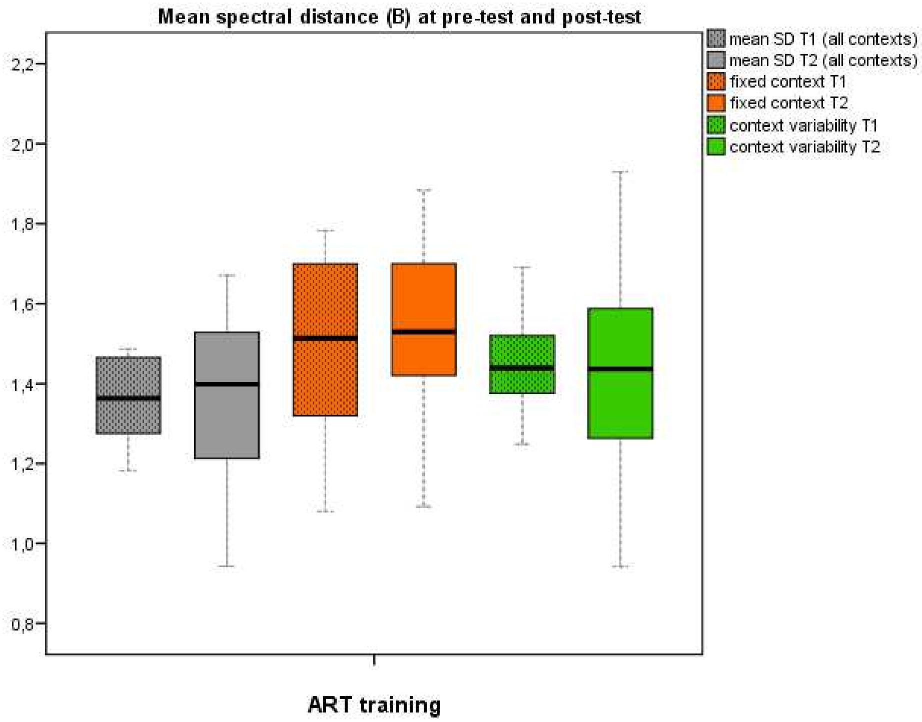


Figure 5.4. Pre-test and post-test spectral distances (SDs) of vowel productions in all conditions, fixed context, and context variability obtained by ART training groups

Types of training	Vowel Production	Vowel Spectral Distance scores (Bark)								
		All conditions			Fixed context /h_d/			Context variability (cvc)		
		T1	T2	Diff.	T1	T2	Diff.	T1	T2	Diff.
ID training	/iː/-/ɪ/	2.34 (0.47)	2.60 (0.71)	0.21 (0.71)	1.59 (0.68)	2.50 (2.54)	0.90 (2.73)	3.56 (0.63)	3.52 (0.70)	0.04 (0.44)
	/ɪ/-/e/	2.08 (0.63)	1.63 (0.51)	0.44 (0.31)	3.49 (1.02)	2.68 (0.86)	0.81 (0.85)	2.08 (0.39)	2.08 (0.40)	0.00 (0.48)
	/e/-/ɜː/	1.71 (0.66)	1.88 (0.58)	0.11 (0.51)	1.99 (0.84)	2.61 (1.63)	0.62 (1.61)	1.57 (0.48)	1.57 (0.56)	0.00 (0.48)
	/æ/-/ʌ/	0.64 (0.46)	0.79 (0.65)	0.10 (0.44)	1.11 (0.58)	1.60 (1.76)	0.49 (1.74)	0.75 (0.41)	0.81 (0.44)	0.06 (0.44)
	/ʌ/-/ɑː/	1.44 (0.68)	1.27 (0.80)	0.11 (0.41)	1.26 (1.62)	1.00 (0.69)	0.27 (1.73)	1.09 (0.47)	1.08 (0.49)	0.01 (0.52)
	/æ/-/ɑː/	1.07 (0.52)	0.88 (0.51)	0.05 (0.49)	1.74 (0.77)	2.21 (2.53)	0.47 (2.51)	1.31 (0.66)	1.14 (0.71)	0.16 (0.49)
	/ɒ/-/ɑː/	1.00 (0.57)	0.90 (0.47)	0.16 (0.52)	1.41 (0.80)	1.42 (1.76)	0.00 (1.83)	0.97 (0.50)	0.90 (0.44)	0.07 (0.56)
	/ʌ/-/ɒ/	1.46 (0.38)	1.41 (0.39)	0.05 (0.38)	1.58 (0.66)	1.86 (2.29)	0.28 (2.30)	1.47 (0.57)	1.45 (0.37)	0.02 (0.44)
	/ɒ/-/ɔː/	1.34 (0.78)	1.45 (0.86)	0.16 (0.47)	1.47 (0.85)	2.03 (1.92)	0.55 (1.71)	1.18 (0.75)	1.18 (0.57)	0.00 (0.45)
	/ʊ/-/uː/	0.98 (0.35)	0.86 (0.60)	0.15 (0.57)	1.00 (0.70)	1.48 (1.88)	0.48 (1.98)	1.21 (0.43)	0.99 (0.59)	0.21 (0.64)
	Mean SD	1.44 (0.27)	1.41 (0.26)	0.03 (0.17)	1.65 (0.30)	2.02 (1.85)	0.37 (1.82)	1.50 (0.27)	**1.46 (0.18)**	**0.04 (0.20)**
ART training	/iː/-/ɪ/	2.31 (0.47)	2.61 (0.60)	0.27 (0.48)	1.49 (0.76)	2.00 (1.02)	0.51 (0.67)	3.41 (0.51)	3.49 (0.59)	0.08 (0.32)
	/ɪ/-/e/	2.33 (0.43)	1.90 (0.33)	0.42 (0.48)	3.62 (0.78)	2.68 (0.77)	0.94 (0.72)	2.06 (0.50)	2.14 (0.34)	0.07 (0.52)
	/e/-/ɜː/	1.47 (0.39)	1.77 (0.36)	0.23 (0.42)	1.71 (0.59)	2.19 (0.56)	0.48 (0.55)	1.52 (0.38)	1.59 (0.35)	0.07 (0.39)
	/æ/-/ʌ/	0.77 (0.43)	0.69 (0.41)	0.11 (0.47)	1.45 (0.53)	1.40 (0.51)	0.04 (0.63)	0.74 (0.38)	0.73 (0.43)	0.00 (0.39)
	/ʌ/-/ɑː/	1.40 (0.59)	1.27 (0.48)	0.19 (0.42)	0.74 (0.48)	0.82 (0.46)	0.07 (0.57)	1.11 (0.35)	0.97 (0.46)	0.14 (0.48)
	/æ/-/ɑː/	0.84 (0.29)	0.78 (0.42)	0.19 (0.46)	1.96 (0.65)	1.85 (0.65)	0.11 (0.69)	1.31 (0.62)	1.06 (0.40)	0.24 (0.57)
	/ɒ/-/ɑː/	1.02 (0.59)	1.15 (0.59)	0.07 (0.55)	0.99 (0.65)	1.11 (0.70)	0.13 (0.76)	0.94 (0.51)	0.93 (0.53)	0.00 (0.49)
	/ʌ/-/ɒ/	1.13 (0.42)	1.24 (0.57)	0.10 (0.44)	1.16 (0.60)	1.33 (0.60)	0.17 (0.58)	1.28 (0.44)	1.28 (0.54)	0.00 (0.43)
	/ɒ/-/ɔː/	1.25 (0.44)	1.26 (0.47)	0.09 (0.71)	1.47 (0.69)	1.37 (0.51)	0.10 (0.85)	1.35 (0.76)	1.25 (0.45)	0.10 (0.83)
	/ʊ/-/uː/	0.89 (0.41)	0.80 (0.36)	0.07 (0.41)	0.92 (0.49)	1.12 (0.56)	0.20 (0.78)	1.25 (0.55)	**1.00 (0.59)**	0.25 (0.52)
	Mean SD	1.38 (0.20)	1.38 (0.21)	0.00 (0.20)	1.52 (0.22)	1.59 (0.25)	0.07 (0.26)	1.45 (0.16)	**1.45 (0.24)**	0.00 (0.16)

Table 5.4. Mean spectral distance scores (Bark) (and standard deviations) for the 10 vowel contrasts in fixed context and context variability conditions, produced by ID and ART training groups at pre-test and post-test.

First, an ANOVA run on mean SD scores (averaged across all vowel distances) with *test condition* (fixed vs. variability) and effect of *training* (T1 vs. T2) as within-subjects factors, and *type of training* (ID vs. ART) as between-subjects factor, yielded a significant effect of *training* (F(1, 30)= 12.65, p=.001, η^2= .297). The effect of the *test condition* (F(1, 30)= 0.01, p=.929, η^2= .000) and *type of training* (F(1, 30)= 1.08, p=.306, η^2= .035) were not statistically significant. None of the interactions reached significance: *test condition x type of training* (F(1, 30)= 0.388, p=.544, η^2= .012), *effect of training x type of training* (F(1, 30)= 0.24, p=.628, η^2= .008), *test condition x effect of training* (F(1, 30)= 0.33, p=.568, η^2= .011), *test condition x training x type of training* (F(1, 30)= 0.06, p=.801, η^2= .002). These results indicated, on the one hand, that there was a significant difference between pre-test and post-test SDs in fixed context and context variability conditions. On the other hand, these meant that SDs did not differ significantly across test conditions, and the ID and ART training groups obtained similar SDs in each test condition, at pre-test and post-test.

The descriptive data showed that the largest SD changes after ID and ART training were observed for high-front vowel contrasts, so it seems that HVPT had a larger impact upon high-front vowels as compared to low and high-back contrasts. A closer inspection of the descriptive data indicates that, in the fixed condition, both ID and ART groups exhibited a visible change in the SDs of the high-front vowels: whereas the SD between /iː/ and /ɪ/ increased at post-test, the SD between /ɪ/ and /e/ was consequently reduced, and the SD between /e/ and /ɜː/ thus decreased (see Table 5.4).

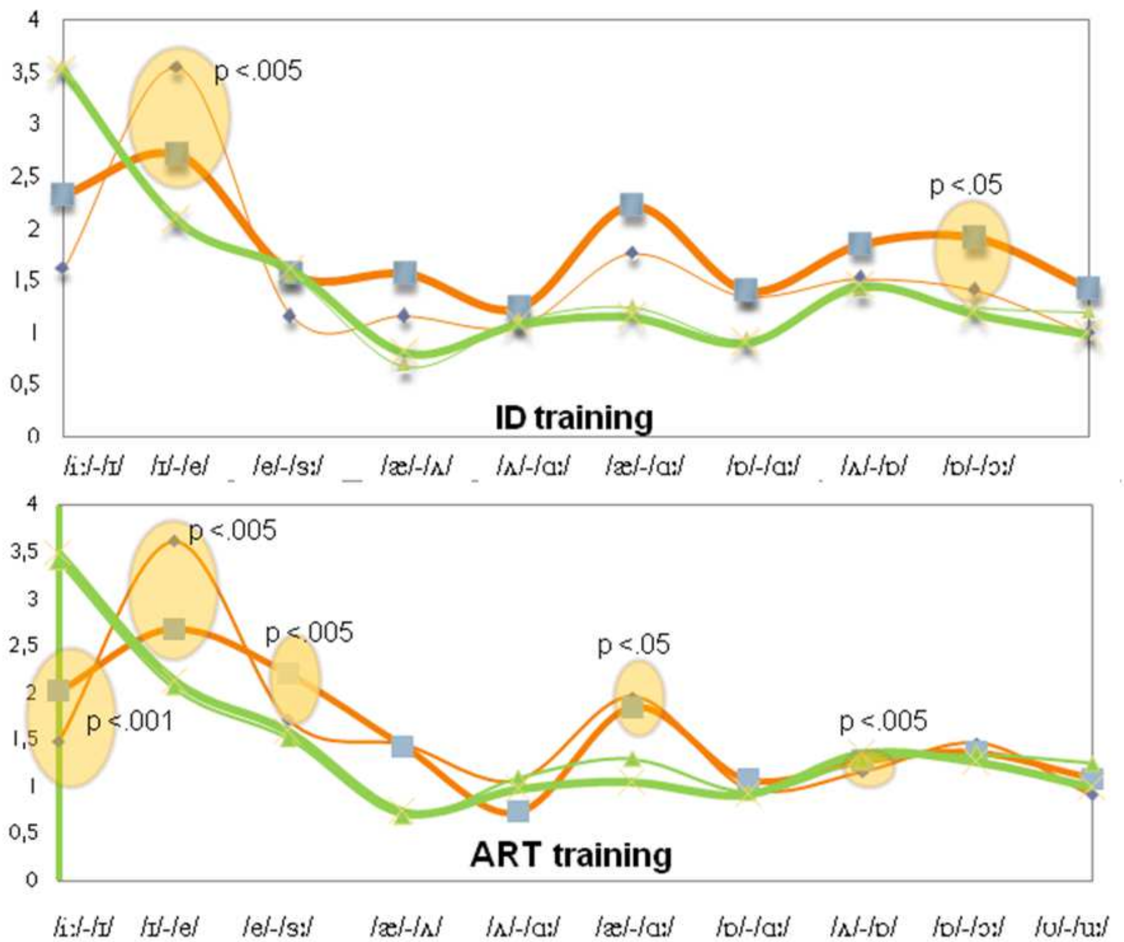
In order to further explore differences between the vowel contrasts after training, separate ANOVAs, with *vowel contrast* (10) as within-subjects factor and *type of training* (ID vs. ART) as between-subjects factor, were conducted for the fixed context and context variability conditions.

In both fixed and context variability conditions, the ANOVAs yielded a significant effect of *vowel contrast* (*fixed context:* F(9, 360)= 42.75, p<. 001, η^2= .517; *context variability:*

F(9, 342)= 144.68, $p<.001$, $\eta^2= .792$) and a significant interaction between *training* and *vowel contrast* (*fixed context:* *F*(9, 360)= 9.93, $p<.001$, $\eta^2= .199$; *context variability:* *F*(9, 342)= 2.24, $p<.05$, $\eta^2= .056$). Other interactions and the main effect of *type of training* (*fixed context:* *F*(1, 40)= 1.70, $p=.200$, $\eta^2= .041$; *context variability:* F(1, 40)= 0.25, p=.620, $\eta^2= .007$) were not significant. The training x vowel contrast interaction confirmed that the effect of training on SDs varied significantly as a function of *vowel contrast* in fixed context and context variability.

However, in the fixed context condition, there was also a significant *training x training type* interaction (*F*(1, 38)= .59, $p=.008$, $\eta^2=.314$), indicating that training in fixed context was only effective for one of the training groups. In fact, ID and ART training groups underwent significantly different changes in SDs in from pre-test to post-test when they produced vowels embedded in fixed /h_d/ context (Table 5.3).

To further explore interaction between training and type of training in the fixed context and context variability condition, paired-samples t-tests were run on SD scores for each training group. In the fixed context condition, the t-tests revealed significant SD changes after ART training for five vowel contrasts: /iː/-/ɪ/ (t(29)=-3.56, $p= .002$), /ɪ/-/e/ (t(29)=7.02, $p< .001$) and /e/-/ɜː/ (t(29)=-4.71, $p< .001$), /æ/-/ɑː/ (t(29)=2.71, $p< .05$) and /ʌ/-/ɒ/ (t(29)=2.66, $p< .05$) (see Figure 5.6). The t-tests revealed, however, that after ID training only the front vowel contrast /ɪ/-/e/ (t(29)=4.73, $p< .001$) had changed significantly from pre-test to post-test in the fixed context condition, as the SD between /ɪ/ and /e/was significantly shortened after training. In the context variability conditions, there were only significant SD changes for /æ/-/ɑː/ (t(29)=2.71, $p< .05$) after ART training, but none of the differences proved significant in context variability after ID training (Figures 5.5-5.6).



Figures 5.5-5.6. Line graph showing spectral distance scores of ten vowel contrasts produced by ID and ART training groups in the fixed context (orange) and context variability (green) conditions, at pre-test (thinner lines) and pos-test (thicker lines).

Summary

In summary, the fixed context conditioned offered a more solid ground for observing the changes in SDs from pre-test to post-test than the context variability condition. Both ID and ART training groups exhibited similar spectral distance changes for the vowel pairs /iː/-/ɪ/, /ɪ/-/e/ and /e/-/ɜː/ in the fixed context condition. At post-test, longer distances for /iː/-/ɪ/ and /e/-/ɜː/ appear to be compensated by shorter distances between /ɪ/ and/e/. Although these changes in spectral distances take place after both types of training in the form of a "compensatory" mechanism, the amount of significant SD changes.

5.1.3. Effects of training and type of training on spectral distances by vowel set

The interest in the analysis of SDs by vowel sets lies in the fact that our HVPT was based on the use of tree vowel sets (high-front, low, high-back) chosen based on difficulty (Nishi & Kewley-Port, 2007). One of the research questions was whether SD changes would vary as a function of *vowel set* (RQ1.4.1b). RQ2.4a explored which type of training led to more positive effects on spectral distances by vowel sets.

The control group did not show any changes in SDs for low vowels, and the SDs of high-front and back vowels slightly decreased at post-test. However, ID and ART training groups showed changes in SD for certain vowel contrasts after training (i.e. SD increased or decreased), suggesting that both training methods were effective in leading to formant movement for some vowels. Overall, ID and ART training groups differed very little from each other with respect to SD changes from pre-test to post-test.

Table 5.1 illustrates the normalized mean SD values for the three vowel sets (high-front, low and back) produced by each of the training groups (ID vs. ART) at pre-test and post-test (T1 vs. T2). Overall, trainees showed higher SDs for high-front vowels than for low and high-back vowels, at pre-test and post-test. The smallest SDs were found for low vowels. Whereas SDs of high-front vowels increased after training for the training groups, of the SDS of low and back vowels decreased at post-test.

Vowel sets	**Spectral Distance scores (Bark)**									
	ID training (*N*=32)			**ART training (*N*=32)**			**Control group (*N*=20)**			**NS (*N*=10)**
	T1	T2	Diff.	T1	T2	Diff.	T1	T2	Diff.	
Vowel set 1 High-front	2.06 (0.27)	2.03 (0.31)	0.03 (0.26)	2.04 (0.22)	2.09 (0.17)	0.06 (0.26)	2.01 (0.38)	1.95 (0.40)	0.06 (0.27)	2.33
Vowel set 2 Low	1.08 (0.29)	1.05 (0.32)	0.03 (0.23)	1.07 (0.31)	1.03 (0.32)	0.04 (0.27)	0.86 (0.20)	0.86 (0.20)	0.00 (0.26)	0.69
Vowel set 3 Back	1.13 (0.47)	1.13 (0.44)	0.00 (0.3)	1.09 (0.31)	1.06 (0.32)	0.03 (0.33)	1.09 (0.20)	1.06 (0.20)	0.03 (0.21)	1.32

Table 5.5. Pre-test and post-test spectral distance scores (Bark) and standard deviations in vowel production for the 3 vowel sets produced by ID training, ART training NS groups.

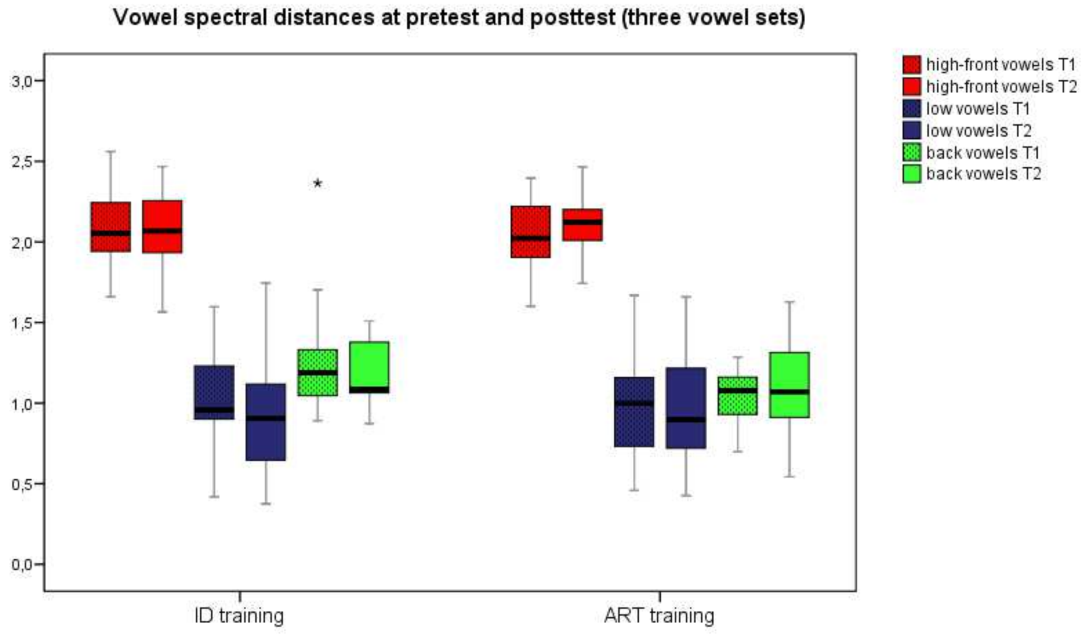


Figure 5.7. Boxplots of pre-test and post-test spectral distances (Bark) in vowel production for the 3 vowel sets produced by ID and ART training groups.

A three-way ANOVA with effect of *training* (T1 vs.T2) and *vowel set* (high-front, low, back) as within-subjects factors, and *training type* (ID vs. ART) as between-subjects factor, was run on SD scores to determine if the *training* had been equally effective in leading to changes in SDs of three vowel sets (high-front: /iː/-/ɪ/,/ɪ/-/e/, /e/-/ɜː/; low/æ/-/ʌ/. /ʌ/-/ɑː/, /æ/-/ɑː/, /ɒ/-/ɑː/; back: /ʌ/-/ɒ/, /ɒ/-/ɔː/, /ʊ/-/uː/) (RQ1.4.1b) and whether differences in production could be due to *type of training* (RQ2.4a).

The results revealed a significant main effect of *vowel set* ($F(2, 70)= 169.4$, $p<.001$, $\eta^2=.829$) but the main effects of *training* ($F(1, 35)= 0.13$, $p=.723$, $\eta^2=.004$) and *type of training* ($F(1, 35)= 0.42$, $p=.519$, $\eta^2=.012$) were not significant. None of the interactions reached statistical significance. These results indicated that the SDs differed as a function of *vowel set* at pre-test and post-test regardless of the type of training. The SD was significantly higher for high-vowels (p< .001) (/iː/-/ɪ/,/ɪ/-/e/, /e/-/ɜː/) than for low and high-back vowels, both at pre-test and post-test (see Table 5.5), suggesting that the degree of spectral overlap in vowel production was higher for low and back vowels than for high-front vowels.

SD changes were higher for /ɪ/-/e/than for the other high-front contrasts. Within the low vowel subset, changes in SD for/æ/-/ɑː/ were significantly greater for the ART group than for the ID training group. The amount of SD changes obtained for high-front and low vowels were more salient for the ART group than for the ID group (Figure 5. 5). For both training groups, the SD between high-front vowels was significantly larger than the SD obtained for back and low vowels. The SD between low vowels was the shortest, both at pre-test and post-test.

There were no significant effects of *training type* on SDs, suggesting that distance in place of articulation for vowel articulation was not significantly affected by the type of training received and that both types of training led to similar patterns in vowel production: the SD increased for high-front vowels and decreased for low and back vowels from pre-test to post-test. However, paired-samples t-tests conducted on SD changes of the three vowel sets revealed that these differences were significant for the three vowel sets for the ART group, whereas none of the pre-test/post-test differences reached significance for the ID group (see Table 5.6).

Vowel Identification	Paired comparisons			
	ID training (*N*=32)		**ART training (*N*=32)**	
	df t p-value	**gains**	**df t p-value**	**gains**
Vowel set 1 High-front	$t(19)=4.34, p< .590$	15.36 (19.87)	$t(30)=3.24, p< .005$*	12.37 (21.70)
Vowel set 2 Low	$t(20)=5.46, p< .499$	21.35 (23.47)	$t(30)=2.72, p< .05$*	13.02 (25.53)
Vowel set 3 Back	$t(20)=6.38, p< .949$	22.40 (20.13)	$t(30)=4.99, p< .001$*	17.32 (19.20)

Table 5.6. Results of paired-samples t-tests on the pre-test and post-test spectral distance scores in vowel production of 3 vowel sets produced by ID and ART training groups.

Summary

The amount of spectral changes from pre-test to post-test varied significantly as a function of the target vowel contrast. After training, higher spectral distances (SDs) were obtained between high-front vowels than between low and between high-back vowels. Within the high-front vowel area, the SDs increased for and /e/-/ɜː/ but decreased for /ɪ/-/e/ after HVPT. In the low and back vowel areas of the mouth, SD remained somehow steady for /æ/-/ʌ/, /ʌ/-/ɑː/, /ɒ/-/ɑː/, /ʌ/-/ɒ/, /ɒ/-/ɔː/ and decreased for /æ/-/ɑː/ and /ʊ/-/uː/. The results of paired-samples t-tests revealed that the three vowel sets underwent significant spectral distance changes only after ART training (p< .05) but not after ID training.

5.1.4. Effects of training and type of training on spectral distances by vowel contrast

One of the research questions for L2 vowel production was whether spectral distance changes would as a function of *vowel* (RQ1.4.1c). RQ2.4a explored which type of training led to more positive effects on spectral distances by vowel contrast.

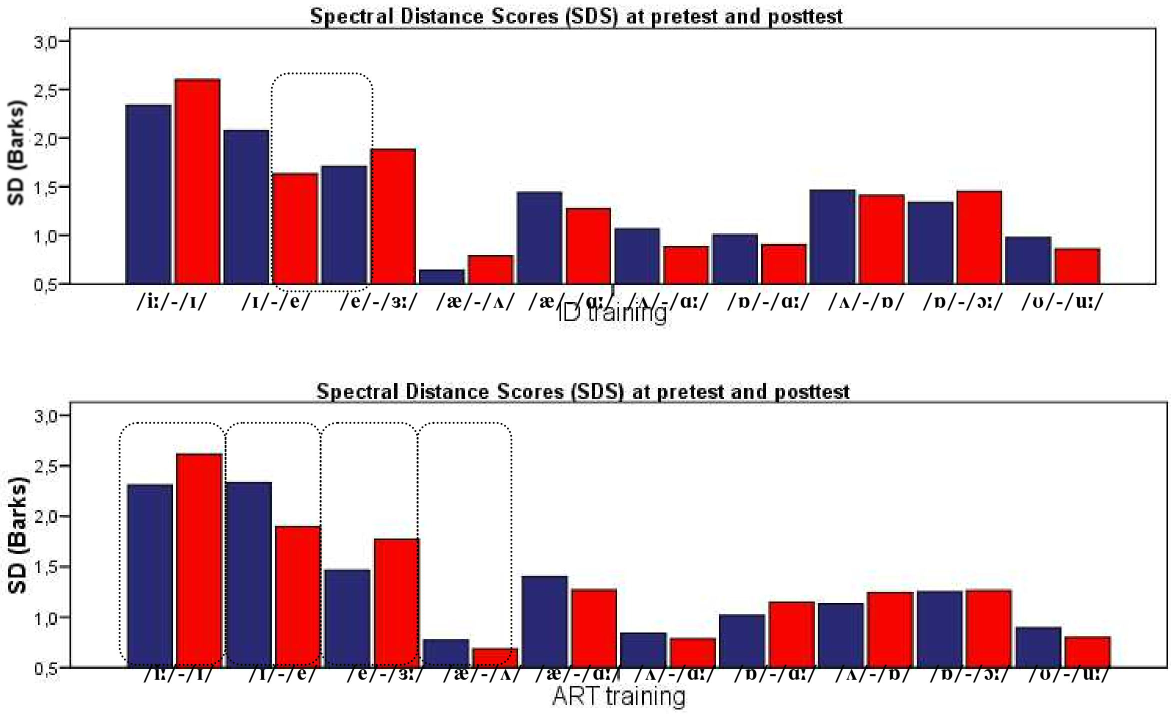
For the control group, SDs did not suffer major changes overall ($p > .05$). A closer inspection of the data revealed that the controls produced the /ɪ/-/e/(t(19)=7.45, $p = .172$) and /ʊ/-/uː/ (t(19)=9.45, $p = .232$) contrasts at post-test with a lower SD in relation to pre-test, and /e/-/ɜː/ (t(19)=-3.72, $p = .072$) were more distant from each other at post-test but, as expected, paired-samples t-tests did not yield statistically significant differences for any vowel contrast (p >.05) for the group who had not received any treatment.

Pre-test and post-test SDs were examined for the 10 vowel contrasts produced by ID and ART training groups (see previous Table 5.4) and three observations were made:

(1) First, trainees showed higher SDs for 6 vowel contrasts (/iː/-/ɪ/, /e/-/ɜː/, /æ/-/ʌ/, /ɒ/-/ɑː/, /ʌ/-/ɒ/, /ɒ/-/ɔː/) and lower SD for the rest (/ɪ/-/e/, /ʌ/-/ɑː/, /æ/-/ɑː/ and /ʊ/-/uː/) after training (Figures 5.8-5.9).

(2) Second, they exhibited a visible change in the SDs of the high-front vowels: whereas the SD between /iː/ and /ɪ/ increased at post-test, the SD between /ɪ/ and /e/ was consequently reduced.

(3) Third, in the low vowel area, they showed higher SD between /æ/ and /ʌ/ at post-test, which led to shorter SD for the other low vowel pairs, /ʌ/-/ɑː/ and /æ/-/ɑː/. SD between back target vowels decreased after training.



Figures 5.8-5.9. Pre-test and post-test spectral distances (Bark) in vowel productions for ID and ART training groups.

In order to explore the SDs of individual vowel contrasts in fixed context and context variability conditions, of a series of repeated-measures ANOVAs were conducted on SD with *type of training* (ID vs. ART) as a between-subjects factor, and *vowel contrast (*10), *effect of training* (T1 vs. T2) and *test condition* (fixed context vs. context variability) as within-subjects factors.

The results yielded a significant effect of *training* for four vowel contrasts: /iː/-/ɪ/ (F(1, 41)= 133.16, p< .001, η^2=.765), /ɪ/-/e/(F(1, 49)= 95.50, p< .001, η^2=.661), /e/-

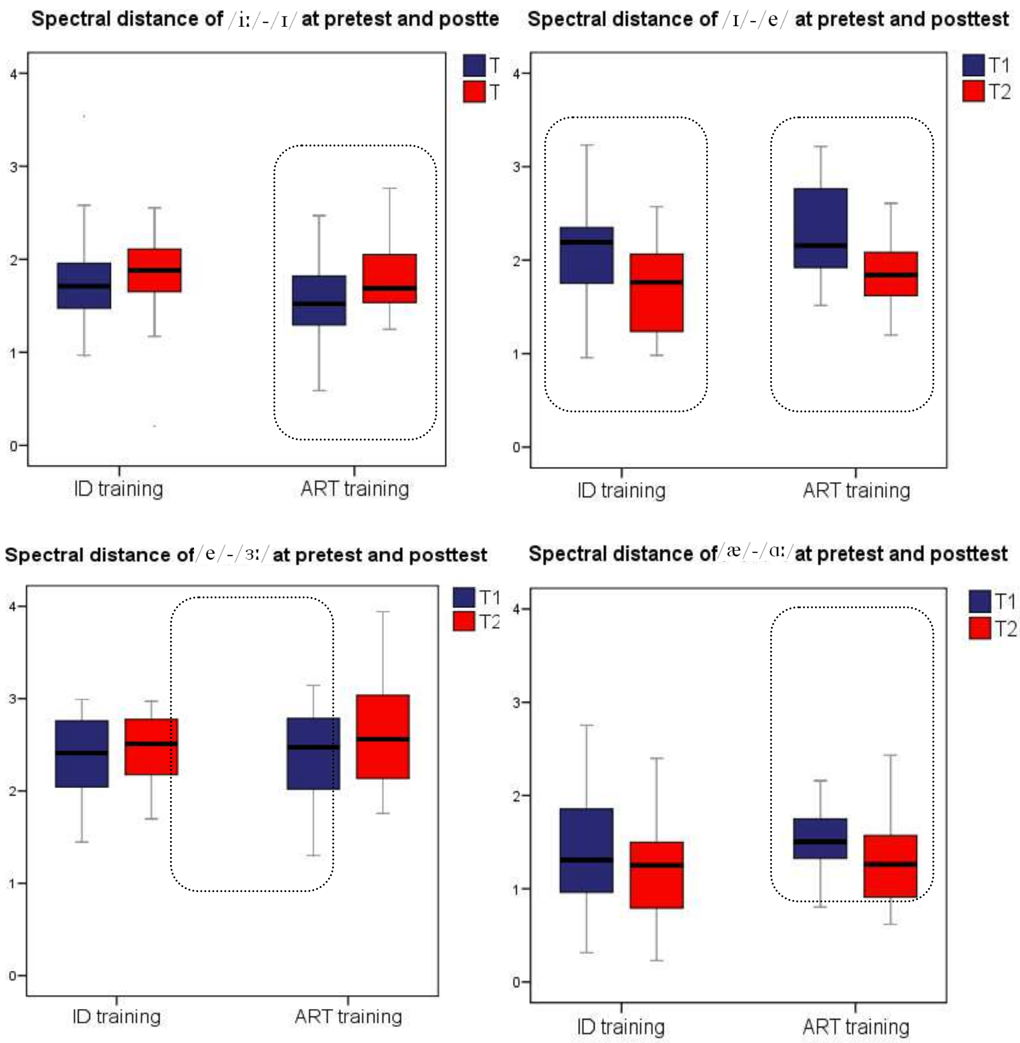
/ɜː/ $F(1, 50) = 35.33$, $p < .001$, $\eta^2 = .414$) and /æ/-/ɑː/ $F(1, 51) = 78.04$, $p < .001$, $\eta^2 = .605$). There was a significant effect of *test condition* for the same: /iː/-/ɪ/ ($F(1, 41) = 16.39$, $p < .001$, $\eta^2 = .286$), /ɪ/-/e/ ($F(1, 49) = 46.38$, $p < .001$, $\eta^2 = .486$), /e/-/ɜː/ $F(1, 50) = 14.75$, $p < .001$, $\eta^2 = .228$) and /æ/-/ɑː/ $F(1, 51) = 4.53$, $p < .05$, $\eta^2 = .082$). There was a significant interaction between *training* and *condition* for the high-front vowel contrasts /iː/-/ɪ/ ($F(1, 41) = 11.94$, $p = .001$, $\eta^2 = .726$), /ɪ/-/e/ ($F(1, 49) = 59.23$, $p < .001$, $\eta^2 = .547$) and /e/-/ɜː/ ($F(1, 50) = 16.30$, $p < .001$, $\eta^2 = .246$), but not for /æ/-/ɑː/ ($F(1, 51) = 1.63$, $p = .207$, $\eta^2 = .031$), suggesting that the impact of training on SD of these three contrasts depended on the test condition observed.

The training had a significant impact on the SD of the three high-front vowel contrasts which varied as a function of the test condition. For all high-front vowel contrasts a significant *training x test condition* indicated that SD changes were greater for /iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/ in the fixed context condition than in context variability (Figure 5.14).

- In the *fixed context* condition, the SD between /iː/ and /ɪ/ increased significantly after ID and ART training. However, ID and ART groups showed opposite behaviours in context variability: the SD between /iː/ and /ɪ/ increased after ART training and decreased after ID training.
- The SD between /ɪ/ and /e/ decreased significantly in all conditions, especially in *fixed context*, for both ID and ART training.
- The SD between /e/ and /ɜː/ increased significantly, especially in the *fixed context* condition, after both ID and ART training.

The effect of *type of training* did not reach significance for any of these vowel contrasts: /iː/-/ɪ/ ($F(1, 41) = 0.08$, $p = .776$, $\eta^2 = .002$), /ɪ/-/e/ ($F(1, 49) = 0.47$, $p = .497$, $\eta^2 = .009$), /e/-

/ɜː/$F(1, 50) = 35.33$, $p < .001$, $\eta^2 = .414$). This meant that ID and ART did not differ significantly in the SD obtained at post-test for any of these vowel contrasts.



Figures 5.10-5.13. Boxplots of pre-test and post-test vowel spectral distances (Bark) for /ɪː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/ and /æ/-/ɑː/ produced by ID and ART training groups, and significant pre-test/post-test differences ($p < .05$).

The ANOVAS run on the different low vowel contrasts confirmed there was a significant effect of *training* on /æ/-/ʌ/($F(1, 48) = 199.03$, $p < .001$, $\eta^2 = .806$), /ʌ/-/ɑː/ ($F(1, 48) = 8.85$, $p < .05$, $\eta^2 = .156$), /æ/-/ɑː/($F(1, 51) = 78.04$, $p < .001$, $\eta^2 = .605$) and /ɑː/-/ɒ/ ($F(1, 49) = 10.14$, $p < .05$, $\eta^2 = .171$). The effect of *condition* was significant for /æ/-/ɑː/($F(1, 51) = 4.53$, $p < .05$, $\eta^2 = .002$), but did not reach significance for the rest of contrasts. There was a significant *training × type of training* interaction for /æ/-/ʌ/(F(1, 48)= 7.74, p< .05, η^2=.139) and a significant *training × condition × type of training* interaction for /ɑː/-/ɒ/ (F(1, 49)= 5.36p< .05, η^2=.099). The effect of *type of training* did not reach significance for any of the low vowel contrasts: /æ/-/ɑː/ (F(1, 51)= 0.94, p= .38, η^2=.018). This meant that ID and ART did not differ significantly in the SD obtained at post-test for /æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/ or /ɑː/-/ɒ/. These results showed the following SD changes for low vowels after training:

- The two-way interaction found for /æ/-/ʌ/ suggested that the effect of training varied as a consequence of the type of training received. In the fixed context and context variability conditions, SDs between /æ/ and /ʌ/ increased after ID training but decreased after ART training.
- After ID and ART training, the effect of training on /ʌ/-/ɑː/ was different across the two test conditions. The SD between /ʌ/ and /ɑː/ decreased in fixed context after both ID and ART training, but increased in context variability, especially after ART training.
- For the low /æ/-/ɑː/, there was a significant decrease in SD from pre-test to post-test, especially in context variability, for both ID and ART groups. Both at pre-test and post-test, the SDs were significantly larger in the fixed context than in the variability condition (Figure 5.15).
- The three-way interaction found for /ɑː/-/ɒ/ indicated that the impact of training on each of the test conditions varied significantly as a result of the type of training. After ID training, SDs between /ɑː/ and /ɒ/decreased in the two conditions. However, SDs between/ɑː/ and /ɒ/ increased after ART training, especially in the fixed context condition and remained steady in context variability. ID and ART training groups showed opposite behaviours with respect to this contrast. At post-

test, the ID group obtained higher SDs for /ɑː/-/ɒ/ in fixed context and in context variability, whereas higher SDs were found in context variability for the ART group.

The effect of *type of training* did not reach significance for any of these vowel contrasts and the interactions involving type of training only reached significance for /ɑː/-/ɒ/. For the other low contrasts, differences between ID and ART in SDs were not considered significant.

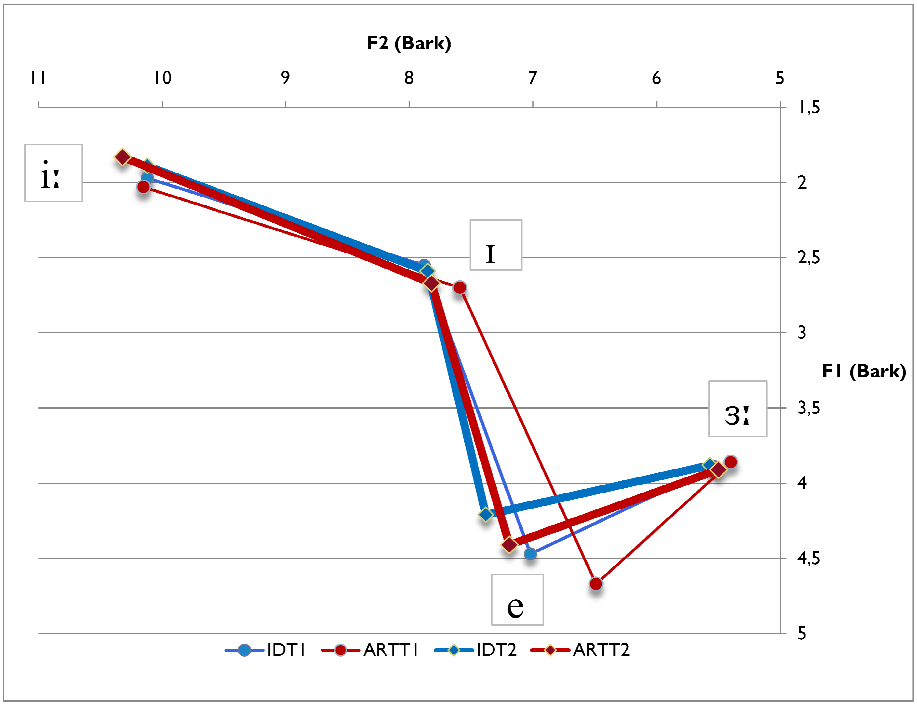


Figure 5.14. Spectral distances of the high-front vowel pairs produced by the ID training group (blue lines) and the ART training group (red lines) at pre-test (thinner lines) and post-test (thicker lines).

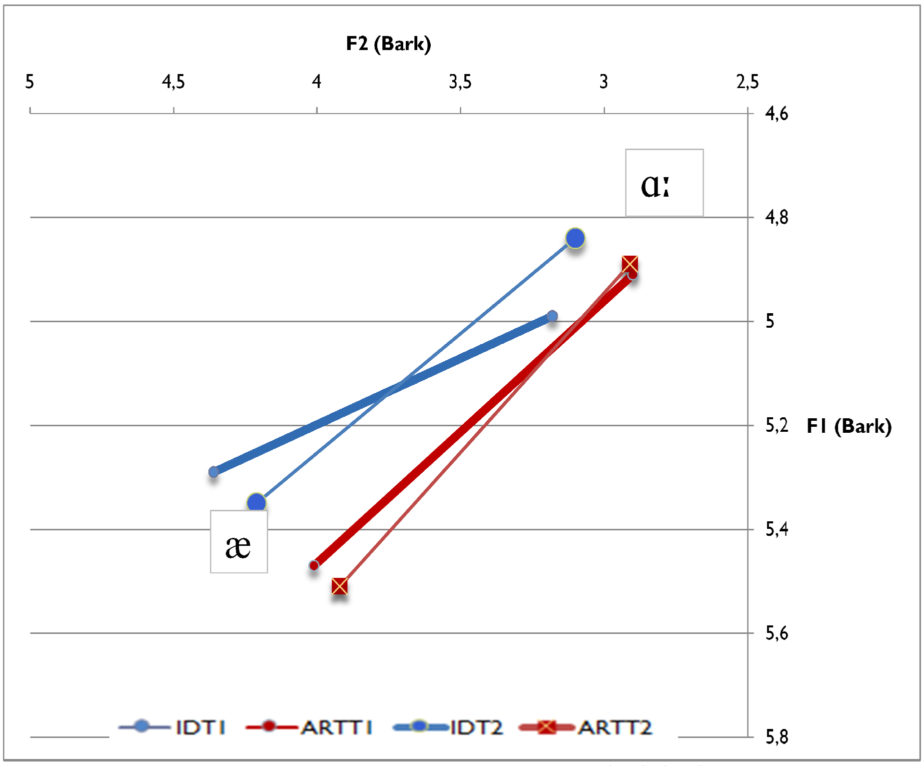


Figure 5.15. Spectral distances of the low vowel pair /æ/-/ɑː/produced by the ID training group (blue lines) and the ART training group (red lines) at pre-test (thinner lines) and post-test (thicker lines).

Finally, the ANOVAS conducted on back contrasts revealed a significant effect of *training* for /ɒ/-/ɔː/ ($F(1, 49)= 10.30$,$p< .05$, $\eta^2=.174$) and a significant *training x condition* interaction for /ʊ/-/uː/ ($F(1, 49)= 6.89$, $p< .05$, $\eta^2=.123$). None of the factors or interactions were significant for /ɒ/-/ʌ/. These are the most remarkable findings concerning back vowels:

- The SD between /ɒ/ and /ʌ/ decreased after ID training and increased after ART training, but SD changes did not reach statistical significance, so the training did not have a significant effect on the production of this vowel contrast.

- The training had a significant impact on /ɒ/-/ɔː/. In the fixed context, ID and ART groups showed opposite behaviours: whereas mean SD increased after ID training, it decreased after ART training. In context variability, there was a significant decrease in the SD between /ɒ/ and /ɔː/.

- The effect of training on the SD between /ʊ/ and /uː/ varied as a function of test condition. In fixed context, both ID and ART groups obtained higher SD between /ʊ/ and /uː/ and lower SD in context variability.

Despite not finding a significant effect of *type of training* on SDs, the results of separate paired-sample t-tests conducted for ID and ART training groups in all conditions, fixed context and context variability pointed to some between-group differences with respect to the amount of spectral changes in vowel production (Table 5.7). The most salient differences between ID and ART training groups are the following:

- Immediately after ID training, the SDs between /ɪ/and /e/ decreased significantly. SDs of the other high-front target vowels increased but SD changes were not significant. However, the ART training group produced the /iː/-/ɪ/ and /e/-/ɜː/ contrasts with significantly higher SD in relation to pre-test and the vowels of the pair /ɪ/-/e/ were significantly closer at post-test (Figure 5.14).

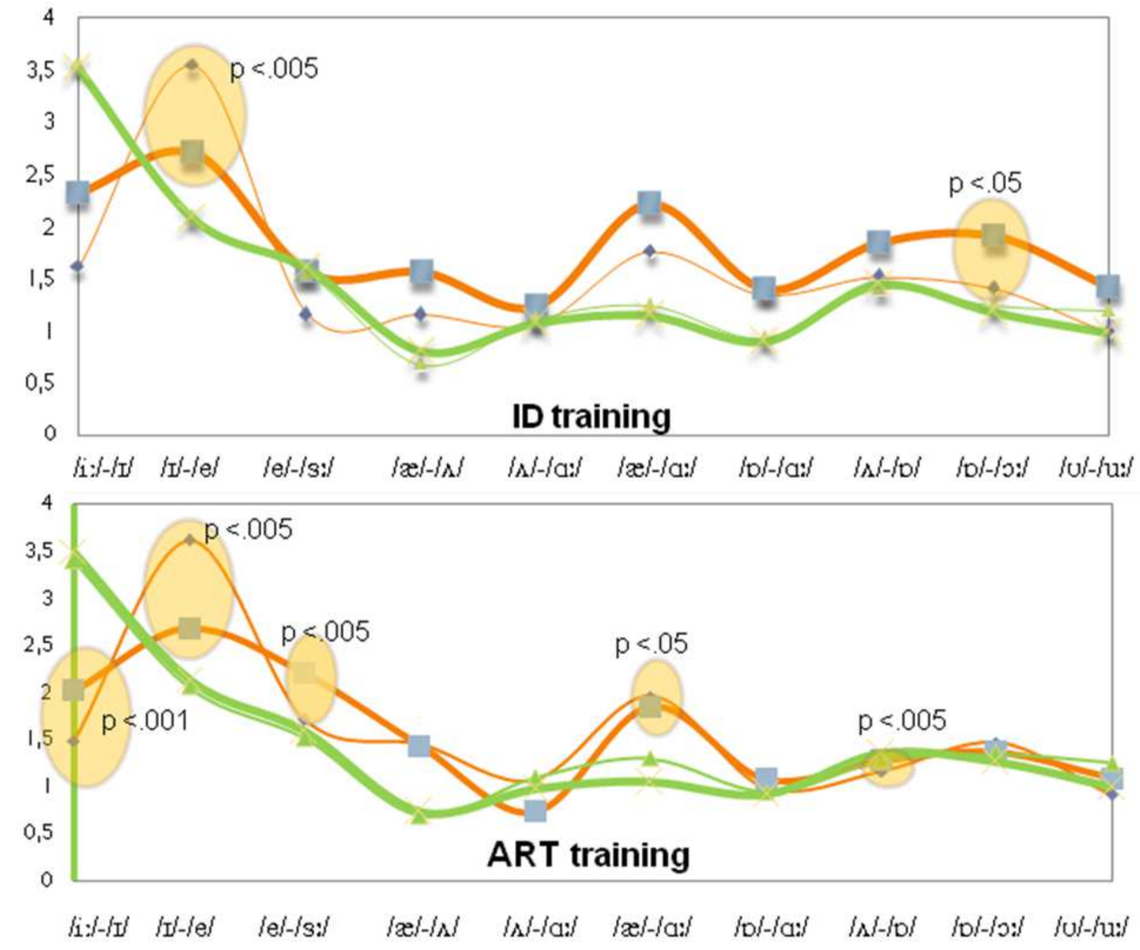
- The SDs between /æ/ and/ɑː/ was significantly shorter in the productions of the ART training group, especially in context variability. The decrease in the SD of this vowel pair was not significant for the ID training group. The amount of significant differences in the SD between vowels was greater in the ART training group than in the ID group (Figure 5.15).

- The SD between /ʌ/ and/ɑː/ increased significantly after ART training in the context variability condition but no significant differences were found after ID training.

- The SD between /ʊ/ and /uː/ decreased significantly after ART training in the context variability condition. However, differences did not turn out significant for the ID training group.

In summary, one important finding was that the effect of HVPT on SDs between non-native vowels was significant. The amount of SDchanges from pre-test to post-test varied significantly as a function of the target vowel contrast. We found that HVPT had a relatively significant impact on the articulation of three high-front vowel and one low vowel contrast. Within the high-front vowel area, the SDs increased for /iː/-/ɪ/and /e/-/ɜː/ but decreased for /ɪ/-/e/ after training. In the low and back vowel areas of the mouth, SD remained somehow steady for /æ/-/ʌ/, /ʌ/-/ɑː/,/ɒ/-/ɑː/, /ʌ/-/ɒ/, /ɒ/-/ɔː/and decreased for /æ/-/ɑː/ and /ʊ/-/uː/. Significant pre-test-post-test differences in SDs revealed that vowel contrasts changed at different rates with respect to SDs (Figure 5.16).

The fact that SD increased after training for /iː/-/ɪ/ & /e/-/ɜː/, decreased for /ɪ/-/e/suggests that only 10 intensive sessions of AV HVPT can produce significant changes in the vowel quality of certain target vowels without explicit instructions about tongue or lip position, openness or duration information, without manipulated or enhanced stimuli in training). The results also revealed that changes in SDs from pre-test to post-test were more visible in the fixed context condition than in the context variability condition, and that there was a significant effect of vowel set on SDs. After training, higher SDs were obtained between high-front vowels than between low and high-back vowels. In the fixed context condition, the target contrast /ɪ/-/e/ underwent the highest amount of spectral change after ID and ART training, which in turn affected the SD of the target vowels within the same subset, /iː/-/ɪ/ and /e/-/ɜː/.



Figures 5.16-5.17. Line graphs showing mean spectral distance scores of ten vowel contrasts produced by ID (blue line) and ART (red line) training groups, at pre-test (thinner lines) and pos-test (thicker lines), and significant pre-test/post-test differences (p< .05).

Test condition	Vowel sets	Vowel Identification	Paired comparisons: ID training (N=32) df t p-value	ID Diff.	ART training (N=32) df t p-value	ART Diff.
All contexts	High-front	/iː/-/ɪ/	t(20)=-1.33, *p*= .198	0.21 (0.71)	t(21)=-2.63, ***p*< . 05***	0.27 (0.48)
		/ɪ/-/e/	t(21)=6.60, ***p*< .001***	0.44 (0.31)	t(28)=4.76, ***p*< . 001***	0.42 (0.48)
		/e/-/ɜː/	t(22)=-1.06, *p*= .300	0.11 (0.51)	t(28)=-2.91, ***p*< . 05***	0.23 (00.42)
	Low	/æ/-/ʌ/	t(20)=-1.05, *p*= .306	0.10 (0.44)	t(28)=1.33, *p*= .195	0.11 (0.47)
		/ʌ/-/ɑː/	t(20)=-.46, *p*= .650	0.04 (0.49)	t(28)=1.38, *p*= .179	0.11 (0.42)
		/æ/-/ɑː/	t(20)= 1.30, *p*= .205	0.11 (0.41)	t(29)=2.26, ***p*< .05***	0.19 (0.46)
		/ɒ/-/ɑː/	t(20)= 0.06, *p*= .952	0.00 (0.38)	t(27)=-1.19, *p*= .245	0.10 (0.44)
	Back	/ʌ/-/ɒ/	t(20)= 0.06, *p*= .952	0.00 (0.38)	t(27)= -1.19, *p*= .245	0.10 (0.44)
		/ɒ/-/ɔː/	t(21)= -1.66, *p*= .112	0.16 (0.47)	t(28)= 0.70, *p*= .489	0.09 (0.71)
		/ʊ/-/uː/	t(21)= 1.22, *p*= .236	0.15 (0.57)	t(28)= 0.97, *p*= .341	0.07 (0.41)
Fixed context	High-front	/iː/-/ɪ/	t(23)= -1.59, *p*= .126	0.90 (2.73)	t(21)= -3.56, ***p*< .05***	0.90 (2.73)
		/ɪ/-/e/	t(24)= 4.73, ***p*< .001***	0.81 (0.85)	t(28)= 7.002, ***p*< .001***	0.81 (0.85)
		/e/-/ɜː/	t(24)= -1.91, *p*= .068	0.62 (1.61)	t(29)= -4.71, ***p*< .001***	0.62 (1.61)
	Low	/æ/-/ʌ/	t(23)=-1.37, *p*= .184	0.49 (1.74)	t(29)= 0.38, *p*= .703	0.04 (0.69)
		/ʌ/-/ɑː/	t(23)= 0.75, *p*= .458	0.27 (1.73)	t(29)= -0.68, *p*= .500	0.07 (0.57)
		/æ/-/ɑː/	t(24)= -0.93, *p*= .361	0.47 (2.51)	t(29)= 0.89, *p*= .381	0.11 (0.69)
		/ɒ/-/ɑː/	t(23)= -0.01, *p*= .992	0.0 (1.83)	t(28)= -0.89*p*= .381	0.13 (0.76)
	Back	/ʌ/-/ɒ/	t(23)=-0.61, *p*= .549	0.28 (1.71)	t(28)=-1.56, *p*= .129	0.17 (0.58)
		/ɒ/-/ɔː/	t(23)= -1.59, *p*= .125	0.55 (1.71)	t(28)= 0.64, *p*= .525	0.10 (0.85)
		/ʊ/-/uː/	t(23)= -1.19, *p*= .247	0.48 (1.98)	t(28)= -1.39, *p*= .176	0.20 (0.78)
Context variability	High-front	/iː/-/ɪ/	t(24)= 0.43, *p*= .669	0.04 (0.44)	t(28)= -1.44, *p*= .160	0.04 (0.44)
		/ɪ/-/e/	t(23)= 0.06, *p*= .951	0.01 (0.48)	t(29)= -0.79, *p*= .436	0.01 (0.48)
		/e/-/ɜː/	t(22)= 0.04, *p*= .972	0.00 (0.48)	t(28)= -1.01, *p*= .319	0.00 (0.48)
	Low	/æ/-/ʌ/	t(20)= -0.63, *p*= .536	0.06 (0.44)	t(22)= 0.09, *p*= .930	0.01 (0.39)
		/ʌ/-/ɑː/	t(24)= 1.65, *p*= .113	0.16 (0.49)	t(29)= 2.31, ***p*< .05***	0.24 (0.57)
		/æ/-/ɑː/	t(23)= 0.11, *p*= .910	0.01 (0.52)	t(28)= 1.61, *p*=.118	0.14 (0.48)
		/ɒ/-/ɑː/	t(24)= 0.63, *p*= .533	0.07 (0.56)	t(29)= 0.11, *p*= .917	0.01 (0.49)
	Back	/ʌ/-/ɒ/	t(23)= 0.26, *p*= .793	0.02 (0.44)	t(28)= 0.00, *p*= .999	0.00 (0.43)
		/ɒ/-/ɔː/	t(24)= -0.05, *p*= .959	0.01 (0.45)	t(29)= 0.66, *p*= .514	0.10 (0.83)
		/ʊ/-/uː/	t(24)= 1.67, *p*=.107	0.21 (0.64)	t(29)= 2.71, ***p*< .05***	00.26 (0.52)

Table 5.7. Results of paired-samples t-tests on the pre-test and post-test spectral distance scores of 10 vowel pairs produced by ID and ART training groups, and pre-test/post.-test differences in spectral distances.

5.1.6. Comparability of NS and NNS groups for post-test spectral distances

When comparing trainees' SDs to NS baseline data, we observed that NS-NNS differences found at pre-test still remained significant at post-test: the SDs of the target vowel contrasts /iː/-/ɪ/, /ɪ/-/e/ and /ʊ/-/uː/produced by trainees were still smaller than those of NS (see Figure 5.8). Despite the changes due to training, and lesser degree of vowel overlapping found at post-test, trainees had somewhat lower SD than NS for the rest of vowel contrasts.

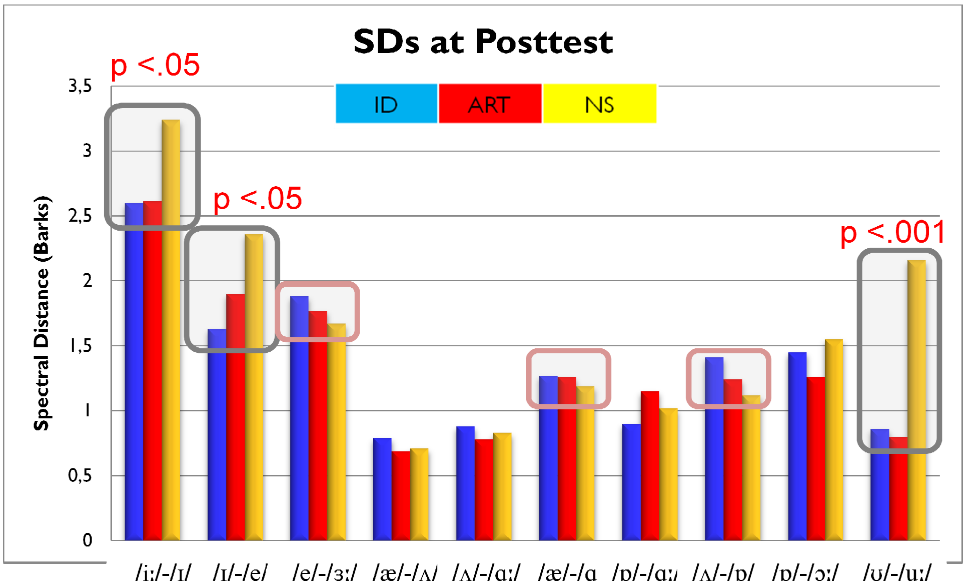


Figure 5. 18. Pre-test and post-test spectral distances (SDs) by vowel contrast for ID training, ART training and NS groups.

To further explore the SD changes as a result of training, a one-way ANOVA was conducted on post-test SDs, with *subject group* (ID, ART, NS) as between-subjects factor. The ANOVA revealed that there was significant effect of *vowel contrast* on SDs ($F(9, 495)= 56.17$, $p<.001, \eta^2=.505$) and a significant *vowel contrast × training* interaction ($F(18, 495)= 2.99$, $p<.001$, $\eta^2=.098$). There was no significant effect of *subject group*. However, significant differences between training and NS groups were obtained for /iː/-/ɪ/(p= .037), /ɪ/-/e/ (p= .009)and /ʊ/-/uː/(p= .000). The results of Posthoc tests adjusted with Bonferroni multiple comparisons confirmed that, at post-test, differences between NS and NNSs were significant for /ɪ/-/e/ (p< .05). At post-test, both ID and ART training groups produced the /ʊ/-/uː/ contrast with significantly shorter SD than NSs (p< .001) and considerable overlapping. For /iː/-/ɪ/, the tests yielded significant differences between ID and NSs (p= .036). Differences between ART and NS did not reach significance (p= .073) for this target contrast.

Spectral Distances (SD) at pre-test and post-test	**ID Training**		**ART Training**		**NS**
	T1	**T2**	**T1**	**T2**	
Mean SDs	1.44 (0.27)	1.41 (0.26)	1.38 (0.20)	1.38 (0.21)	1.61 (0.45)
C1 /iː/-/ɪ/	2.34 (0.47)	2.60 (0.71)	2.31 (0.47)	2.61 (0.60)	3.24 (0.53)
C2 /ɪ/-/e/	2.08 (0.63)	1.63 (0.51)	2.33 (0.43)	1.90 (0.33)	2.89 (0.97)
C3 /e/-/ɜː/	1.71 (0.66)	1.88 (0.58)	1.47 (0.39)	1.77 (0.36)	1.57 (0.93)
C4 /æ/-/ʌ/	0.64 (0.46)	0.79 (0.65)	0.77 (0.43)	0.69 (0.41)	0.71 (0.35)
C5 /æ/-/ɑː/	1.44 (0.68)	1.27 (0.80)	1.40 (0.59)	1.27 (0.48)	1.19 (0.49)
C6 /ʌ/-/ɑː/	1.07 (0.52)	0.88 (0.51)	0.84 (0.29)	0.78 (0.42)	0.66 (0.47)
C7 /ɒ/-/ɑː/	1.00 (0.57)	0.90 (0.47)	1.02 (0.59)	1.15 (0.59)	1.02 (0.52)
C8 /ʌ/-/ɒ/	1.46 (0.38)	1.41 (0.39)	1.13 (0.42)	1.24 (0.57)	1.12 (0.27)
C9 /ɒ/-/ɔː/	1.34 (0.78)	1.45 (0.86)	1.25 (0.44)	1.26 (0.47)	1.55 (0.98)
C10 /ʊ/-/uː/	0.98 (0.35)	0.86 (0.60)	0.89 (0.41)	0.80 (0.36)	2.16 (0.58)

Table 5.8. Pre-test and post-test spectral distances (Bark) in the vowel productions of ID training, ART training and NS groups.

5.1.5 Spectral Distance (SD) changes

The descriptive data indicated that the amount of SD changes (e.g. increase or decrease in SD) in each test condition (fixed context and context variability) from pre-test to post-test varied significantly as a function of vowel contrast (see Figure 5. 19). In the fixed context condition, the ART training group displayed larger SD changes than the ID group for vowel contrasts /iː/-/ɪ/,/æ/-/ɑː/, /ɒ/-/ɑː/, /ʌ/-/ɒ/, and /ʊ/-/uː/, whereas the rest (/ɪ/-/e/, /æ/-/ʌ/. /ʌ/-/ɑː/, /ɒ/-/ɔː/) were produced with larger SD changes at post-test by the ID group. In the context variability condition, the ART group obtained larger SD changes than ID for (/iː/-/ɪ/,/ɪ/-/e/, /e/-/ɜː/, /ɒ/-/ɑː/, /ʌ/-/ɒ/, /ɒ/-/ɔː/), whereas the rest (/æ/-/ʌ/. /ʌ/-/ɑː/, /æ/-/ɑː/, /ʊ/-/uː/) were produced with larger SDs after ID training.

The results of the repeated-measures ANOVAS run on SD changes, with *type of training* (ID vs. ART) as between-subjects factor, and *vowel contrast* (10) and *condition* (fixed context vs. context variability) as within-subjects factors, showed a significant effect of *vowel contrast* on SD changes in all conditions ($F(9, 270)= 8.17$, $p< .001$,$\eta^2=.214$) and a significant interaction between *condition* and *vowel contrast* ($F(9, 270)= 9.18$, $p< .001$,$\eta^2=.234$). The effect of *type of training* on SD change did not reach significance ($F(1, 30)= .38$, $p= .544$,$\eta^2=.012$).

In the fixed context condition, posthoc analysis using pairwise comparisons adjusted with Bonferroni showed that the SD changes obtained for /ɪ/-/e/ were significantly higher than those obtained for the other nine vowels (p<. 001). The target contrast /ɪ/-/e/ underwent the highest amount of SD change after ID and ART training, and therefore the distribution of the target vowels within the same vowel set, /iː/-/ɪ/ and /e/-/ɜː/, was affected (Figures 5.20-5.21).

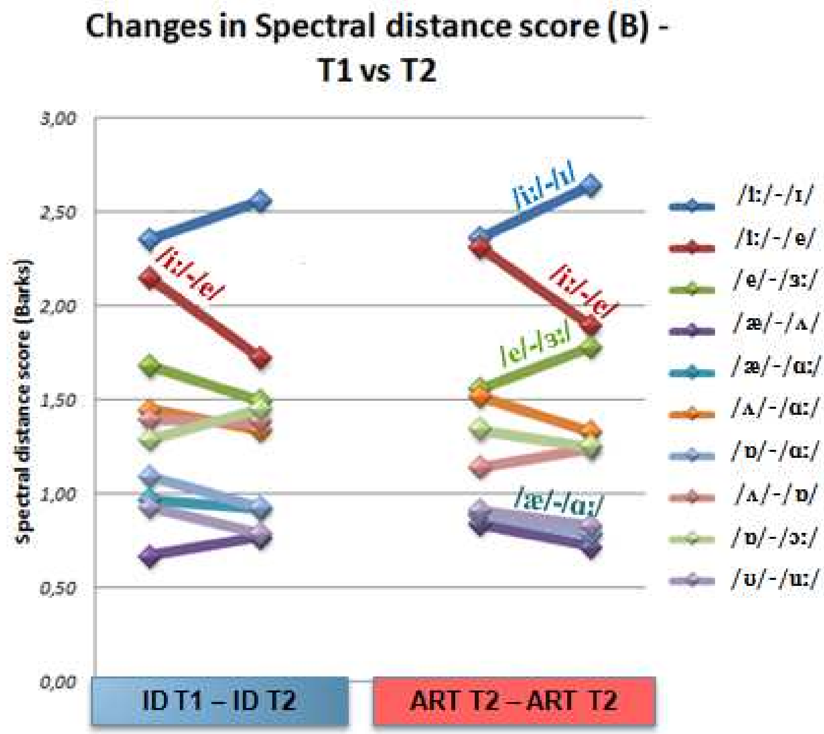
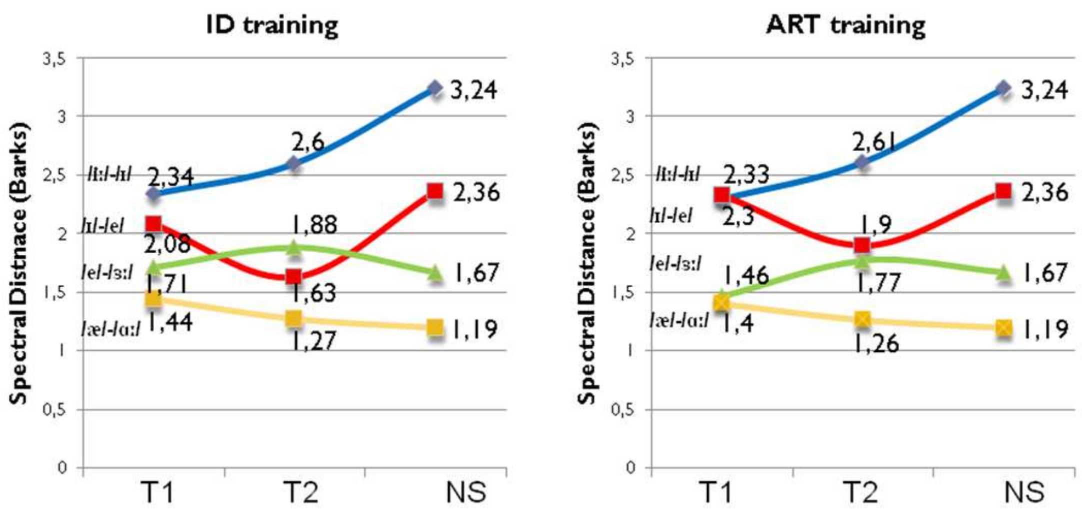


Figure 5.19. Line graph showing amount and direction of spectral distance change computed for ten vowel pairs from pre-test to post-test, after ID and ART training.



Figures 5.20-5.21. Line graphs showing amount and direction of spectral distance change for /iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/ and /æ/-/ɑː/ after ID training (left) and ART training (right).

Summary

The effect of HVPT on SDs between non-native vowels was significant. SD changes varied significantly as a function of vowel set and vowel contrast. Despite not finding a significant effect of type of training on mean SDs, we observed some salient between-group differences with respect to SD changes in vowel production. After ART training, the SDs of /iː/-/ɪ/ and /e/-/ɜː/ contrasts increased significantly, whereas the SDs of /ɪ/-/e/ and /æ/-/ɑː/ were significantly shorter. After ID training, the SDs of /ɪ/-/e/ decreased significantly, but other changes in high-front or low vowel contrasts were not significant. The amount of significant SD changes obtained for vowels was greater in the ART training group than in the ID group. The findings suggested that ART training somehow had a larger impact on the whole vowel system, esp. when vowels are embedded in a fixed context (/h_d/).

5.2. Duration Ratios (DR) and duration

The following analyses aimed at examining the effects of HVPT on the production of the non-native (trained) vowels, particularly the effects of HVPT (RQ1.4.2.) and type of training (RQ2.4b) on the use of the duration cue for more accurate vowel articulation. To this end, we used the duration (msec) measure (Table 5.9) to calculate *tense/lax duration ratios* (DR) of six tense-lax vowel pairs (/iː/-/ɪ/, /iː/-/e/, /e/-/ɜː/, /æ/-/ɑː/, /ʌ/-/ɑː/, /ʊ/-/uː/), by dividing the mean duration of the tense vowel (/iː/, /ɜː/, /ɑː/) by the mean duration of the lax vowel (/ɪ/, /e/, /æ/, /ʌ/, /ʊ/) (Wang, 2008).

Inspection of the production data at pre-test had revealed that Catalan-Spanish speakers produce all vowels with unnaturally longer duration than NSs. Taking the learners' over-reliance on duration to contrast vowels, the pre-test/post-test analysis of this measure was expected to reveal information about the trainees' use of duration differences between the vowels of these six contrasts. We wanted to better understand training-related changes in the use of the duration cue in non-native vowel production, and expected decreased DRs and decreased vowel overlap at post-test.

The analysis was executed using a series of repeated-measures ANOVAs with mean *DRs* and *duration* as the dependent variable, with effect of *training* (T1 vs. T2), and *vowel contrast* (10) as within-subjects factors each time, and *type of training* (ID vs. ART) as between-subjects factor.

Vowel duration (ms)	**ID training**	**ART training**	**NS control**
Mean duration	261.82 (31.30)	269.15 (31.95)	273.98 (20.98)
/iː/	278.69 (39.25)	282.68 (41.77)	253.92 (58.37)
/ɪ/	143.54 (11.43)	155.56 (30.78)	131.76 (10.19)
/e/	188.93 (17.05)	190.53 (24.58)	170.90 (23.69)
/ɜː/	359.60 (51.27)	368.75 (56.83)	308.57 (53.67)
/æ/	226.59 (29.92)	234.70 (33.73)	207.59 (69.38)
/ʌ/	164.51 (19.52)	170.04 (32.48)	146.97 (24.96)
/ɑː/	391.31 (55.37)	405.50 (54.67)	349.64 (75.22)
/ɒ/	207.07 (38.31)	202.77 (24.61)	181.44 (28.80)
/ɔː/	367.04 (58.86)	384.88 (55.68)	363.80 (55.05)
/ʊ/	149.04 (15.29)	156.24 (22.08)	154.98 (13.07)
/uː/	303.60 (79.69)	294.33 (39.52)	275.89 (52.07)

Table 5.9. Mean duration (ms) (and standard deviations) for the 11 vowel monophthongs produced by ID training, ART training, and NS groups at pre-test.

5.2.1. Comparability of subject groups at pre-test

ID training, ART training and control groups had a similar range of mean DR scores at pre-test (see Figure 5.22). A one-way ANOVA conducted on mean tense-lax DR (averages across the six vowel contrasts), with *subject group* (ID, ART, control) as a between-subjects factor, revealed a non-significant between-groups effect ($F(2, 58)= 0.24$, p=. 787), Another ANOVA conducted on mean DR of six tense-lax vowel contrasts yielded no significant effect of *subject group*. These results confirmed that the three subject groups did not differ significantly with respect to duration ratios at pre-test and were indeed comparable (Table 5.9).

A mixed between-within ANOVAs conducted on DRs, with *vowel contrast* (6) as within-subjects factor and *subject group* (ID, ART, control) as between-subjects factor, yielded a non-significant effect of *subject group* (F(2, 57)= 0.52, p=. 598, η^2= .018) and a significant effect of *vowel contrast* on DR (F(5, 285)= 71.79, p<.001, η^2= .557). The results confirmed, first, that pre-test duration ratios differed as a function of *vowel contrast* (e.g. the highest DRs were found for back vowels /ʌ/-/ɑː/and /ʊ/-/uː/) and, second, that the subject groups did not differ with respect to duration ratios at pre-test and were indeed comparable (Table 5.11).

Tense-lax Duration ratios (RT)	ID training	ART training	NNS control
Mean DR	1.93 (0.21)	1.86 (0.20)	1.89 (0.22)
/iː/-/ɪ/	2.01 (0.33)	1.85 (0.34)	1.99 (0.29)
/iː/-/e/	1.53 (0.24)	1.52 (0.20)	1.48 (0.28)
/e/-/ɜː/	1.94 (0.25)	1.95 (0.35)	1.85 (0.26)
/æ/-/ɑː/	1.71 (0.19)	1.75 (0.29)	1.54 (0.14)
/ʌ/-/ɑː/	2.44 (0.33)	2.38 (0.49)	2.37 (0.43)
/ʊ/-/uː/	2.12 (0.64)	1.86 (0.21)	2.15 (0.56)

Table 5.10. Mean normalized tense-lax duration ratios (DR) (and standard deviations) for six tense-lax vowel contrasts, produced by ID training, ART training, NNS control and NS groups at pre-test.

Duration Ratio (DR) Tense-lax contrasts	One-way ANOVA between-group differences
/iː/-/ɪ/	(F(2, 65)= 0.84, p=. 435
/iː/-/e/	(F(2, 65)= 0.02, p=. 978
/e/-/ɜː/	(F(2, 65)= 0.26, p=. 768
/æ/-/ɑː/	(F(2, 65)= 3.15, p=. 049
/ʌ/-/ɑː/	(F(2, 65)= 0.10, p=. 906
/ʊ/-/uː/	(F(2, 65)= 1.38, p=. 259

Table 5.11. A one-way ANOVA on pre-test duration ratios (DR) computed for six vowel contrasts produced by ID, ART and control groups.

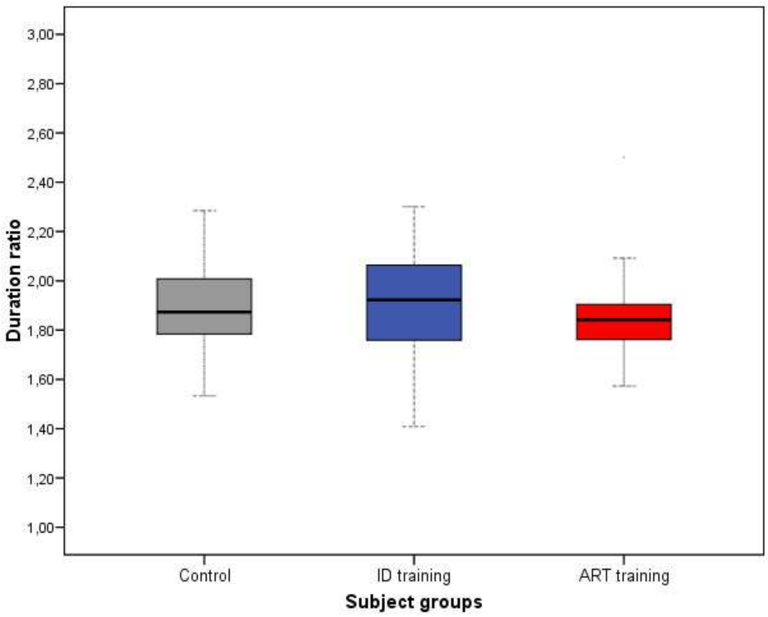


Figure 5.22. Boxplots of mean pre-test duration ratio (RT) of English tense-lax vowel contrasts produced by ID training, ART training, and control groups at pre-test.

5.2.3. Effects of training, type of training, and vowel contrasts on duration ratios (DR) and duration (ms)

One of the research questions for L2 vowel production was to what extent training would affect duration ratio as a function of *vowel* (RQ1.4.2a). RQ2.4b explored which type of training led to more positive effects on tense-lax duration ratios by vowel contrast.

The descriptive data revealed that the duration ratios (DRs) of tense-lax vowel contrasts (/iː/-/ɪ/, /iː/-/e/, /e/-/ɜː/, /æ/-/ɑː/, /ʌ/-/ɑː/, /ʊ/-/uː/) increased immediately after training, that is, trainees produced a greater temporal contrast between tense and lax vowels at post-test regardless of the type of training received (Table 5.12). In general, the DRs of tense-lax contrasts tended to be slightly larger than at pre-test, which indicates that trainees produced vowels of each pair with longer durational values. As expected, changes in DR varied as a function of vowel contrast. In particular, ID and ART training differed only for high-back contrast, /uː/-/ʊ/. At post-test the ID group produced /ʊ/-/uː/ with lower DR than at pre-test, whereas the ART group showed increased DR for the same contrast after training.

Vowel production	Duration ratio (DR)					
	ID training(*N*=32)			**ART training(*N*=32)**		
	T1	T2	Diff.	T1	T2	Diff.
/iː/-/ɪ/	2.01 (0.33)	2.06 (0.42)	0.11 (0.54)	1.85 (0.34)	2.03 (0.35)	0.17 (0.36)
/iː/-/e/	1.53 (0.24)	1.57 (0.29)	0.07 (0.32)	1.52 (0.20)	1.62 (0.25)	0.10 (0.18)
/e/-/ɜː/	1.94 (0.25)	2.00 (0.33)	0.07 (0.22)	1.95 (0.35)	2.07 (0.36)	0.15 (0.34)
/æ/-/ɑː/	1.71 (0.19)	1.77 (0.18)	0.07 (0.55)	1.75 (0.29)	1.80 (0.30)	0.08 (0.21)
/ʌ/-/ɑː/	2.44 (0.33)	2.47 (0.59)	0.08 (0.57)	2.38 (0.49)	2.64 (0.72)	0.25 (0.50)
/ʊ/-/uː/	2.12 (0.64)	2.01 (0.28)	0.01 (0.28)	1.86 (0.21)	2.09 (0.48)	0.22 (0.50)
Mean Duration Ratio (all vowel pairs)	1.93 (0.21)	1.94 (0.26)	0.15 (0.24)	1.86 (0.20)	2.02 (0.32)	0.15 (0.24)

Table 5.12. Pre-test and post-test tense-lax duration ratios (ms) and standard deviations of six tense- lax vowel contrasts produced by ID and ART training groups.

An ANOVA with *type of training* (ID vs. ART) as the between-subjects factor, and *effect of training* (T1 vs. T2) and *vowel contrast* (6) as the within-subjects factors, yielded a significant effect of *training* (F(1, 38)= 4.65, p= .05, η^2= .109) and a significant effect of *vowel contrast* (F(5, 190)= 92.68, p<.001, η^2= .709) on DR. None of the interactions were significant and the effect of *type of training* did not reach significance either. These results suggested, first, that training had a significant impact on mean duration ratios and, second, that the changes in mean duration ratios from pre-test to post-test varied significantly as a function of vowel contrast.

In order to further explore changes in DRs after ID and ART types of training, paired-samples t-tests were run on pre-test and post-test DRs for each of the tense-lax contrasts (Table 5.13). Interestingly, the tests revealed a significant increase in DR for all vowel contrasts after ART training. However, mean duration ratios of the ID group were also found to increase from pre-test to post-test but t-tests revealed no significant differences (*p*> .05).

Duration ratio (DR)	Paired comparisons			
	ID training (*N*=32)		**ART training (*N*=32)**	
	df t p-value	Diff.	df t p-value	Diff.
Mean DR	t(31)=-2.87, *p*= .869	0.15 (0.24)	t(31)=-2.87, *p*= .010*	0.15 (0.24)
/iː/-/ɪ/	t(31)=-1.02, *p*= .336	0.11 (0.54)	t(31)=-2.22, *p*= .037*	0.17 (0.36)
/iː/-/e/	t(31)=-1.01, *p*= .318	0.07 (0.32)	t(31)=-2.83, *p*= .010*	0.10 (0.18)
/e/-/ɜː/	t(31)=-1.35, *p*= .321	0.07 (0.22)	t(31)=-2.42, *p*= .022*	0.15 (0.34)
/æ/-/ɑː/	t(31)=-0.60, *p*= .190	0.07 (0.55)	t(31)=-2.01, *p*= .054*	0.08 (0.21)
/ʌ/-/ɑː/	t(31)= 0.68, *p*= .557	0.08 (0.57)	t(31)=-2.74, *p*= .010*	0.25 (0.50)
/u/-/uː/	t(31)=-0.17, *p*= .505	0.01 (0.28)	t(31)=-2.31, *p*= .029*	0.22 (0.50)

Table 5.13. Results of paired-samples t-tests on pre-test and post-test duration rations of six tense-lax vowel contrasts and pre-test/post-test differences (diff.) in duration ratios.

To further explore the changes in DR from pre-test to post-test (i.e. whether duration had increased for tense or lax vowels, or for both), similar separate analyses were run on the mean duration scores (ms) of tense and lax vowels. An ANOVA with effect of *training* (T1 vs. T2) and *type of vowel* (tense vs. lax) as within-subjects factors, and *type of training* (ID vs. ART) as between-subjects factor, yielded a significant effect of *training* on mean duration ($F(1, 38)= 862.88$, $p< .001$, $\eta^2= .958$) and a significant effect of *vowel type* (tense vs.lax) (F(1, 38)= 3.98, p= .053, η^2= .095). The effect of *type of training* did not reach significance (F(1, 38)= 0.43, p= .517, η^2= .011) but the significant interaction between *training* and *vowel type* (F(1, 38)= 4.55, p= .<05, η^2= .107) revealed two different patterns in the use of the duration cue for non-native vowel production according to type of training (Figure 5. 5). After training, the ID group consistently produced "longer" vowels regardless of the type of vowel ($tense_{T1}$: 342; $tense_{T2}$: 356; t(20)=-1.48, p= .153 or lax_{T1}: 178; lax_{T2}: 184; t(20)=-1.13, p= .273), whereas the ART group produced "longer" tense vowels (T1: 349; T2: 367; t(21)=-2.6, p< .05) and slightly "shorter" lax vowels (T1: 188; T2: 183; t(25)=0.86, p= .399) after training (Table 5.14).

Vowel production	**Acoustic measures: Duration (ms)**						
	ID training(*N*=32)			**ART training(*N*=32)**			**NS (*N*=10)**
	T1	T2	Diff.	T1	T2	Diff.	M
Tense (/iː, ɜː, ɑː, ɔː, uː/)	342.00 (48.13)	356.92 (50.87)	14.92 (49.12)	348.99 (42.23)	366.50 (49.64)	17.51 (47.34)	371.45 (14.17)
Lax (/ɪ, e, æ, ʌ, ɒ, ʊ/)	178.80 (17.42)	184.14 (22.56)	5.34 (32.33)	186.39 (21.96)	183.11 (23.42)	-3.28 (29.39)	163.07 (15.67)
Mean Vowel Duration	261.82 (31.30)	270.76 (31.32)	8.94 (26.25)	269.15 (31.95)	275.75 (32.97)	6.6 (28.12)	273.98 (20.98)

Table 5.14. Pre-test and post-test duration values (ms) (and standard deviations) of tense and lax vowels produced by ID training, ART training and NS baseline groups.

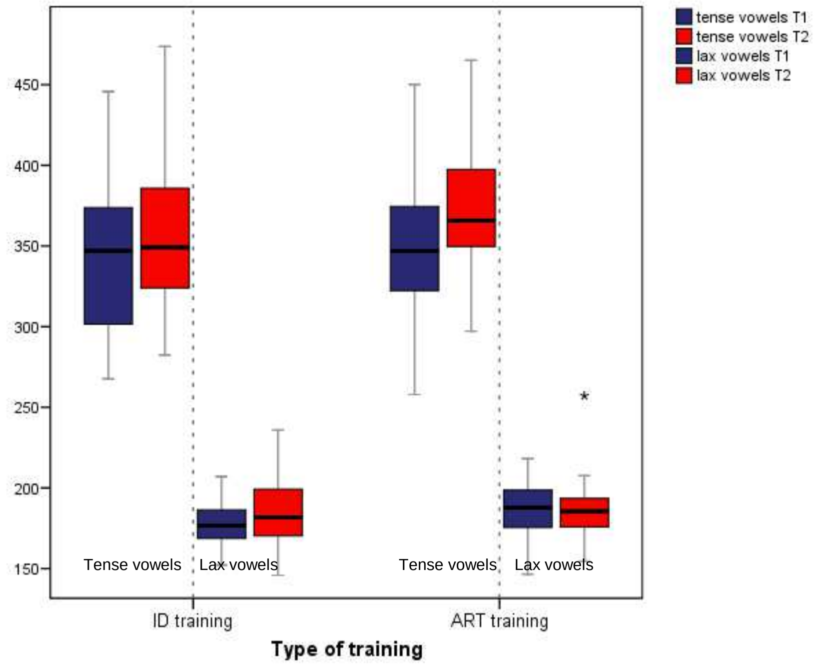


Figure 5.23. Boxplots of mean vowel duration (ms) for the tense (/iː, ɜː, ɑː, ɔː, uː/) and lax (/ɪ, e, æ, ʌ, ɒ, ʊ/) vowels produced by the ID and ART training groups, showing magnitude of durational change in vowel production from pre-test to post-test.

Summary

The findings indicate that HVPT had a significant effect on the use of the duration cue in L2 vowel production. Immediately after training, the duration ratios of the five target tense-lax vowel contrasts increased significantly (/iː/-/ɪ/, /iː/-/e/, /e/-/ɜː/, /æ/-/ɑː/, /ʌ/-/ɑː/) and decreased for the high-back vowel contrast (/ʊ/-/uː/) That confirms that, as the training progressed, spectral distance changes (Section 5.1) went hand in hand with a general increase of durational differences between tense and lax vowels. The results indicate differences between ID and ART training groups with respect to DRs: whereas the DRs changed significantly from pre-test to post-test after ART training ($p < .05$), changes in DR did not turn out significant a for the ID training group ($p > .05$).

The analyses of mean Duration scores (ms.) revealed two different patterns in the use of the duration cue as a function of type of training: Whereas the ID training group consistently produced both tense and lax vowels with increased vowel duration at post-test, the ART group produced the tense vowels with *significantly longer duration ($p< .05$) (T1: 349; T2: 367) and the lax vowels with slightly shorter duration (T1: 188; T2: 183).*

5. 3. Vowel acoustic measures: height, frontness and duration

To further examine the effects of training (RQ1.4.3) and type of training (RQ2.4c) on spectral changes in vowel production, a series of analyses were conducted to be better able to understand changes in the height and frontness dimensions of vowels produced by the ID and ART training groups at pre-test and post-test.

5.3.1. Comparability of subject groups at pre-test

Separate repeated-measures ANOVAs with *subject group* (ID, ART, control) as a between-subjects factor and *vowel* (11) as a within-subjects factor were run on height and frontness values at pre-test.

At pre-test, the ANOVA yielded a significant effect of *vowel type* ($F(10, 550)=672.84$, $p< .001$, $\eta^2=.924$) and no effect of *subject group* ($F(2, 55)=1.14$, $p= .327$, $\eta^2=.040$) or *vowel x subject group* interaction ($F(20, 550)=1.19$, $p= .255$, $\eta^2=.042$) on mean vowel height, suggesting that the two groups were well-matched prior to training and height differed significantly depending on vowel.

Similarly, for mean frontness values at pre-test the ANOVA yielded a significant effect of *vowel type* ($F(10, 560)=974.89$, $p< .001$, $\eta^2=.946$) but the *vowel x subject group* interaction was not significant, ($F(20, 560)=1.38$,$p= .124$, $\eta^2=.047$). The effect of *subject group* on vowel frontness did not reach significance ($F(1, 56)=0.49$,$p= .612$, $\eta^2=.017$), suggesting that the three subject groups were well-matched prior to training and frontness differed significantly depending on the vowel.

5.3.2. Effects of training and type of training on height

Table 5.15 illustrates the normalized mean vowel height (B1-B0) values for vowels produced by each of the training groups (ID vs. ART) at pre-test and post-test (T1 vs. T2). Figure 5.24 reflects changes in mean vowel height from pre-test to post-test for each of the training groups (RQ1.4.3). Overall, ID and ART training groups **produced higher vowels after training (i.e. vowels produced with lower height values**), suggesting that both training methods were effective in changing vowel articulation with respect to the height dimension.

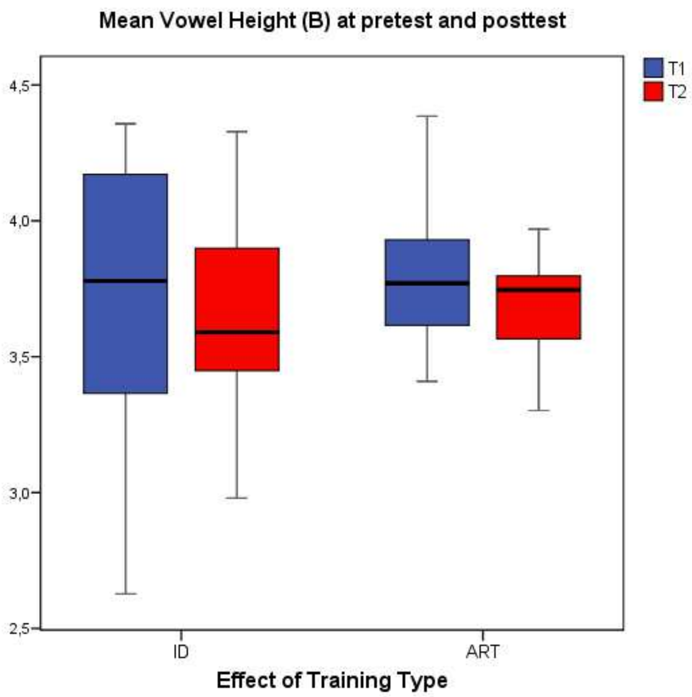


Figure 5.24. Boxplots of pre-test and post-test height (Bark) values in vowel production for ID and ART training groups.

A repeated-measures ANOVA with *effect of training* (ID vs. ART) as a within-subjects factor, and *type of training* (ID vs. ART) as a between-subjects factor, was conducted on pre-test and post-test mean height values. The ANOVA yielded a near-significant *effect of training* ($F(1, 35)=3.91$, $p= .056$, $\eta^2=.100$) and a significant effect of *vowel* ($F(10,$

350)=54731, $p < .001$, $\eta^2=.940$). The *training x vowel* interaction was significant ($F(10, 350)=3.71$, $p < .001$, $\eta^2=.096$), but the effect of *type of training* ($F(1, 35)=0.70$, $p= .407$, $\eta^2=.020$) and the *training x type of training* interaction ($F(1, 35)=0.00$, $p= .965$, $\eta^2=.000$) did not reach statistical significance. These results confirmed that the effect of training on height varied as a function of the target vowel, so vowels varied significantly along the height dimension. However, the type of training did not have a significant effect on vowel articulation with respect to height, suggesting there were no significant differences between the ID and ART groups with respect to the height dimension at pre-test and post-test.

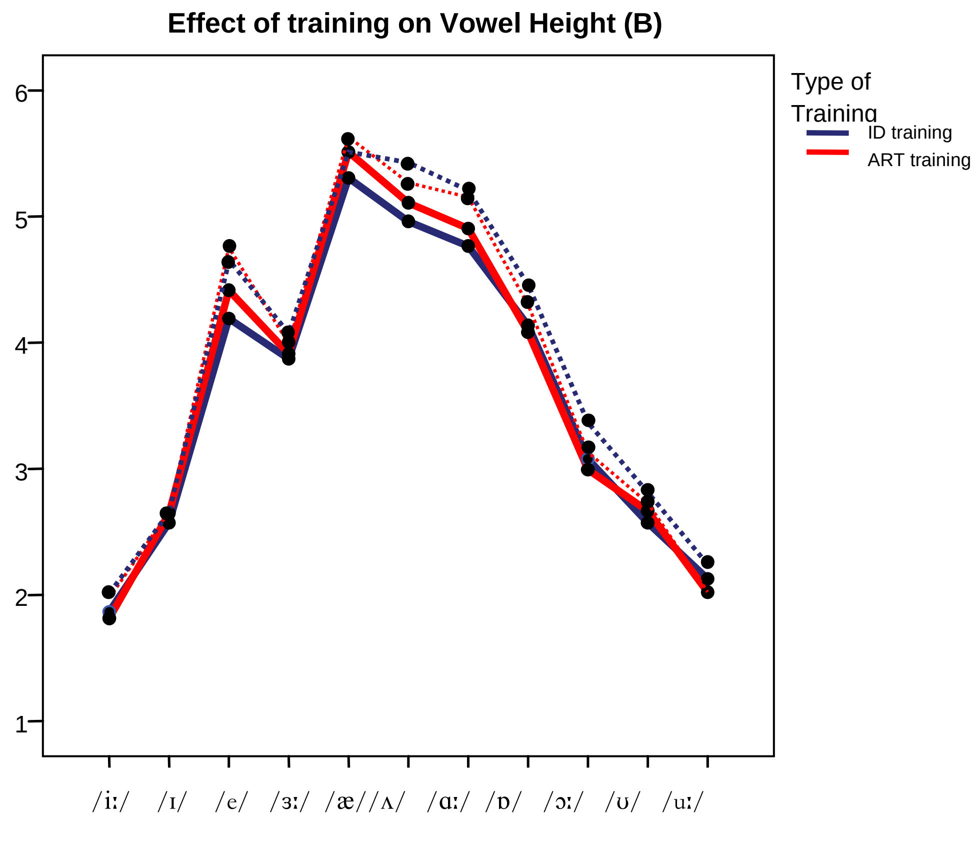


Figure 5.25. Degree of height (Bark) in the English vowel system of participants from the ID training and ART training groups, at pre-test (broken line) and post-test (solid line).

However, there was some indication of differences between the post-test height values of the ID and ART training groups: at post-test, the ID group produced the front mid vowel /e/ and most low vowels (/æʌɑː/) with lower height values than the ART group (i.e. /eæʌɑː/ were higher in the spectrum for the ID compared to the ART group) (Figure 5.25).

Vowel Production	ID training (N=17) T1	ID training (N=17) T2	ID training Diff.(p)	ART training (N=20) T1	ART training (N=20) T2	ART training Diff.(p)	NS (*N*=10)
	Acoustic measures: Height (B1-B0)						
/iː/	1.98 (0.40)	1.87 (0.41)	0.11	2.05 (0.34)	1.81 (0.31)	0.24	2.28
/ɪ/	2.59 (0.51)	2.57 (0.45)	0.02	2.70 (0.34)	2.65 (0.30)	0.05	2.85
/e/	**4.49 (0.48)**	**4.19 (0.43)**	**0.30**	**4.69 (0.36)**	**4.41 (0.28)**	0.28	4.43
/ɜː/	3.94 (0.52)	3.87 (0.53)	0.07	3.97 (0.37)	3.91 (0.34)	0.06	4.09
/æ/	5.30 (0.58)	5.51 (0.43)	-0.21	5.52 (0.25)	5.51 (0.43)	0.01	5.17
/ʌ/	5.22 (0.67)	4.96 (0.67)	0.26	5.18 (0.48)	5.11 (0.48)	0.07	5.00
/ɑː/	5.02 (0.85)	4.77 (0.70)	0.25	5.07 (0.62)	4.90 (0.61)	0.17	4.40
/ɒ/	4.30 (0.83)	4.13 (0.49)	0.17	4.26 (0.51)	4.08 (0.46)	0.18	4.02
/ɔː/	3.27 (0.68)	3.08 (0.47)	0.19	3.15 (0.51)	2.99 (0.45)	0.16	2.61
/ʊ/	2.74 (0.52)	2.57 (0.45)	0.17	2.77 (0.33)	2.67 (0.27)	0.10	2.97
/uː/	2.20 (0.54)	2.13 (0.47)	0.07	2.10 (0.38)	2.02 (0.31)	0.08	2.34
Mean vowel height	**3.70 (0.53)**	**3.61 (0.44)**	0.09	**3.81 (0.25)**	**3.71 (0.27)**	0.10	**3.65 (0.33)**

Table 5.15. Pre-test and post-test height values (Bark) (and standard deviations) for the 11 vowels produced by ID and ART training groups.

A second set of analyses included separate t-tests for each *subject group* to further examine changes in vowel height of the 11 target vowels from pre-test to post-test. Separated paired-samples t-tests conducted on post-test vowel height of the 11 vowels for ID and ART groups. As shown in Figures 5.26-5.27, **/iː/ and /e/ were produced significantly higher** in the spectrum at post-test, both in *fixed context* and *context variability* conditions, by both ID and ART groups (Tables 5.16 and 5.17). However, **/uː/ (in *fixed context*) was significantly higher after training for the ART**, but pre-test/post-test changes in height for /uː/ did not reach statistical significance for the ID group. At post-test, **vowels /ʌ/ and /ʊ/ were significantly raised for the ID group** but pre-test/post-test differences for these vowels did not reach significance for the ART group.

Test condition	Vowel Type (11)	Acoustic measures: Height (B) *(fixed context, context variability)* ID training (N=32) T1	T2	Diff.(p)	ART training (N=32) T1	T2	Diff.(p)
Fixed context	/iː/	1.58 (.044)	1.15 (1.50)	0.43	1.63 (0.41)	1.30 (0.44)	0.33
	/ɪ/	2.36 (0.54)	2.62 (1.17)	0.26	2.43 (0.41)	2.51 (0.45)	-0.08
	/e/	4.54 (0.48)	4.01 (0.50)	0.53	4.63 (0.45)	4.11 (0.31)	0.52
	/ɜː/	4.01 (0.50)	3.64 (0.73)	0.37	3.93 (0.48)	4.02 (0.33)	-0.09
	/æ/	5.83 (0.70)	5.51 (2.24)	0.32	6.02 (0.41)	6.10 (0.54)	-0.08
	/ʌ/	5.32 (0.80)	5.01 (0.74)	0.31	5.10 (0.60)	5.09 (0.47)	0.01
	/ɑː/	5.20 (0.81)	5.00 (0.81)	0.20	4.95 (0.74)	4.93 (0.58)	0.02
	/ɒ/	4.31 (0.94)	3.98 (1.91)	0.33	4.15 (0.58)	4.07 (0.50)	0.08
	/ɔː/	3.19 (0.77)	2.79 (1.75)	0.40	3.06 (0.59)	2.94 (0.51)	0.12
	/ʊ/	2.39 (0.46)	2.39 (0.91)	-	2.32 (0.25)	2.27 (0.24)	0.05
	/uː/	2.09 (0.69)	1.84 (0.50)	0.25	1.93 (0.38)	1.73 (0.41)	0.20
Context Variability	/iː/	2.31 (0.45)	2.33 (0.48)	0.02	2.43 (0.33)	2.36 (0.30)	0.07
	/ɪ/	2.77 (0.50)	2.80 (0.46)	0.03	2.97 (0.40)	2.84 (0.28)	0.13
	/e/	4.47 (0.54)	4.45 (0.50)	0.02	4.71 (0.36)	4.72 (0.31)	-0.01
	/ɜː/	3.78 (0.59)	3.77 (0.60)	0.01	3.79 (0.49)	3.81 (0.43)	-0.02
	/æ/	4.74 (0.45)	4.75 (0.52)	0.01	4.92 (0.35)	4.89 (0.43)	0.03
	/ʌ/	5.07 (0.57)	4.93 (0.69)	0.14	5.05 (0.57)	5.11 (0.56)	-0.06
	/ɑː/	4.77 (0.90)	4.74 (0.76)	0.03	4.87 (0.63)	4.86 (0.66)	0.01
	/ɒ/	4.06 (0.80)	4.15 (0.55)	0.09	4.19 (0.54)	4.11 (0.51)	0.08
	/ɔː/	3.19 (0.60)	3.22 (0.60)	0.03	3.14 (0.59)	3.09 (0.46)	0.05
	/ʊ/	3.04 (0.51)	2.97 (0.56)	0.07	3.15 (0.59)	3.09 (0.46)	0.06
	/uː/	2.21 (0.48)	2.51 (0.65)	0.30	2.16 (0.42)	2.35 (0.32)	-0.19

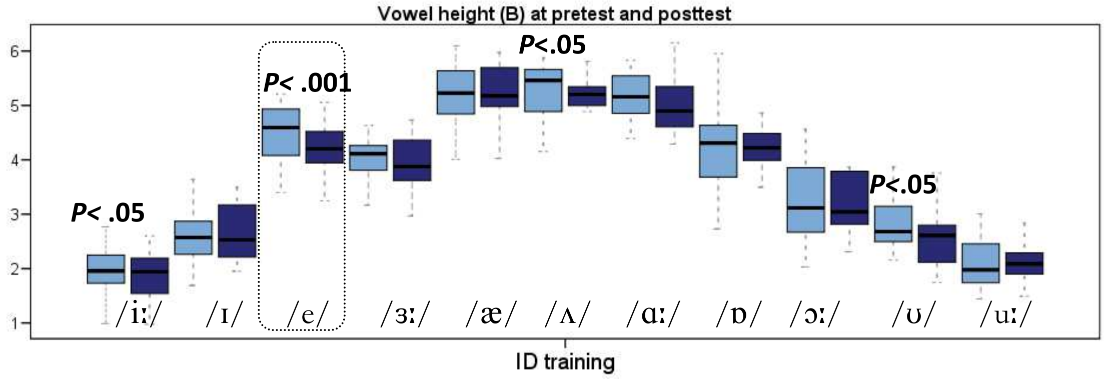
Table 5.16. Mean pre-test and post-test height values (Bark) (and standard deviations) for the 11 vowels produced by ID and ART training groups, in fixed context and context variability.

Test condition	Vowel production	Paired comparisons (vowel height) ID training (N=32) df t p-value	Diff.	ART training (N=32) df t p-value	Diff.
All contexts	/iː/	t(21)= 1.12, *p*= .275	0.79	t(21)= 3.35, ***p*= .003***	0.18
	/ɪ/	t(22)= -0.45, *p*= .655	0.38	t(28)= 0.38, *p*= .704	0.02
	/e/	t(22)= 4.98, ***p*< .001***	0.29	t(29)= 4.47, ***p*< .001***	0.25
	/ɜː/	t(22)=0.22, *p*= .827	0.02	t(28)= -0.88, *p*= .387	0.06
	/æ/	t(22)= -0.78, *p*= .444	0.05	t(29)= -0.38, *p*= .703	0.02
	/ʌ/	t(20)= 2.42, ***p*< .05***	0.19	t(28)= -0.22, *p*= .826	0.02
	/ɑː/	t(22)= 1.72, *p*= .100	0.17	t(29)= 0.15, *p*= .885	0.02
	/ɒ/	t(21)= -0.51, *p*= .612	0.06	t(28)= 1.01, *p*= .323	0.08
	/ɔː/	t(22)= 0.31, *p*= .198	0.03	t(29)= 1.41, *p*= .169	0.11
	/ʊ/	t(21)= 2.27, ***p*< .05***	0.14	t(28)= 121, *p*= .236	0.6
	/uː/	t(22)= 0.08, *p*= .931	0.01	t(29)= 0.08, *p*= .934	0.00
Fixed context	/iː/	t(22)= 1.44, *p*= .165	0.43	t(22)= 4.57, ***p*< .001***	0.33
	/ɪ/	t(24)= -1.06, *p*= .301	0.26	t(28)= -0.95, *p*= .352	0.08
	/e/	t(24)= 6.97, ***p*< .001***	0.52	t(29)= 6.03, ***p*< .001***	0.51
	/ɜː/	t(24)= 1.01, *p*= .321	0.36	t(29)= -1.00, *p*= .326	0.09
	/æ/	t(24)= 0.73, *p*= .474	0.32	t(29)= -0.94, *p*= .357	0.08
	/ʌ/	t(23)= 2.54, ***p*< .05***	0.30	t(29)= 0.07, *p*= .957	0.01
	/ɑː/	t(24)= 1.76, *p*= .091	0.20	t(29)= 0.15, *p*= .885	0.2
	/ɒ/	t(23)= -0.80, *p*= .432	0.32	t(28)= 0.72, *p*= .477	0.08
	/ɔː/	t(24)= 1.17, *p*= .254	0.40	t(29)= 1.27, *p*= .213	0.13
	/ʊ/	t(23)= 0.00, *p*= .993	0.00	t(28)= 1.00, *p*= .325	0.04
	/uː/	t(24)= 1.71, *p*= .100	0.25	t(29)= 2.50, ***p*< .05***	0.20
Context variability	/iː/	t(25)= -0.21 *p*= .839	0.02	t(29)= 1.12, *p*= .270	0.06
	/ɪ/	t(24)= -0.31, *p*= .756	0.02	t(29)= 2.32, ***p*< .05***	0.13
	/e/	t(24)= 0.38, *p*= .706	0.03	t(29)= -0.18, *p*= .857	0.01
	/ɜː/	t(22)= 0.05, *p*= .960	0.00	t(29)= -0.36, *p*= .721	0.03
	/æ/	t(24)= -0.19, *p*= .849	0.01	t(29)= 0.42, *p*= .680	0.03
	/ʌ/	t(23)= 1.61, *p*= .120	0.13	t(29)= -0.54, *p*= .590	0.05
	/ɑː/	t(24)= 0.23, *p*= .822	0.02	t(29)= 0.12, *p*= .908	0.01
	/ɒ/	t(24)= -0.99, *p*= .334	0.09	t(29)= 0.87, *p*= .391	0.08
	/ɔː/	t(24)= -0.32, *p*= .755	0.03	t(29)= 1.13, *p*= .267	0.09
	/ʊ/	t(24)= 0.74, *p*= .467	0.07	t(29)= 0.52, *p*= .607	0.05
	/uː/	t(24)= -2.68, ***p*< .05***	0.30	t(29)= -3.17, ***p*< .05***	0.19

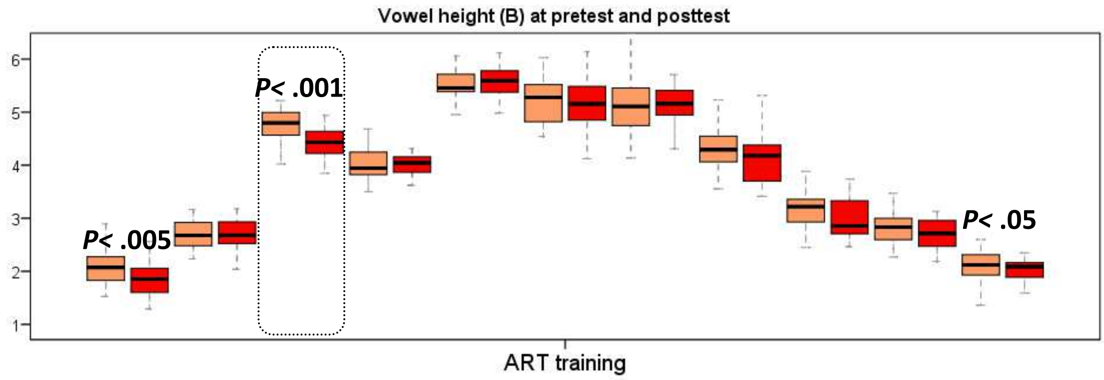
Table 5.17. Results of paired-samples t-tests on the pre-test and post-test height values (Bark) of 11 vowel contrasts produced by ID and ART training groups.

Next, an independent-sample t-test conducted on the post-test height values to examine differences between ID and ART training groups revealed **significant differences between ID and ART training groups for /e/,** t(54)= -2.09, *p*< .05. Although both ID and ART groups produced a significantly higher /e/ after training, /e/

was produced significantly higher at post-test by the ID group than the ART group (Figures 5.26-5.27).



Figures 5.26. Boxplots of mean vowel height (Bark) for the 11 vowel monophthongs (in all contexts) produced by ID training group, at pre-test (light blue) and post-test (dark blue), and the ART group.



Figures 5.27. Boxplots of mean vowel height (Bark) for the 11 vowel monophthongs (in all contexts) produced by ID training group, at pre-test (light blue) and post-test (dark blue), and the ART group.

In order to compare NS and NNS height values, separate one-way between-groups ANOVAs, with *subject group* (ID, ART, NS) as a between-subjects factor and *vowel* (11) as a within-subjects factor, were run on vowel height values at pre-test and post-test. At pre-test, there was a statistically significant effect of subject group for/e/ (p= .009). Post-hoc comparisons indicated that ID training and the NS groups differed significantly for /e/ in all conditions and *fixed context*. /e/ was significantly lower in the vowel spectrum of NSs compared to the ID group, in *all contexts* ($F(2, 62)= 5.03$; $p< .05$) and *fixed context* conditions

(F(2, 62)= 4.00; p< .05). In the *fixed context* condition, the NS group produced /e/ with significantly lower height values (higher vowel in the spectrum) than the ID group. The ART group did not differ significantly from the NS group.

Further one-way between-groups ANOVAS were conducted to explore the effect of *subject group* (ID, ART, NS) on vowel height in *fixed context* (Figure 5.30). At post-test, there was a statistically significant difference for the groups for /e/ (p= .030) and /ʊ/ (p= .006) in *fixed context*. Post-hoc comparisons indicated that, at post-test, there was a statistically significant difference between the ID training and NS groups for vowel height for /e/ (p= .031) in the *fixed context* condition, and both ID and ART groups differed from the NS group for /ʊ/ at the p< .05 and p< .005 level, respectively. However, pre-test between-group differences for /iːɪ/ in *fixed context* did not remain statistically significant at post-test.

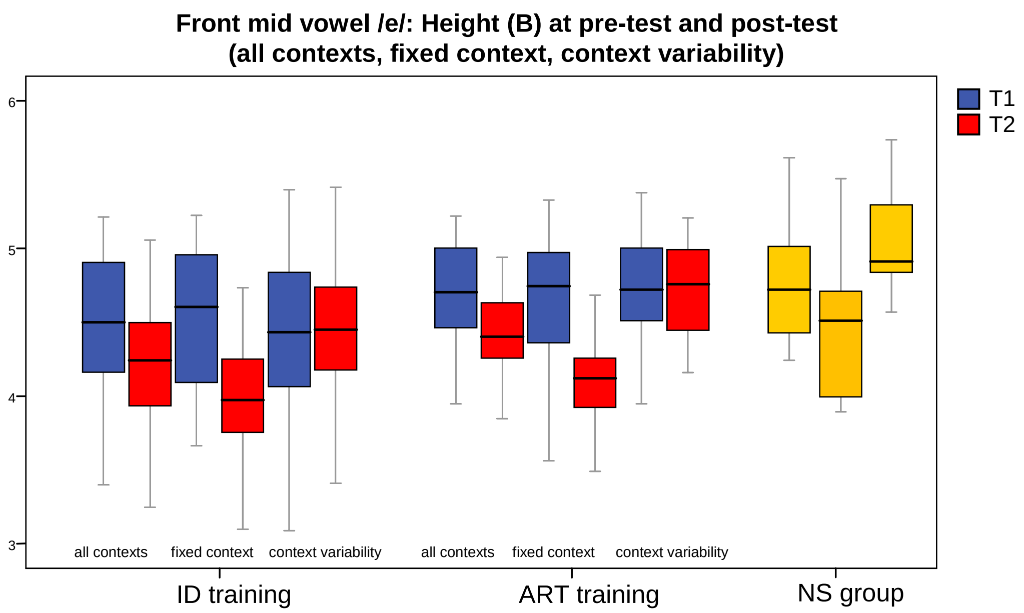


Figure 5.28. Pre-test and post-test height values (Bark) of English /e/ produced in all conditions, fixed context and context variability conditions, by the ID training, ART training and NS groups.

Although the effect of training and type of training were not significant, there was some indication of changes in the position of the tongue during the articulation. Overall, ID and ART training groups produced higher vowels after training (Figures 5.29). Particularly, /iː/ and /e/ were produced significantly higher in the spectrum at post-test, in *fixed context* and *context variability* conditions (Figures 5.29-5.30), for both ID and ART groups, suggesting that both training methods were effective in changing vowel articulation with respect to the height dimension. However, /uː/ in *fixed context* was significantly higher after training for the ART, but changes did not reach statistical significance for the ID group.

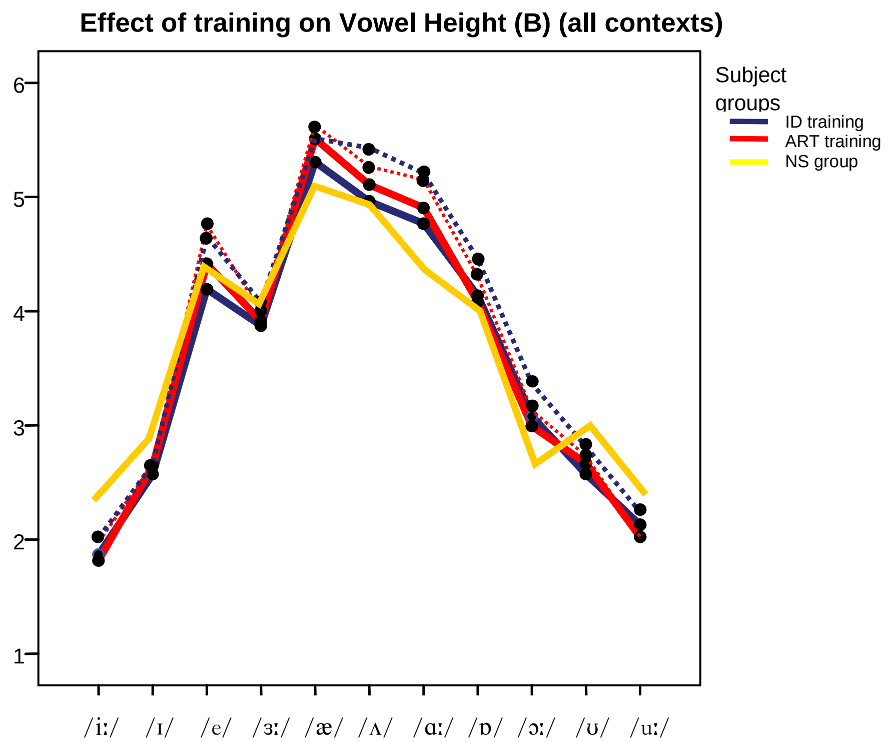


Figure 5.29. Degree of height (Bark) in all contexts in the English vowel system of ID training and ART training groups, at pre-test (broken line) and post-test (solid line), and the NS group.

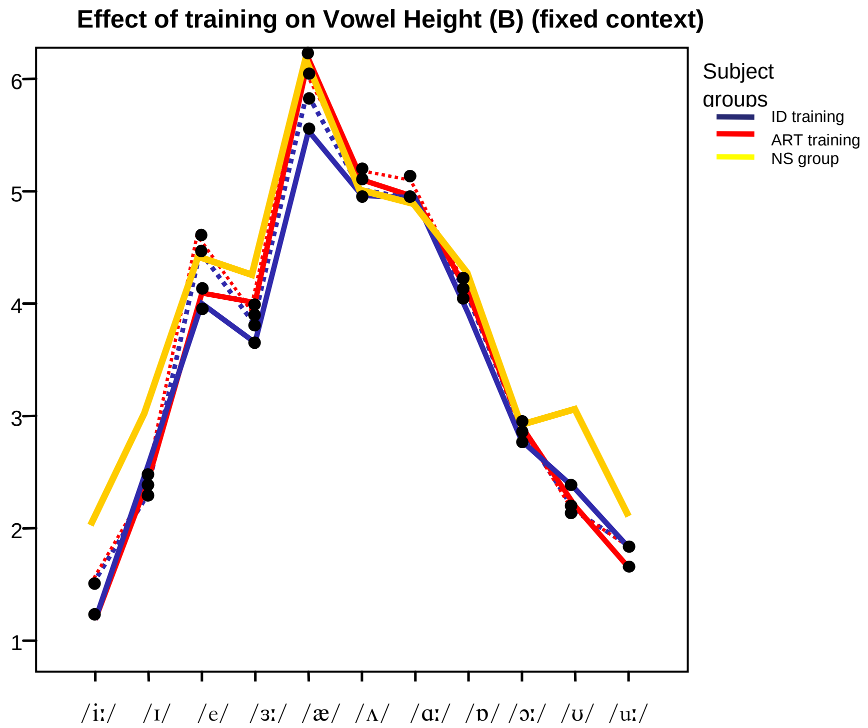


Figure 5.30. Degree of height (Bark) in fixed context in the English vowel system of ID training and ART training groups, at pre-test (broken line) and post-test (solid line), and the NS group.

Summary

Although the effect of training and type of training on vowel height was not significant, there was some indication of changes in the position of the tongue during the articulation. Overall, ID and ART training groups produced higher vowels after training. Particularly, /iː/ and /e/ were produced significantly higher in the spectrum at post-test, in fixed context and context variability conditions, by both ID and ART groups, suggesting that both training methods were effective in changing vowel articulation with respect to the height dimension. However, /uː/ in fixed context was significantly higher after training for the ART, but changes did not reach statistical significance for the ID group.

5.3.3. Effects of training and type of training on frontness

Table 5.18 illustrates the normalized mean vowel frontness (B1-B0) values for vowels produced by ID and ART training groups at pre-test and post-test (T1 vs. T2). Figure 5.31 reflects changes in mean vowel frontness from pre-test to post-test for each of the training groups. The descriptive data showed two different patterns for changes in vowel frontness at post-test. ID and ART training groups produced more fronted or more retracted vowels (i.e. vowels were produced with higher or lower height values) after training as a function of vowel. High vowels /iːeʊuː/ were generally produced with higher frontness values after training (i.e. slightly fronted vowels after training), whereas low vowels /æʌɑːɒɔː/ and mid-central vowels /ɪɜː/ became more retracted vowels after training .

Vowel Production	Acoustic measures: Frontness (B2-B1)						
	ID training (N=32)			ART training (N=32)			NS (*N*=10)
	T1	T2	Diff.	T1	T2	Diff.	
/iː/	10.14 (0.57)	10.22 (0.62)	0.08	10.15 (0.56)	10.31 (0.54)	0.16	10.88 (0.86)
/ɪ/	7.86 (0.46)	7.82 (0.34)	0.04	7.59 (1.49)	7.82 (0.47)	0.23	9.37 (0.71)
/e/	7.07 (0.61)	7.36 (0.43)	0.29	6.49 (1.81)	7.18 (0.53)	0.69	6.48 (1.31)
/ɜː/	5.52 (0.87)	5.58 (0.70)	0.06	5.38 (1.23)	5.50 (0.45)	0.12	4.87 (1.11)
/æ/	4.41 (0.60)	4.22 (0.87)	0.19	4.01 (1.51)	3.95 (0.45)	0.06	3.85 (0.67)
/ʌ/	3.98 (0.62)	3.97 (0.63)	0.01	3.53 (1.40)	3.56 (0.46)	0.03	3.26 (0.59)
/ɑː/	3.17 (0.84)	3.12 (0.78)	0.05	2.90 (1.05)	2.92 (0.47)	0.02	2.74 (0.57)
/ɒ/	3.38 (0.63)	3.19 (0.58)	0.19	3.02 (1.22)	3.02 (0.47)	-	2.81 (0.59)
/ɔː/	3.98 (0.70)	3.96 (1.01)	0.02	3.59 (0.81)	3.36 (0.50)	0.23	3.10 (0.63)
/ʊ/	5.83 (0.91)	5.84 (0.88)	0.01	5.96 (0.63)	5.91 (0.69)	0.05	5.97 (0.88)
/uː/	5.75 (0.77)	6.11 (0.82)	0.36	5.96 (1.37)	6.10 (0.72)	0.14	6.68 (1.04)
Mean	5.57 (0.42)	5.61 (0.39)		5.47 (0.29)	5.39 (0.27)		

Table 5.18. Pre-test and post-test frontness values (Bark) (and standard deviations) for the 11 vowels produced by ID and ART training and NS groups.

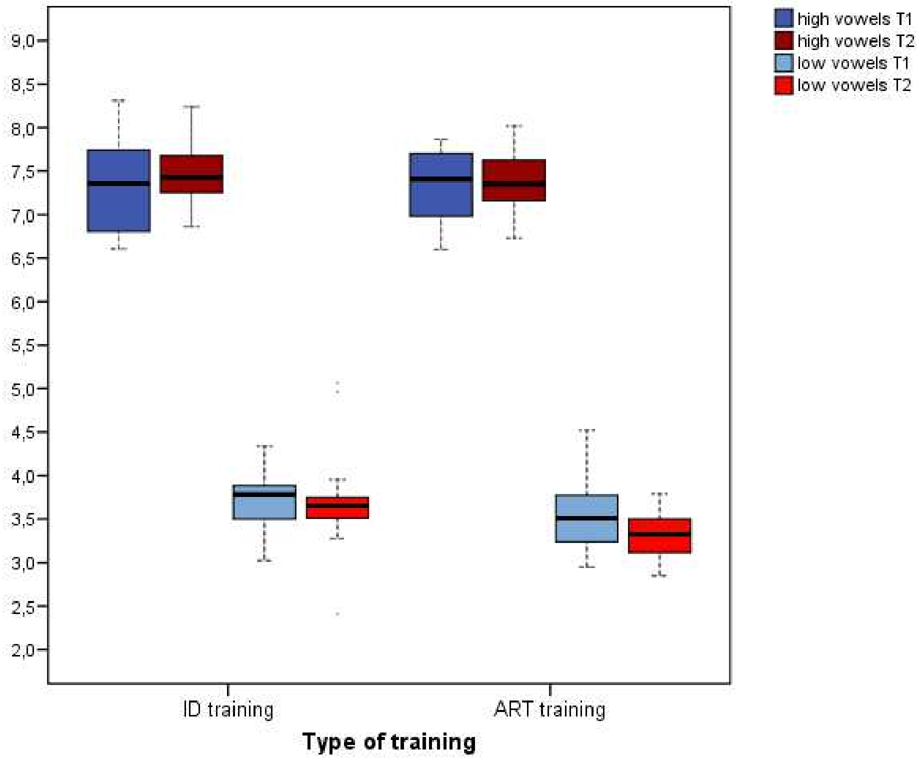


Figure 5.31. Boxplots of pre-test and post-test frontness (Bark) values of high (/iːeʊuː/) and low vowels /æʌɑːɒɔː/ produced by ID and ART training groups.

Separate repeated-measures ANOVA with *training* (T1 vs. T2) as the within-subjects factor, and *type of training* (ID vs. ART) as the between-subjects factor, were conducted separately on the pre-test and post-test frontness of high (/iːeʊuː/), mid (/ɪɜː/) and low (/æʌɑːɒɔː/) vowels (Table 5.19).

Vowel Production	Acoustic measures: Frontness (B2-B1)						
	ID training (*N*=32)			**ART training (*N*=32)**			**NS (*N*=10)**
	T1	T2	Diff	T1	T2	Diff.	
high	5.75 (0.47)	5.90 (0.39)	fronted	5.83 (0.35)	5.91 (0.39)	fronted	5.39 (0.58)
low	3.74 (0.53)	3.62 (0.58)	retracted	3.36 (1.27)	3.36 (0.34)	same	2.88 (0.46)
mid	6.69 (0.52)	6.65 (0.35)	retracted	6.48 (1.34)	6.65 (0.35)	fronted	6.09 (0.39)
Mean vowel frontness	5.57 (0.42)	5.47 (0.29)	retracted	5.60 (0.39)	5.39 (0.27)	retracted	4.47

Table 5.19. Pre-test and post-test frontness values (Bark) (and standard deviations) for high (/iːeʊuː/), low (/æʌɑːɒɔː/) and mid-central (/ɪɜː/) vowels produced by ID training, ART training, and NS groups.

For high vowels, the ANOVA yielded a significant effect of *training* ($F(1, 41)=4.99$, $p< .05$, $\eta^2=.109$). However, there were no significant *vowel type x training type* interaction ($F(3, 123)=1.992$, $p= .130$, $\eta^2=.05$) or effect of *training type* ($F(1, 41)=0.12$, $p= .728$, $\eta^2=.003$). The results suggested that training had a significant effect on the degree of frontness of /iːeʊuː/, which became significantly more fronted at post-test (i.e. after training, mean frontness was significantly raised for high vowels). Although high vowels were slightly more fronted for the ART group after training, differences between training groups were not statistically significant.

For low vowels, the results of the ANOVA showed a significant effect of *training type* ($F(1, 49)=5\text{-}52$, $p< .05$, $\eta^2=.101$). However, the effect of *type of training* ($F(1, 49)=0.26$, $p= .609$, $\eta^2=.005$), and *vowel x type of training* interaction ($F(4, 196)=0.82$, $p= .515$, $\eta^2=.016$) did not reach statistical significance, which indicated that changes in vowel frontness were non-significant for /æʌɑːɒ/. At post-test, /æʌɑːɒ/ were significantly more retracted for the ART than for the ID group but differences were not significant.

For mid-central vowels, hat the effect of *training* ($F(1, 50)=0.47$,$p= .497$, $\eta^2=.009$) and *type of training* (F (1, 50)= 0.68, $p= .414$, $\eta^2=.010$) were not significant, nor was the *vowel x type of training* interaction ($F(1, 50)=0.07$ $p= .785$, $\eta^2=.001$). These indicated that changes in vowel frontness were non-significant for /ɪɜː/ (i.e. after training, mean frontness was only slightly reduced), and the *type of training* did not have an effect on frontness. At post-test, /ɪɜː/ remained more retracted for the ART than for the ID group, but between-groups differences did not reach statistical significance.

Next, separated paired-samples t-tests conducted on the vowel frontness values of the 11 vowels for ID and ART groups (Table 5.21) (Figures 5.33-5.34) demonstrated that only /iː/ was produced significantly more fronted at post-test by the ART group. Changes found for /iː/ for the ID group were statistically non-significant (i.e. /iː/ was in a more fronted position after ID training)./iːeʊuː/ were found to be more fronted after both ID and ART training but changes did not reach statistical significance (Figure 5.33-5.34). The rest were in a more retracted position after both types of training but changes were not significant either ($p > .05$).

A second set of analyses included separate paired-sample t-tests to further examine changes in vowel frontness in fixed context and in context variability (Table 5.18). Separated paired-samples t-tests conducted on vowel frontness in all conditions demonstrated that /e/ was significantly fronted (produced with significantly higher frontness values) at post-test by both ID and ART training groups. /uː/ was also found to be significantly fronted (produced with significantly higher frontness values) for the ID group but pre-test-post-test differences did not reach significance for the ART group.

In *context variability,* the results demonstrated that /ɪ/ was significantly less fronted ($p = .053$) (produced with significantly lower frontness values) at post-test by the ID training group. In the *fixed context* condition, /iː/ was significantly more fronted (produced with significantly higher frontness values) at post-test by the ART group.

An independent-sample t-test revealed significant differences between the two training groups for the low vowels /ʌ/ and /ɔː/, in both *fixed context*(/ʌ/($t(56) = 2.35$, $p = .022$) and /ɔː/ ($t(56) = 3.02$, $p = .004$) and *context variability*(/ʌ/ ($t(54) = 2.74$, $p = .008$) and /ɔː/ ($t(54) = 1.97$, $p = .054$) conditions. These vowels were at a significantly more fronted position for the ID group as compared to the ART group in both conditions (Figure 5.32).

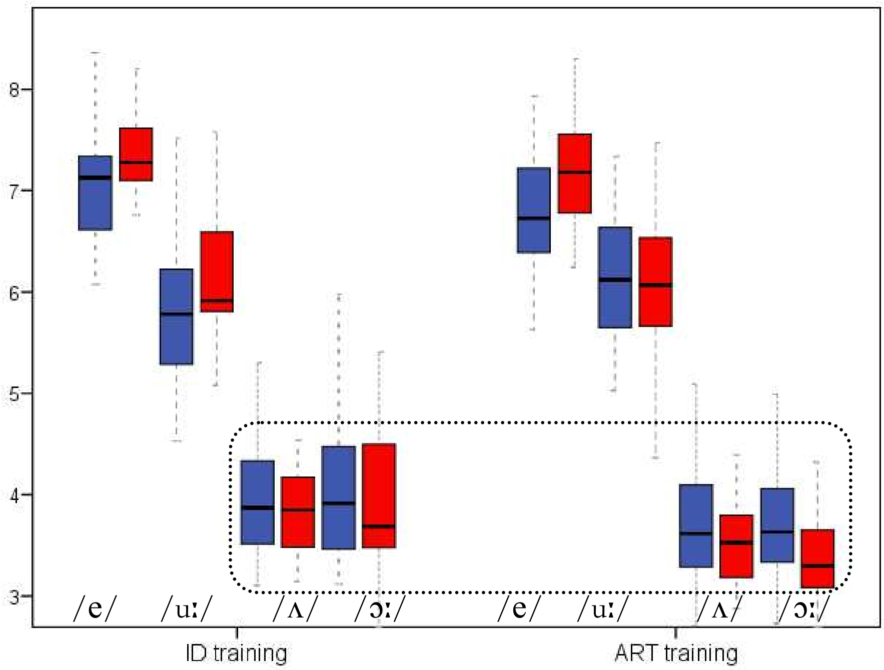


Figure 5.32. Pre-test and post-test frontness (Bark) values of English /e/, /uː/, /ʌ/and/ɔː/produced by the ID and ART training groups.

Test condition	Vowel Type (11)	Acoustic measures: Frontness (B) *(fixed context, context variability)* ID training (N=32) T1	ID training (N=32) T2	ART training (N=32) T1	ART training (N=32) T2	NS (*N*=10)
Fixed context	/iː/	10.12 (0.55)	10.29 (0.61)	10.15 (0.56)	10.32 (0.53)	10.88 (0.86)
	/ɪ/	7.88 (0.45)	7.85 (0.35)	7.59 (1.49)	7.82 (0.46)	9.37 (0.71)
	/e/	7.03 (0.60)	7.38 (0.42)	6.49 (1.81)	7.19 (0.52)	6.48 (1.31)
	/ɜː/	5.53 (0.82)	5.57 (0.70)	5.40 (1.21)	5.50 (0.44)	4.87 (1.11)
	/æ/	4.36 (0.57)	4.21 (0.83)	4.01 (1.51)	3.92 (0.46)	3.85 (0.67)
	/ʌ/	4.00 (0.58)	3.95 (0.59)	4.53 (1.40)	3.55 (0.47)	3.26 (0.59)
	/ɑː/	3.18 (0.77)	3.10 (0.76)	2.90 (1.05)	2.91 (0.47)	2.74 (0.57)
	/ɒ/	3.37 (0.61)	3.15 (0.55)	3.02 (1.22)	3.01 (0.47)	2.81 (0.59)
	/ɔː/	3.95 (0.72)	3.91 (0.99)	3.59 (0.81)	3.35 (0.50)	3.10 (0.63)
	/ʊ/	5.84 (0.88)	5.84 (0.82)	5.96 (0.63)	5.93 (0.67)	5.97 (0.88)
	/uː/	5.75 (0.70)	6.08 (0.79)	5.96 (1.37)	6.11 (0.71)	6.68 (1.04)
Context Variability	/iː/	9.40 (0.80)	9.31 (0.69)	8.78 (2.29)	9.32 (0.68)	8.63 (0.45)
	/ɪ/	5.93 (0.45)	5.82 (0.23)	5.67 (1.15)	5.86 (0.34)	5.48 (0.84)
	/e/	6.97 (0.54)	6.94 (0.54)	6.28 (2.04)	6.73 (0.50)	5.94 (0.78)
	/ɜː/	5.66 (0.74)	5.61 (0.73)	5.44 (1.43)	5.52 (0.47)	5.02 (0.80)
	/æ/	4.48 (0.55)	4.35 (0.86)	4.12 (1.62)	4.02 (0.40)	3.65 (1.08)
	/ʌ/	4.36 (0.52)	4.29 (0.71)	3.85 (1.71)	3.85 (0.48)	3.03 (0.42)
	/ɑː/	3.50 (0.78)	3.36 (0.81)	3.02 (1.26)	3.08 (0.49)	2.54 (0.21)
	/ɒ/	3.57 (0.68)	3.42 (0.53)	3.14 (1.47)	3.21 (0.44)	2.55 (0.45)
	/ɔː/	4.12 (0.71)	3.96 (0.77)	3.74 (0.89)	3.62 (0.53)	2.50 (0.45)
	/ʊ/	6.12 (0.91)	5.97 (0.94)	5.64 (1.79)	5.86 (0.70)	5.37 (0.96)
	/uː/	5.59 (0.73)	5.73 (0.71)	5.50 (1.54)	5.53 (0.63)	6.26 (0.78)

Table 5.20. Pre-test and post-test frontness values (Bark) (and standard deviations) for the 11 vowels produced by ID and ART training groups, in fixed context and context variability conditions.

Test condition	Vowel production	Paired comparisons (vowel frontness) ID training (N=32) df t p-value	Diff.	ART training (N=32) df t p-value	Diff.
All contexts	/iː/	t(21)= -0.61, *p*= .549	0.09	t(22)= -1.51, *p*= .145	0.16
	/ɪ/	t(22)= 0.39, *p*= .696	0.38	t(29)= -0.86, *p*= .394	0.23
	/e/	t(22)= -2.32, ***p*< .05***	0.29	t(29)= -2.00, ***p*= .055***	0.69
	/ɜː/	t(22)=-0.57, *p*= .577	0.06	t(28)= -0.50, *p*= .617	0.12
	/æ/	t(22)= 1.93, *p*= .067	0.19	t(29)= 0.22, *p*= .829	0.06
	/ʌ/	t(21)= 0.18, *p*= .857	0.02	t(29)= -0.12, *p*= .901	0.03
	/ɑː/	t(22)= 0.43, *p*= .673	0.04	t(29)= -0.09, *p*= .932	0.02
	/ɒ/	t(21)= 1.77, *p*= .091	0.19	t(28)= 0.02, *p*= .985	0.00
	/ɔː/	t(22)= 0.13, *p*= .901	0.02	t(29)= 1.39, *p*= .176	0.22
	/ʊ/	t(21)= -0.11, *p*= .911	0.01	t(28)= 0.43, *p*= .670	0.05
	/uː/	t(22)= -3.40, ***p*< .05***	0.36	t(29)= -0.54, *p*= .591	0.14
Fixed context	/iː/	t(21)= -0.38, *p*= .704	0.09	t(22)= -1.70, *p*= .102	0.20
	/ɪ/	t(24)= 0.86, *p*= .398	0.32	t(29)= -0.79, *p*= .436	0.26
	/e/	t(24)= -0.72, *p*= .476	0.27	t(29)= -2.97, ***p*< .05***	0.92
	/ɜː/	t(24)=0.47, *p*= .644	0.18	t(29)= -0.59, *p*= .559	0.12
	/æ/	t(24)= 1.37, *p*= .182	0.16	t(29)= 0.20, *p*= .843	0.05
	/ʌ/	t(23)= 0.02, *p*= .982	0.00	t(29)= -0.24, *p*= .809	0.05
	/ɑː/	t(24)= 0.86, *p*= .398	0.29	t(29)= 0.20, *p*= .846	0.04
	/ɒ/	t(23)= 1.75, *p*= .093	0.56	t(28)= 0.48, *p*= .637	0.11
	/ɔː/	t(24)= -0.71, *p*= .482	0.15	t(29)= 1.82, *p*= .079	0.33
	/ʊ/	t(23)= -0.04, *p*= .967	0.01	t(28)= 0.05, *p*= .963	0.01
	/uː/	t(24)= -0.04, *p*= .967	0.02	t(29)= -0.92, *p*= .363	0.25
Context variability	/iː/	t(25)= 0.67, *p*= .507	0.11	t(29)= -1.31, *p*= .200	0.55
	/ɪ/	t(25)= 2.03, ***p*= .053***	0.11	t(29)= -0.93, *p*= .362	0.19
	/e/	t(24)= 0.30, *p*= .768	0.03	t(29)= -1.17, *p*= .251	0.45
	/ɜː/	t(24)=0.56, *p*= .578	0.07	t(29)= -0.45, *p*= .656	0.12
	/æ/	t(24)= 1.42, *p*= .168	0.17	t(29)= 0.22, *p*= .829	0.07
	/ʌ/	t(24)= 0.49, *p*= .632	0.06	t(29)= -0.04, *p*= .971	0.01
	/ɑː/	t(24)= 1.54, *p*= .137	0.15	t(29)= 0.32, *p*= .754	0.07
	/ɒ/	t(24)= 1.27, *p*= .217	0.14	t(29)= 0.33, *p*= .745	0.09
	/ɔː/	t(24)= 1.31, *p*= .201	0.16	t(29)= 0.63, *p*= .533	0.12
	/ʊ/	t(24)= 1.03, *p*= .312	0.14	t(29)= -0.62, *p*= .538	0.21
	/uː/	t(24)= -0.96, *p*= .347	0.12	t(29)= -0.13, *p*= .900	0.04

Table 5.21. Results of paired-samples t-tests on the pre-test and post-test height values of 11 vowels produced by ID and ART training groups.

In order to explore the distance between ID and ART training groups and NS control groups, separate repeated-measures ANOVAs with s*ubject group* (ID, ART, NS) as a between-subjects factor and *vowel* (11) as a within-subjects factor were run on vowel frontness at post-test, in all test conditions (all contexts, fixed context, context variability). At post-test, the vowels /eɜːæʌɔː/ were still considerably more fronted in the spectrum in the fixed context condition for both ID and ART training groups as compared to NSs. However; differences between the training groups and the NS group became less salient for the rest of vowels (/iːɪɑːɒʊ/) after training. /uː/ continued more fronted for NS talkers than for the training groups after training. The ANOVA revealed a significant effect of *vowel* (F(10, 630)=437.27, p< .001, η^2=.874), and significant *vowel x subject group* interaction (F(2, 630)=3.13, p< .001, η^2=.090), in the fixed context condition at post-test. The effect of *subject group*(F(2, 63)=0.55, p= .582, η^2=.197) did not reach significance at post-test, suggesting that none of the training groups differed significantly from the NS control group in *fixed context* after each training method. Pairwise comparisons indicated pre-test significant differences between the training groups and the NS group had disappeared for vowel frontness in the *fixed context* condition at post-test, suggesting that post-test frontness values were slightly more homogeneous across groups after training.

The ANOVA conducted to explore the impact of subject group (ID, ART, NS) on vowel frontness in *context variability* (Figure 5.34) revealed a significant effect of *vowel type* (F(10, 580)=419.13, p< .001, η^2=.895) at post-test, but the effect of *subject group* was significant at post-test (F(1, 58)=9.17, p< .001, η^2=.240). Both at pre-test and post-test, the vowels /eɜːæʌɑːɔːɒʊ/ were still considerably more fronted in the spectrum in the *context variability* condition for ID and ART training groups than for the NS group (Figure 5.34).

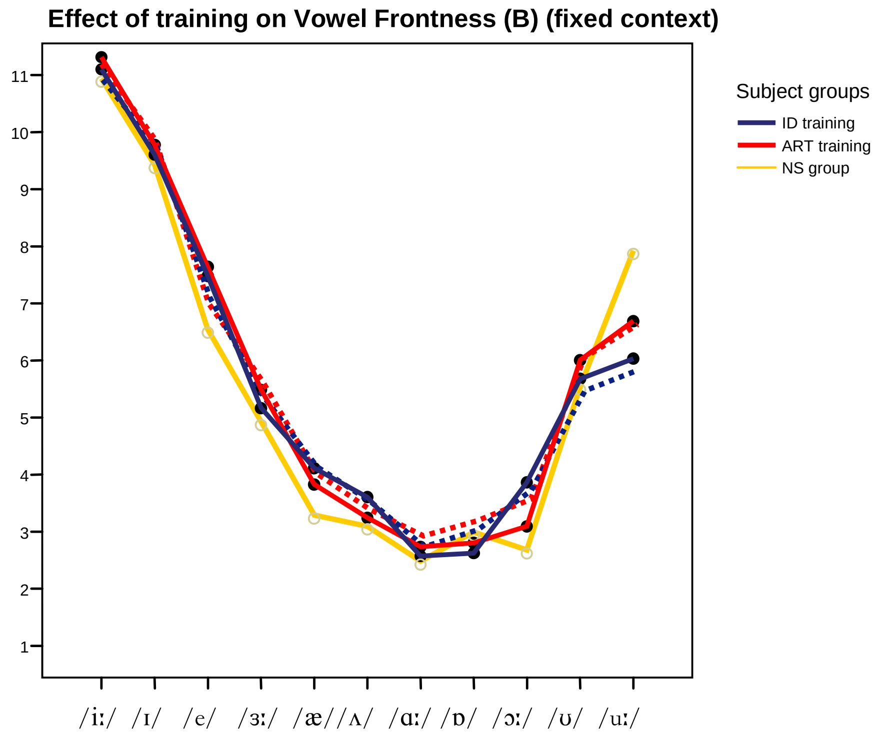
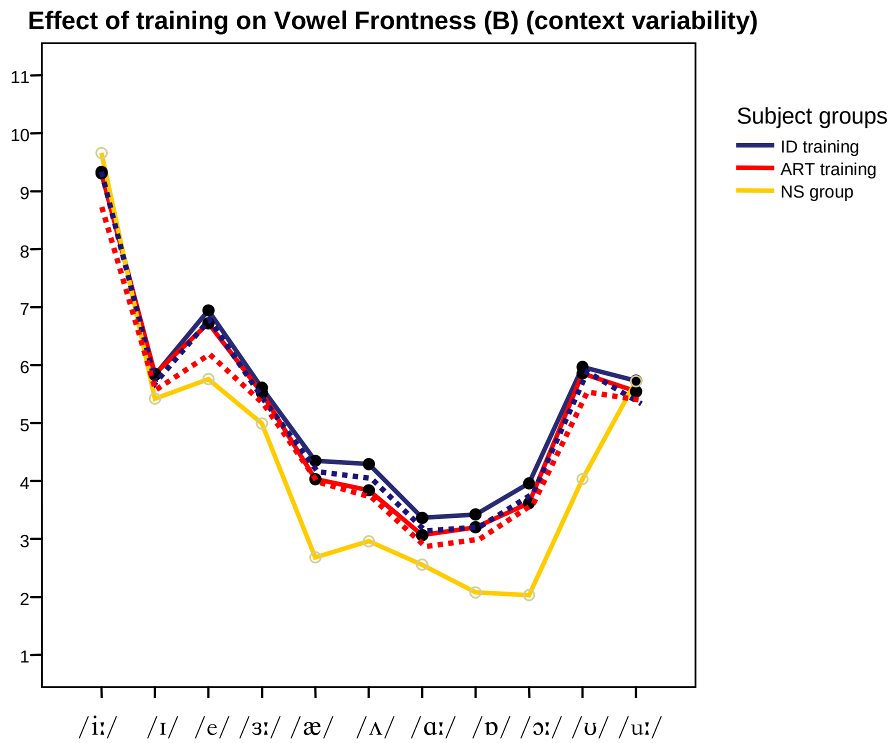


Figure 5.33. Degree of frontness (Bark) in fixed context in the English vowel system of ID training and ART training groups, at pre-test (broken line) and post-test (solid line), and the NS group.



Figures 5.34. Degree of frontness (Bark) in context variability in the English vowel system of ID training and ART training groups, at pre-test (broken line) and post-test (solid line), and the NS group.

Next, separate one-way between-groups ANOVAs with *subject group* (ID, ART, NS) as a between-subjects factor and *vowel type* (11 vowels) as a within-subjects factor were carried out on vowel frontness at pre-test and post-test. At pre-test, there was a statistically significant between-group difference in /ɔː/ (p= .000) and /uː/ (p= .017) and a near-significant difference in /e/ (p= .061) (Figure 5.3.12). At pre-test, post-hoc comparisons using Bonferroni indicated that the differences between the ID training and the NS groups were significant for /e/ in all conditions. /e/ was almost significantly more retracted in the vowel spectrum of NSs compared to the ID group in *all contexts*(F(2, 62)= 3.097; p= .052) and *fixed context* conditions (F(2, 62)= 4.006; p= .023). In both contexts, the NS group produced /e/ with significantly lower height values (higher vowel in the spectrum) than the ID group. The ART group did not differ significantly from the NS group for /e/. At post-test, between-group differences in /e/ (p= .000)./ɔː/ (p= .000) and /uː/ (p= .006) remained significant, but further significant differences arose between NS and each of the training groups after training, for /æ/ (p< .05) and /ʌ/ (p< .005) in all conditions, *fixed context* and *context variability* conditions. Both ID and ART training groups produced vowels /e æʌɔː/ with significantly higher frontness values (i.e. in a more fronted position in the spectrum) than the NS group after training.

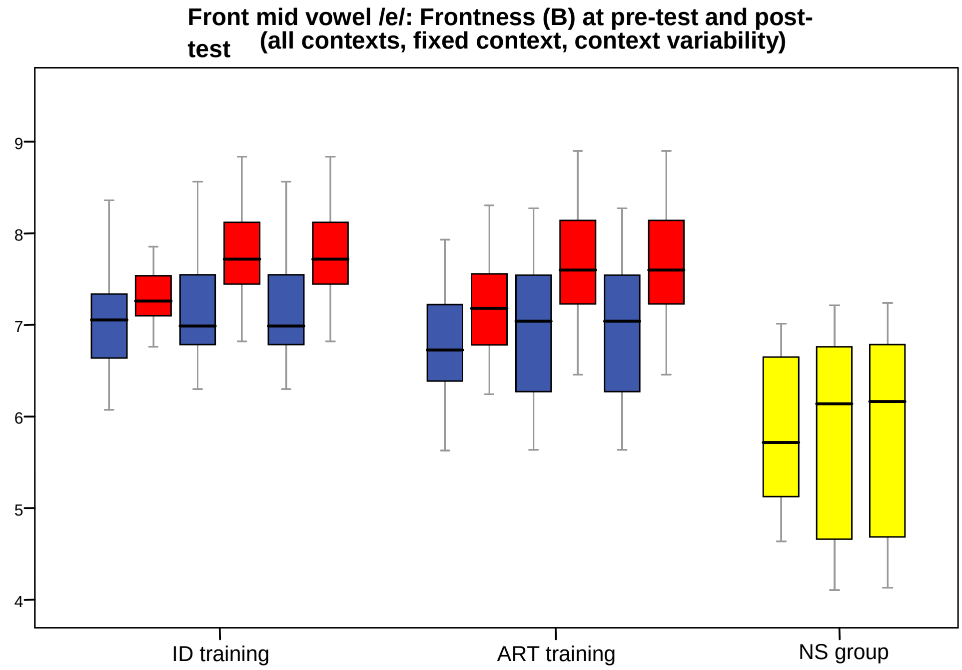


Figure 5.35. Pre-test and post-test height (Bark) values of English /e/ produced in all conditions, fixed context and context variability conditions, by the ID and ART training, and the NS group.

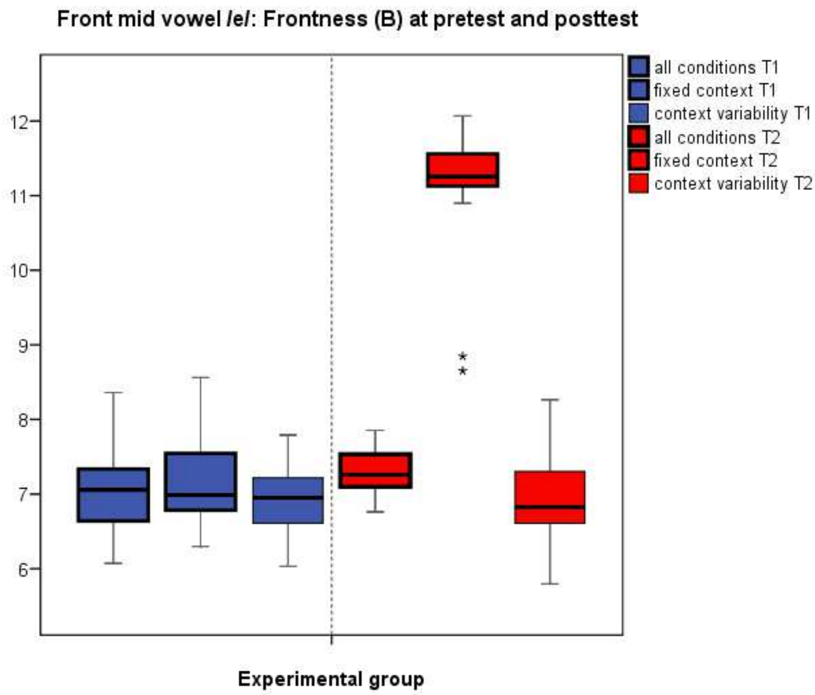


Figure 5.36. Pre-test and post-test frontness (Bark) values of English /e/ produced by the training group (*N*= 64) at pre-test (blue) and post-test (red) in all conditions, fixed context and context variability conditions.

Summary

Although the effect of training and type of training on vowel frontness was not significant, there was some indication of changes in the position of the tongue during the articulation. After training, there was a significant change in the articulation of high vowels towards a more fronted position and non-high vowels towards a more retracted position, suggesting that both training methods were effective in changing vowel articulation with respect to the frontness dimension. Particularly, there was a large effect size of training on the articulation of /e/, becoming significantly fronted vowel after training. For high vowels, there was a significant effect of training on vowel frontness: /iːeʊuː/were significantly more fronted as a result of training. In the fixed context condition, /iː/ was significantly more fronted by the ART training group, but the effect of type of training did not reach significance. After training, high vowels /iːeʊuː/ were generally slightly fronted after training), whereas low vowels /æʌɑːɒɔː/ and mid-central vowels /ɪɜː/ became more retracted vowels.

5.4. Summary of the production results

In the present chapter we have examined vowel production data from a group of Catalan-Spanish speakers who had been found to rely on (secondary) temporal cues rather than (primary) spectral cues to produce English vowels prior to training. The task used to access vowel production data in this study was a 30-minute production task based on delayed repetition consisting of 160 CVC natural words in isolation for repetition, in fixed context and context variability conditions, produced by 2 new talkers (1 male, 1 female).

In order to assess accuracy of the vowel productions before and after training, the main effects of *training* (T1 vs. T2) and *type of training* (ID vs. ART) were first reported for three production measures: spectral distances (SD) (of /iː/-/ɪ/,/ɪ/-/e/, /e/-/ɜː/, /æ/-/ʌ/,

/ʌ/-/ɑː/, /æ/-/ɑː/, /ɒ/-/ɑː/, /ʌ/-/ɒ/, /ɒ/-/ɔː/, /ʊ/-/uː/), and tense-lax duration ratio (DRs) of 6 tense-lax vowel contrasts (/iː/-/ɪ/, /iː/-/e/, /e/-/ɜː/, /æ/-/ɑː/, /ʌ/-/ɑː/, /ʊ/-/uː/). Next, training-related changes in frontness (Bark) and height (Bark) were explored at the end of the chapter.

The most important finding of this chapter was that the effect of *training* on the SDs between non-native vowels (RQ1.4.1). There was a significant effect of *vowel contrast* on SDs as well as a significant *vowel contrast x training* interaction indicating that the amount of spectral distance changes from pre-test to post-test varied significantly as a function of the vowel contrast. Within the high-front vowel area, the SDs increased for /iː/-/ɪ/ (T1: M=2.35 vs. T2: M=2.60; p<.05) and /e/-/ɜː/(T1: M=1.64 vs. T2: M=1.81; p<.05) but decreased for /ɪ/-/e/(T1: M=2.24 vs. T2: M=1.80; p<.001) after HVPT. In the low and back vowel areas of the mouth, SD remained somehow steady for /æ/-/ʌ/, /ʌ/-/ɑː/,/ɒ/-/ɑː/, /ʌ/-/ɒ/, /ɒ/-/ɔː/ and decreased for /æ/-/ɑː/ (T1: M=1.49 vs. T2: M=1.33; p<.05) and /ʊ/-/uː/(T1: M=0.92 vs. T2: M=0.81; p>.05). There was a significant *vowel contrast x type of training* interaction suggesting that the type of training had a significantly different effect on SD depending on the target vowel contrast. The results revealed a significant effect of test condition on SDs, suggesting that changes in SDs from pre-test to post-test were more visible in the fixed context condition than in the context variability condition. It is also interesting to note that there was a significant effect of vowel set on SDs. After training, higher SDs were obtained between high-front vowels than between low and high-back vowels. Although the effect of *type of training* on SD was not significant, the amount of SD changes obtained for high-front and low vowels was more salient for the ART group than for the ID group.

The results revealed that, in the fixed context condition, the target contrast /ɪ/-/e/ underwent the highest amount of spectral change after ID and ART training, which in turn affected the SD of the target vowels within the same subset, /iː/-/ɪ/ and /e/-/ɜː/. Despite not finding a significant effect of *type of train*ing, this chapter outlined salient between-group differences with respect to the amount of significant spectral changes in vowel production (RQ2.4a):

- After ART training, the /iː/-/ɪ/ and /e/-/ɜː/contrasts were produced with significantly higher SD in relation to pre-test, whereas the SDs of /ɪ/-/e/ and /æ/-/ɑː/ were significantly shorter in the productions. After ID training, the SDs of /ɪ/-/e/ decreased significantly, but other changes in high-front or low vowel contrasts were not significant.

- The amount of significant differences in the SD between vowels was greater in the ART training group than in the ID group.

- The findings suggested that ART training somehow had a larger impact on the whole vowel system, esp. when vowels are embedded in a fixed context (/h_d/).

The second section of this chapter reported a significant effect of *training* and a significant effect of *vowel contrast* on tense-lax duration ratios (DRs) (RQ1.4.2). After training, DRs of ID and ART training groups were higher than at pre-test for the six tense-lax vowel contrasts (/iː/-/ɪ/, /iː/-/e/, /e/-/ɜː/, /æ/-/ɑː/, /ʌ/-/ɑː/, /ʊ/-/uː/). DR varied significantly as a function of the vowel contrast. Although there was a non-significant effect of *type of training*, the chapter outlined some salient between-group differences (RQ2.4b):

- First, the ART group showed significantly higher tense-lax DRs for all tense-lax contrasts.
- However, the ID training showed increased DR for all vowel contrasts but changes in DR from pre-test to post-test did not reach significance.

To sum up, a significant effect of *training* and tense-lax *vowel contrast* on DR was reported, besides a significant *contrast x type of training* interaction on DR. Overall, the DR of tense-lax contrasts was greater at post-test regardless of the *type of training* trainees underwent. In particular, ID and ART training groups differed for the high-back /uː/-/ʊ/ contrast.Furthermore, the analyses of mean duration scores (msec.) helped to explain changes in DR ad revealed two different patterns in the use of the duration cue for non-native vowel production according to type of training. Whereas the ID training group consistently produced both tense and lax vowels with increased vowel duration at post-test, the ART group produced the tense vowels with significantly longer duration ($p < .05$) (T1: 349; T2: 367) and the lax vowels with slightly shorter duration (T1: 188; T2: 183).

Finally, the third section of this chapter described some changes in the height and frontness dimensions of vowels produced by the ID and ART training groups to better understand spectral changes in vowel production (RQ1.4.3). Although the effect of training and type of training were not significant, there was some indication of changes in the position of the tongue during the articulation. Overall, ID and ART training groups produced higher vowels after training. Particularly, /iː/ and /e/ were produced significantly higher in the spectrum at post-test, in *fixed context* and *context variability* conditions, by both ID and ART groups, suggesting that both training methods were effective in changing vowel articulation with respect to the height dimension. However, /uː/ in *fixed context* was significantly higher after training for the ART, but changes did not reach statistical significance for the ID group. Along the height dimension, there was a significant change in the articulation of high vowels towards a more fronted position and non-high vowels towards a more retracted position (i.e. vowels were produced with higher or lower height values depending on vowel type), suggesting that both training methods were effective in changing vowel articulation with respect to the frontness dimension. After training, high vowels /iːeʊuː/ were generally produced with higher frontness values (i.e. slightly fronted vowels after training), whereas low vowels /æʌɑːɒɔː/ and mid-central vowels /ɪɜː/ became more retracted vowels. Particularly, there was a large effect size of training on the articulation of /e/, indicating a significantly fronted vowel after training. For high vowels, there was a main effect of *training* on the degree of frontness: /iːeʊuː/were significantly more fronted as a result of training. At post-test, high vowels were slightly more fronted for the ART group than for the ID group and, in the *fixed context* condition, /iː/ was significantly more fronted ($p= .006$) by the ART training group, but the effect of type of training did not reach significance (RQ2.4c).

PART III

DISCUSSION AND CONCLUSIONS

Chapter 6

General discussion

The purpose of this thesis was to evaluate two training methods that might be used to improve learners' perception and articulation of the 11 English RP monophthongal vowels (/iː ɪ e ɜː æ ʌ ɑː ʊ uː/) by bilingual Catalan-Spanish learners of English. Two groups of learners of English (*N*=64) were assigned to different types of audiovisual (AV) High Variability Phonetic Training (HVPT) and exposed to natural CVC words from multiple talkers through either identification (ID) or articulatory (ART) training. Whereas the ID training required learners to focus on audiovisual cues to perceptually identify the vowel category within a vowel subset, ART training required learners to do so to produce the vowels. Auditory feedback provided assistance in correcting identification, or to change erroneous articulations. This thesis compares the effects of perceptual and production-based HVPT on L2 learners' perception and production of the fullset of English monophthongs.

The aim of this chapter is to offer an interpretation and comprehensive discussion of the perception and production results presented in Chapters 4 and 5.

The most important finding of the study was that the two HVPT groups showed accuracy gains in vowel perception and less error dispersion per vowel. Both ID and ART training methods were effective in leading to significant formant movement for some vowels, which were produced with less spectral overlap at post-test.

The first research question (RQ1) investigated whether audiovisual HVPT was an effective method to improve L2 vowel perception and production. The second research question (RQ2) assessed differential gains in L2 vowel perception and production as a function of *type of training* (Identification vs. Articulatory training).

The discussion of the findings is divided into the discussion of the perception results (RQ 1.1, RQ 1.2, RQ 1.3; RQ 2.1, RQ 2.2, RQ 2.3) and the discussion of the production results (RQ 1.4 and RQ 2.4). Sections 6.1.1 and 6.1.2 are organized around the two main research questions, effects of training (RQ1) and effects of type of training (RQ2).

6.1. Effects of ID and ART training on Vowel Perception

This thesis investigates if audiovisual HVPT leads to changes in the Catalan-Spanish speakers' perception of English vowels (RQ 1.1, 1.2 and 1.3), and whether there are any differential gains in L2 vowel perception as a function of *type of training* (Identification vs. Articulatory training) (RQ 2.1, 2.2 and 2.3).

The results from the pre- and post-test **vowel identification task (ID1-task1)** have demonstrated that the two training methods were successful in improving listeners' ability to identify English vowels (RQ 1.1). Vowel identification scores at post-test were significantly higher than at pre-test, suggesting that ID and ART training methods were generally effective in improving vowel identification accuracy. Trainees showed significant improvements in vowel identification of 17.63% (ID training) versus 12.88% (ART training). ID and ART types of training were similarly effective in improving vowel identification accuracy (RQ 2.1).

RQ1.1.1a investigated whether the effect of training and identification gains varied as a function of ***vowel set***. There was a significant improvement in vowel identification of 10-18% for the three vowel sets, so improvement in the identification of natural vowels was generalized to high-front, low and back vowels. However, results indicated that the *vowel set* seems to play a role in the overall training effect. Both ID and ART types of training seemed to have a more visible impact on the high-front and low vowel sets, whereas back –especially high-back vowels- were less accurately identified and probably more resistant to training. One explanation offered for this is that, being both types of training audiovisual, vowels in a more retracted position, that is, with a high degree of "backness", were less visually salient than front vowels, for instance, in both types of training. Rounded vowels articulated at the back of the tongue such as /ʊ/ and /uː/ were probably less visually salient than front vowels like /iː/ and /e/, for which mouth openness, lip position or tongue movement were more visually salient in these two types of HVPT. ID and ART types of training did not differ with respect to vowel identification by vowel set (**RQ2.1a**).

RQ1.1.1b investigated whether the effect of training and identification gains varied as a function of ***vowel.*** Both types of training improved learners' identification performance about 14.87% and learning generalized to the 11 English vowel monopthongs. However, some English vowels were thus more accurately perceived than others after training (Elliot 1995, 1997). Post-test identification performance revealed that the lowest percentage of identification after training was found high back vowels /ʊ/ and /uː/ were the most confused at post-test. The best post-test performance was found for the mid vowels /e/ and /ɜː/, a notable improvement was seen for /ɑː/ and /ɪ/, and only moderate improvement was reported for /ɔː/ and /ʊ/. Trainees improved considerably for /uː/, /ɜː/, /ɪ/ and improved about 12% points for /æ/, /ʌ/ and /ɑː/. No evidence of a significant advantage for one of the training groups was shown in vowel identification at post-test, except for the fact that the ID training group identified /e/ significantly better than the ART training group (**RQ2.2b**).

Two tentative explanations are offered to account for the fact that the 11 vowels were identified at different accuracy rates. First, the varying degree of visual saliency of vowels in relation to point of articulation (degree of frontness-backness) seemed to be an explanation why /ɒɔːʊuː/ were still less accurately perceived than /iːɪeɜː/ at post-test. However, the fact that the vowel least accurately perceived (/uː/) received the highest percentage of gains confirms the efficiency of the two training methods. This is in support of Ortega-Llebaria et al. (2001), which suggests that audiovisual training could only be effective more effective in improving vowel perception when vowels with very distinctive visual feature (very dissimilar lip configurations) were paired and presented to learners. It is hypothesized that the physical differences in the articulation of the different vowel stimuli can affect the visual saliency of stimuli and might determine, therefore, the efficiency of AV training for improving the perception of a large range of different vowels. Secondly, the different degree of (dis)similary between L1 and L2 vowel systems might have played a role in the success with which L2 vowel monophthongs are identified after training. It should be notice the fact that, at pre-test /i/ was more accurately identified than /ɪ/, which could probably due to the fact that Catalan/Spanish /i/ is spectrally closer to the English /iː/ than /ɪ/, but /ɪ/ is more accurately perceived than /i/ after ID and ART training. Third, high-back vowels are much harder to learn than high-front and low vowels due to the extremely scarce minimal instances than can be found. These may serve as explanations of why HVPT provides *different learning effects* to *different vowels*.

RQ1.1.1c explored whether the effect of training and identification gains varied as a function of *stimulus presentation* (auditory vs. audiovisual). Audiovisual HVPT significantly improved English learners' identification performance in auditory and audiovisual conditions, which meant that training successfully trained Catalan-Spanish learners of English to integrate auditory and visual cues in vowel perception for accurate identification. This findings corroborate the previous findings concerning the effectiveness of visual cues in improving the perception of British English vowels by Spanish learners of English (Ortega-Llebaria et al., 2001) or other learners (Akahane-Yamada et al., 1997; Hardison, 2000, 2003; Hazan et al., 2006; Hazan et al., 2002; Hardison, 1999, 2006). Both ID and ART training groups performed similarly better in both modalities at post-test and both performed better with audiovisual vowel stimuli than auditory stimuli at post-test **(RQ2.2c)**. Although ID trainees obtained higher gains than ART trainees in audiovisual and auditory modalities, differences between the types of training were not considered significant.

RQ1.1.1d explored whether improvement in vowel identification be generalized to *new words* and *untrained talkers*. The findings point to a successful transfer of learning to the perception of the same vowel in new words and untrained talkers. Generalization of improvement to untrained words and untrained talkers is a valuable finding of ID and ART types of HVPT. It must noticed, though, that the ID training group outperformed the ART group in the identification of *trained words* at post-test. ID (improvement of 18%) and ART (improvement of 9%) training groups differed in the amount of improvement they made with respect to trained words (**RQ2.2d**). After ID training, learners identified trained words more accurately than new words, whereas ART trainees performed similarly in the two word type conditions. It is hypothesized that trainees might be more likely to remember words previously included in identification tasks used for training (ID training) than words that they had been previously asked to imitate during the articulatory training sessions

(ART training). Bearing in mind that ART trainees perceived similarly better the vowels in trained and new words, and also performed similarly with trained and untrained at post-test, it is speculated that articulatory training could be even more efficient than ID in promoting generalization to new words and contexts due to the self-corrective and self-awareness nature of its feedback component, which involves continuous comparison of the trainees' vowel productions against the NS target. The fact ID trainees also outperformed the ART training for trained words and trained talkers makes us thing that type of corrective feedback administered through perceptual training might have played a role in the degree of success in the trained and trained talkers.

RQ1.1.2 investigated whether HVPT was effective in reducing *mean error dispersion* in vowel identification. The findings revealed a number of differences between ID and ART training groups with respect to the degree of error dispersion in vowel identification (**RQ2.2f**). The ID training group obtained lower error dispersion than the ART group despite overall changes in the distribution of error responses reported for both groups, more concretely, ID and ART groups showed a significantly different pattern of confusions when they identified /iːeæuː/ at post-test. In summary, the ID training group experienced a significantly higher decrease in the number of mistaken "similar" categories assigned to certain vowel stimuli at post-test than the ART group did. Three explanations are offered to account for this difference between training methods. First, the fact tht the ID training was based on an identification task may have contributed to increasing the degree of perceived dissimilarity between the English vowels offered as possible responses better than the ART training (responses were based on three vowel sets: high-front, low and high-back vowels). Another possible explanation for the fact is the similarity between the categorization task used for ID training and testing, which may have led to familiarity effects. Finally, this result simply raises a question about the possible benefits of identification training for better reducing the kind of confusions appearing in vowel identification. The results also suggest that perceptual training may be highly effective and help learners distinguish ambiguous vowel sets that will trigger less possible response categories in vowel identification experiments and therefore better select the response.

The results from the pre- and post-test **identification task with duration-manipulated vowels (ID2-task2)** have demonstrated that the two training methods were successful in improving listeners' ability to identify synthesized English vowels (**RQ 1.2**), and helping learners reduce reliance on duration cues and increase attention to spectral cues for accurate identification of duration-manipulated *tense* and *lax* vowels (**RQ1.2.1**). The he fact that ID and ART groups identified synthesized vowels significantly better after training and showed lower duration effect scores for 6 of the 8 synthesized vowels (/iɑɪæʌʊ/) is meaningful because the findings suggest that:

(6) The two types of high-variability training facilitate and contribute to the formation of robust vowel phonetic categories, as the results demonstrate that learners were able to apply what they have learned to vowels that had been durationally manipulated. In other words, they were capable of generalizing their learning with natural vowels to vowels with manipulated duration. Ten sessions of HVPT training (with high variability of talkers, contexts, vowels, and words and high-quality input) might have contributed to making the application of existing or learned categories to durationally-manipulated vowels more automatic and efficient.

(7) Given the Catalan-Spanish learners' over-reliance on the duration cue, ID and ART training can shift language learners' attention to weight more relevant cues (spectral cues) when the duration cue is ambiguous (e.g. "short" /iːɑː/ and "long" /ɪæʌʊ/). Learners can benefit from high-variability training without cue-manipulated stimuli in order to obtain a more accurate perception of vowel pairs, based on more relevant cues. The findings suggest that high-variability training may be needed to focus

learners on how efficiently use the spectral cues and diminish the listeners' over-sensitivity to duration differences. However, there was no clear advantage of one type of training over the other **(RQ2.2)**.

(8) The improved performance on duration-manipulated tokens at post-test suggests that training on natural and variable tokens (focusing on spectral differences), including the whole vowel category inventory, can be more effective or just as effective as training on both quality and duration differences.

(9) Improved perception on durationally manipulated vowels is probably crucial for L2 production as well, in order to achieve a native-like production of L2 vowels in the long term.

The results from the pre- and post-test **discrimination task (DIS-task3)** have demonstrated that the two training methods were successful in improving that the two training methods were successful in improving listeners' ability to distinguish contrasting English vowels (**RQ 1.3**). Discrimination scores were significantly higher at post-test than at pre-test, for all vowel sets (high-front, low, high-back), vowel contrasts (16), and test conditions (fixed context and context variability). Trainees showed significant improvements in vowel discrimination (ID: 12.72%; ART : 15.09%), and both types of training were similarly effective in improving vowel discrimination accuracy (**RQ 2.3**).

Importantly, the findings revealed that short-term high-variability training (based on identification or imitation), not based on a discrimination task, significantly improved vowel discrimination. The fact that the training included multiple talkers besides context variability in the stimuli might have a played a crucial role in discrimination improvement when discriminating vowels produced by untrained talkers in varied contexts.

It must be noted that the effect of training on discrimination accuracy varied as a function of *vowel set* (**RQ1.3a**). ID and ART training groups showed a similar pattern of discrimination results at pre-test and post-test (**RQ2.3a**). Contrary to identification results, the back vowel contrasts (/ɒ/-/ɔː/, /ʊ/-/uː/, /ɔː/-/uː/) got the highest accuracy in discrimination, followed by high-front (/iː/-/ɪ/, /ɪ/-/e/, /e/-/ɜː/, /ɪ/-/ɜː/), and low (/e/-/æ/,/ɜː/-/æ/, /æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɜː/-/ʌ/, /ɜː/-/ɑː/, /ɒ/-/ɑː/, /ʌ/-/ɒ/) contrasts before and after training.

The effect of training on discrimination accuracy varied as a function of *vowel contrast* (**RQ1.3b**). Trainees listeners still had great difficulty at post-test with contrasts such as /ɜː/-/ɑː/ and /ɒ/-/ɑː/ (<70%) and could only partially distinguish the /iː/-/ɪ/ contrast well. However, trainees performed at ceiling (>95%) for the high-front /ɪ/-/ɜː/ contrast and four low contrasts (/ɜː/-/æ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɜː/-/ʌ/). On average, listeners distinguished the other contrasts relatively well (80-90%): /ɪ/-/e/, /e/-/ɜː/, /e/-/æ/, /æ/-/ʌ/, /ʌ/-/ɒ/, /ɒ/-/ɔː/, /ʊ/-/uː/. ID and ART groups did not differ with respect to discrimination by vowel contrast (**RQ2.3b**).

Finally, the effect of training on discrimination varied as a function of the *test condition* (**RQ1.3c**). For both ID and ART groups different contexts lowered discrimination performance. No differences were found between ID and ART groups, who performed significantly better and obtained grater discrimination gains in fixed context than in context variability conditions after training (**RQ2.3c**). The reason for this is that context variability makes it more difficult for some listeners to make reliable quality distinctions and distinguish spectral differences as accurately as for vowels in isolation (Kewley-Port, 1995, 2001). Moreover, vowel discrimination is generally affected by the nature of the surrounding phonetic context.

As shown by results, listeners' ability to make reliable vowel quality distinctions is lowered when the vowels are placed in different consonant contexts rather than in a fixed /h*V*d/. In fact, post-test performance decreased by 31.65% for ID and 30.03% for ART with context variance.

The following results confirm that audiovisual identification and articulatory types of training (HVPT) are effective and useful paradigms to promote higher accuracy in L2 vowel perception:

(1) Trainees improved their perception of L2 vowels generally regardless of the high-variability training method (ID vs. ART). This finding corroborates previous findings showing that perceptual learning can occur in non-naturalistic environments (i.e., in a FL classroom, in a phonetic laboratory) by means of specific **high variability** –perceptual or articulatory– training (Logan et al. 1991; Lively et al,. 1993; Bradlow et al., 1997; Iverson et al, 2005; Iverson & Evans, 2009; Lambacher et al,. 2005; Nishi & Kewley-Port, 2007, 2008; Wang, 2002; Wong, 2003), within a short period of time. It also supports previous findings on the **malleability of L2 adult learners' perceptual systems**. HVPT seems to have had an immediate impact on vowel categorization and may have favoured automatic processing. We believe that training may have helped learners automatize and establish clearer perceptual routines for successful vowel identification through variability training. Improvement in labeling ability suggests that some change in the representations of L2 categories is possible through intensive training, as evidenced in (a) overall better identification and discrimination abilities (better performance on ID1-task1, ID2-task2 and DIS-task3), and (b) lower error-dispersion in ID1-task1.

(2) **There was an interaction between the effect of training and the effect of vowel.** Some vowels did not improve as much as other despite overall improvement in vowel identification and discrimination. The data suggests that learners, at least after 10 training sessions, are malleable in terms of perception for most phones. The fact that accuracy in vowel categorization differs according to the vowel suggests that 10 AV training sessions may be insufficient to train certain target vowels which seem more difficult and less visually salient in terms of articulation to learners. The data also

confirmed the claims of previous studies showing that some L2 phones responded to treatment better than others, which were more resistant to short-term instruction (Elliot 1995, 1997). High back vowels were the most confused vowels and identified less successfully than the rest at post-test (/ʊ/: 56.25%; /uː/: 66.60%); low vowel contrasts (/e/-/æ/,/æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/, /ɜː/-/ʌ/, /ɒ/-/ɑː/, /ʌ/-/ɒ/) were discriminated more poorly than the rest at post-test.

(3) **Perception gains as a transfer from production training**

The finding of a transfer of articulatory training to improved vowel identification and discrimination suggests that the learning and practice of articulatory patterns and articulatory gestures also helps to alter or establish perceptual representations. This suggests that changes and improvement in vowel perception can be related to changes in vowel production. The findings in this study support the findings of previous training studies showing that training in production leads to improvement in perception (Eg. Lacabex et al., 2009). Learning may be **robust** enough to transfer to untrained modalities. Our data suggests that 10 intensive training sessions based on imitation and self-correction during articulatory training lead to cross-over effects in perception and production

(4) **Improvement generalizes to synthesized vowels.** Results generalize to synthetic speech (ID2-task2), which further indicates successful learning. Ten sessions of HVPT can improve identification consistency for L2 vowel categories. Trainees showed greater reliance on vowel quality rather than duration in synthesized vowel categorization, greater proportion of /iː, ɪ, æ, ʌ, ɑː, ʊ, uː/ responses regardless of duration manipulation, and **less reliance on duration** in the identification of tense and lax vowels at post-test.

(5) Learning may be robust enough to generalize to more sounds, new talkers, contexts and tasks. This transfer of learning supports the efficacy of ID and ART types of training for higher accuracy in vowel identification and discrimination.

6.2. Effects of ID and ART training on Vowel Production

This thesis also investigates if audiovisual HVPT leads to changes in the Catalan-Spanish speakers' production of English vowels (**RQ 1.4**), and whether there are any differential gains in L2 vowel production gains as a function of *type of training* (Identification vs. Articulatory training) (**RQ 2.4**). Improvement in vowel production could be observed as changes in the place of articulation of vowels, particularly in the spectral distances between contrasting vowel pairs and the duration ratios of tense-lax vowels.

This study aimes at complementing previous perceptual training studies showing their effects on production (Bradlow et al., 1999; Lengeris & Hazan, 2010; Wong, 2013, 2014, 2015) or previous production studies (Kartushina et al., 2015) by showing what aspects of the English vowel production the subjects have changed after HVPT training.

We tested for changes in vowel production accuracy and found that HVPT was effective in making the production of contrastive vowels more distinct. This suggests a conscious attempt of learners to produce more acoustically distinct vowel quality targets, with less spectral overlap in their non-native vowel systems. Among the most important findings in production, we found that HVPT had a significant impact on the articulation of three high-front vowel contrasts and one low vowel contrast. Immediately after training, spectral distances increased for /iː/-/ɪ/ & /e/-/ɜː/, decreased for /ɪ/-/e/ and increased for /æ/-/ɑː/. These findings suggest that only 10 intensive sessions of AV HVPT can produce significant changes in the vowel quality of certain target vowels without explicit instruction about tongue or lip position, openness or duration information, without manipulated or enhanced stimuli in training, through ID or ART training.

The effect of HVPT on SDs between non-native vowels was significant. This study showed that the amount of spectral changes from pre-test to post-test varied significantly as a function of the target vowel contrast (**RQ1.4.1c**) and vowel set (**RQ1.4.1b**).Within the high-front vowel area, the SDs increased for /iː/-/ɪ/ and /e/-/ɜː/ but decreased for /ɪ/-/e/ after HVPT. In the low and back vowel areas of the mouth, SD remained somehow steady for /æ/-/ʌ/, /ʌ/-/ɑː/,/ɒ/-/ɑː/, /ʌ/-/ɒ/, /ɒ/-/ɔː/ and decreased for /æ/-/ɑː/ and /ʊ/-/uː/. The results also revealed that changes in SDs from pre-test to post-test were more visible in the fixed context condition than in the context variability condition (**RQ1.4.1a**), and that there was a significant effect of vowel set on SDs (**RQ1.4.1b**). After training, higher SDs were obtained between high-front vowels than between low and high-back vowels. In the fixed context condition, the target contrast /ɪ/-/e/ underwent the highest amount of spectral change after ID and ART training, which in turn affected the SD of the target vowels within the same subset, /iː/-/ɪ/and/e/-/ɜː/.

Despite not finding a significant effect of *type of training* on mean SDs , we observed ssalient between-group differences with respect to spectral distance changes in vowel production (**RQ2.4a**):

(1) First, the findings suggested that **ART training had a larger impact on the whole vowel system**, especially when vowels were embedded in a fixed context (/h_d/). After ART training, the SDs of /iː/-/ɪ/ and /e/-/ɜː/contrasts increased significantly, whereas the SDs of /ɪ/-/e/ and /æ/-/ɑː/ were significantly shorter. After ID training, the SDs of /ɪ/-/e/ decreased significantly, but other changes in high-front or low vowel contrasts were not significant. The amount of significant differences in the

SD between vowels was greater in the ART training group than in the ID group (**RQ2.4a**).

(2) Second, the analyses of tense-lax duration ratios (RT) revealed **two consistently different patterns in the use of the duration cue for non-native vowel production as a function of type of training** (**RQ2.4b**). The ART group showed significantly higher tense-lax DR for all tense-lax vowels at post-test. The ID training showed increased DR for all tense-lax contrasts but changes in DR from pre-test to post-test were not significant. Moreover, the analyses of the duration measure revealed crucial information about the use of duration by trainees: whereas the ID group consistently produced both tense and lax vowels with increased vowel duration at post-test, the ART group produced the tense vowels with significantly longer duration and the lax vowels with slightly shorter duration, closer to native-like patterns in vowel production.

(3) Finally, when the training methods were compared in terms of changes in the height and frontness dimensions of vowels, high vowels were observed to be slightly more fronted for the ART group than for the ID group and, in the *fixed context* condition, /iː/ was significantly more fronted ($p = .006$) for the ART training group, but the effect of type of training did not reach significance./uː/ in *fixed context* was significantly higher after training for the ART, but changes did not reach statistical significance for the ID group (**RQ2.4c**).

In conclusion, HVPT had a significant effect on the pronunciation accuracy of the target vowels, particularly in terms of vowel quality, measured through SD and measures such as height and frontness. These results show that **10 hours of AV HVPT are effective in making confusable L2 vowels more distinct in production**. Although 10 training sessions are insufficient to immediately achieve native-like vowel production, a clear trend is observed towards improvement in production accuracy for some vowels, in terms of increases and decreases in SDs in the L2 vowel system. This trend is clearer in the case of ART trainees and this is due to the type of articulatory training used, consisting of visual stimuli and feedback based on imitation, repetition, self-correction, NS-NNS comparisons, error detection, requiring awareness of NS-NNS dissimilarity and continuous articulatory adjustments and to get close to the target sound.

On the other hand, **some gains in production are observed as a transfer from identification training.** ID training may be effective in raising awareness of L1-L2 vowel dissimilarity and in directing learners' attention to spectral cues in L2 vowel production. This findings is in accordance with previous research that suggests that high-variability perceptual learning can be transferred to the production domain and subjects could generally improve their production in terms of vowel quality (Wong, 2013). However, it must be highlighted that the ART training was considered slightly superior to ID training with respect to spectral chances in L2 vowel production. We hypothesize that, after ART training, the awareness of the spectral and temporal differences among L2 vowels was better raised. The fact that learners were more sensitive to ART training (i.e. more significant SD changes were found) suggests that ART practice may help learners to overcome the limitations that perception might impose on production.

Our data supports the claim that stimulus variability (NS voices, contexts and task conditions) plays a crucial role in learning. Stimulus variability reflects the reality outside the FL classroom and raises the students' awareness of vowel quality differences, contributing to the formation of new phonemic categories for non-native sounds. Although there is not conclusive evidence that actual reweighting occurs due to training, the findings confirm that exposure to phonetic categories and practice have created modest changes in the L2 learners use of spectral and duration cues. These changes revealed a conscious attempt of learners to produce acoustically distinct vowel quality targets for L2 vowels, with a great deal less spectral overlap. The results demonstrated that both HVPT approaches offered a type of learning that allows learners to focus their attention on phonetic information (Wong, 2014), which is different from what is learned in an FL classroom.

Chapter 7

Conclusions

In this final chapter, the concluding remarks to this dissertation are provided. We will begin with recapitulating the main findings of the study, and then discuss their implications for L2 speech learning and pedagogical implications. The last section presents the limitations of the study and some suggestions for future research.

7.1. Summary of results and contribution

This dissertation aimed at contributing to expand the research on the differential effects of perception- and production-based training methods on learners' perceptual and productive competence in the L2. It was also hoped that findings would have implications to L2 pronunciation instruction. To the author's knowledge, no study has directly compared the effectiveness of identification versus articulatory training methods on the entire L2 vowel system. The present study confirms that:

(1) First, this audiovisual high-variability training method is effective in improving non-native vowel perception and production. The data suggested that adult learners are

malleable in terms of perception and production for most L2 vowels. Therefore, the current experiment adds to the literature exploring phonetic training in adults. The current experiment adds to the literature demonstrating that L2 learners can improve their identification, discrimination and articulation of a non-native vowels phonetic contrast via computerized ID and ART phonetic training. Within the cue-weighting field, results also demonstrate that trainees learn to spectrally differentiate English vowels in perception and production.

(2) Second, identification and articulatory training methods lead to improvement in L2 perception and changes in L2 speech production to different extent. Training purely on production improves perception and production. Training purely on perception affects positively the perception and production domains but does not lead to many spectral changes in production as training purely on production does. Finally, it appears that improvement in L2 production and perception do not systematically progress at equal rates within individuals following each training method. Articulatory training seems more effective for inducing spectral changes in L2 vowel production.

(3) Our findings concurred with previous studies in that some L2 phones responded to treatment more than others (e.g. Elliot, 1995, 1997). High-front and mid vowels were generally more sensitive to either type of HVPT, whereas back, esp. high-back vowels were more resistant to training than other phones and were perceived with lower accuracy than the rest at posttest.

After training, the best identification performance was found for the mid vowels /e/ (T2: 96.42 %) and /ɜː/ (T2: 94.20%) and a notable improvement was seen for /iː/, /ɪ/ and/ɑː/. Learning generalized to the perception of synthetic speech reflected in shallower identification slopes and higher duration effect scores

(DES), especially for /iː/, /ɪ/, /ʌ/ and /ɑː/. After training, the vowel contrasts /ɪ/-/ɜː/, /ɜː/-/æ/ and /æ/-/ɑː/ (81-86%) were by far the most accurately discriminated among the 16 vowel contrasts. High back vowels were the most confused vowels at post-test (/ʊ/: 56.25; /uː/: 66.60), although it is important to note that the highest percentage of ID gains was obtained for the high-back vowel /uː/ (14.87%)

Results demonstrate that the (natural and synthetic) vowels with the highest percentage of correct identification and discrimination were generally the same phones undergoing the highest amount of spectral changes from pretest to posttest. The high-front vowel set was dramatically affected by training: spectral distances increased significantly for /iː/-/ɪ/ (T1: M=2.35 vs. T2: M=2.60) and /e/-/ɜː/(T1: M=1.64 vs. T2: M=1.81) but decreased for /ɪ/-/e/(T1: M=2.24 vs. T2: M=1.80) after HVPT. Within low vowels, SD decreased significantly for /æ/-/ɑː/ (T1: M=1.49 vs. T2: M=1.33). SDs remained steady or did not change significantly for the rest. After ART training, the /iː/-/ɪ/ and /e/-/ɜː/contrasts were produced with significantly higher SD in relation to pretest, whereas the SDs of /ɪ/-/e/ and /æ/-/ɑː/ were significantly shorter in the productions. ID and ART training led to significantly higher /iː/ and /e/.

Finally, empirical studies such as this one can help teachers and learners choose more efficient techniques for reaching their goals despite insufficient time and limited resources. Yet not only do phonetic training studies have important pedagogical implications for L2 pronunciation instruction, but the findings obtained through this research paradigm have a bearing on crucial issues in L2 speech learning. First, our findings confirm the malleability of the adult L2 perceptual system and its ability to remain adaptive to new amount of input administered in FL instruction settings or training. Second, these results suggest that quantity and quality of L2 input play a role in L2 speech learning both in perception and production. It offers controlled exposure to L2 input in FL instructional settings to enhance success and improvement in L2 speech learning.

7.2. Limitations and directions for future research

The present study is subject to a number of limitations which need to be kept in mind when interpreting the results. At least four important limitations can be identified involving the methodology of the study, the training programme and the analysis of the production data.

The first limitation is the lack of a delayed post-est that would test long-term retention of learning in L2 perception and production. Given the fact that improvement in L2 production and perception do not systematically progress at equal rates within individuals, it would have been useful to explore whether retention is found given the robustness of the effects of the training obtained (generalization to new talkers; transfer to synthetic speech; generalization to new contexts and novel tokens; identification, articulation and discrimination improvement). This was mainly due to practical reasons, as retention testing would have required following all participants time after the training had finished and that was an impossible task at the time of data collection, when the university courses had already finished. Still, future research could include a delayed post-test to assess long-term gains.

Secondly, the types of feedback administered within each training method were not entirely comparable. ID training included immediate acoustic and written feedback after each response. The type of feedback ensured that learners corrected their own mistakes and that the training sessions were adaptive to the learners' individual needs: immediate acoustic and written feedback was provided after each response. The type of feedback provided ensured learners would correct their own mistakes and that the ID training sessions were adaptive to the learners' individual needs and difficulties. Negative feedback obliged participants to try once again. Participants immediately heard the incorrect response immediately followed by the correct word played again and a written message saying *"Try again"* indicated the stimulus word would be heard again and the alternative responses would be displayed again. Trainees could hear each stimulus a maximum of four times. Conversely, ART training sessions were divided into four blocks preceded and following by feedback sessions based on self-monitoring and self-correction (*'Listen and compare'*) with the aim of activating the learners' feedback sensory systems so that rapid motor adjustments could be made (Guenther et al., 2006) after detecting salient differences between the NS production and their own output. When learners were satisfied with their own productions, they pressed *"Next word"* and continued listening to the other tokens produced in the previous block and included in that feedback section. If not satisfied after judging their own productions, they pressed *"Try again"*. Then they were presented with the same stimulus and they proceeded to imitate it, and then judge their own productions. Whereas ID training included immediate feedback with correct/incorrect judgements, ART training included feedback based on self-perception, self-monitoring and self-correction. The number of repetitions determined by negative feedback in ID and ART training could vary significantly, as several factors determined the amount of repetitions and trials during ART training, such as degree of self-awareness, motivation, commitment, self-demand and perfectionist attitude.

As regards the measurement and analysis of the production data, a NS rating of NNS vowel productions was not included as part of the analysis. In evaluating the production accuracy of our participants, we opted for an objective measure of production, exclusively based on acoustic measurements (height, frontness, duration) and derived measures such as spectral and duration distances and analyses of pretest-posttest differences. Future research could provide information on NS' perception of the quality of NNS vowels. Ideally, both objective and subjective measures should be combined to offer complementary approaches to the assessment of L2 vowel production accuracy. In fact, very few L2 production training studies have used the two approaches (e.g. Akahane-Yamada et al., 1998).

Thirdly, the limitations involving a subset of tasks (e.g. phonological awareness tasks and phonological short-term memory tasks) (Bradlow et al., 1998) that account for individual differences must be mentioned. Although trainees and controls belonged to the same context and were a group of adult bilingual Catalan-Spanish learners of English in their second and third year of a degree in English studies at the University of Barcelona (with similar AOL, similar amount of L2 experience, co and apparently high motivation), further studies are needed to address the roles of other factors that may influence L2 production skills, including individual differences in pronunciation, imitation talent, sensorimotor control, and the actual number of learners that improve in perception and production. There may be various reasons why improvement production of the trained vowels did not generalize to all vowel contrasts or did not reach native-like performance levels: the amount of training could have been insufficient, the nature of the type of feedback could have been inappropriate. For those reasons, additionally having NS judges to rate the productions for accentedness will be particularly relevant in this context.

Furthermore, another limitation is the absence of a correlation analysis to see the relation between improvement in perception and production that contribute to further explain this complex L2 perception and production link in SLA. Our findings on the transfer of improved production after ID training and improved perception after ART training suggest that learning to improve the articulation of non-native vowels results in tuning of perceptual representations for such vowels. Further analyses on the relationship between perception and production measures and gains and further *perception vs. production training studies* can help us provide us with better methods for improved L2 sound production.

Finally, we did not analyse the progress throughout the 10-day HPVT sessions due to the size of the study. It remains an empirical question whether fewer or more training sessions would have replicated the findings reported here or would have led to greater improvement in perception and production. Future studies will analyse identification scores obtained after each training sessions so as to give more information about the amount of exposure to input and practice needed for improvement, with a view to designing more effective training programs to improve L2 pronunciation.

These directions will be pursued in future studies.

7.3. Pedagogical implications: ID and ART training and L2 pronunciation

First, the results confirmed that both ID and ART HVPT methods are highly useful paradigms in improving accuracy in L2 vowel perception and production. Secondly, our findings also demonstrate that HVPT led to a generalization trained words produced by unfamiliar talkers.

In the present study, the two HVPT groups showed higher accuracy in vowel perception, but a clear advantage of the ID group was seen in the identification of trained words and a lesser degree of error dispersion per vowel. ID training based on a multiple-choice categorization task with immediate AV feedback was found to be a more effective approach than the ART training in decreasing the amount of error dispersion in vowel categorization. This implies that ID training consisting of highly variable AV stimuli including the fullset of English vowels (in different phonetic environments and produced by multiple talkers) promotes the listeners' selective attention to the primary acoustic cues which are relevant to a specific vowel sound and, moreover, it reduces the number of distinct error categories per specific vowel stimuli in categorization tasks. Although both ID and ART training methods demonstrated remarkable positive effects on the perception of the vowel fullset, the ID training was found to be more effective in ensuring a higher degree of perceptual dissimilarly among sounds within the same vowel subset (i.e. (/iː/-/ɪ/-/e/-/ɜː/; /æ/-/ʌ/-/ɑː/-/ɒ/-/ɒ/) and reducing the scope of error responses associated with one specific vowel stimuli presented within a large of pool of stimuli with wider variability.

These findings suggest that **the efficacy of the HVPT in laboratory settings can be transferred to FL instructional settings**. This study, by comparing and contrasting the ID and ART HVPT paradigms, confirms the efficacy of high-variability studies and offers more solid support to the claim that stimulus variability administered through perception and production-based training tasks can benefit both FL teachers and learners practically. The great success of generalization of the HVPT method can be attributable to the fact that, after a considerable amount of sessions, HVPT subjects start to draw from a larger pool of tokens stored in their memory during the process of vowel identification. The success of generalization of the HVPT can definitely be a strong rival against the FL classroom where L2 listeners are exposed to “a single talker” and hence develop more limited an shallower perceptual strategies than in a FL setting where they can be trained under high variability conditions. The present study demonstrates that HVPT be used inside the FL classroom by attributing the subjects more criteria to better categorize the vowels produced by the different speakers, and by facilitating the learners with strategies to ignore the between-speaker variability that might present obstacles to the perceptual learning. Moreover, the HVPT could provide listeners with successful perceptual strategies in identifying new words from new speakers that could be successfully transferred to production. Consequently, training under high variability conditions could contribute to more robust phonetic representations, categorization and generalization in the perception domain.

On the other hand, the findings have also provided evidence that training in perception alone can also be useful for improving the production of L2 vowel contrasts among Catalan-Spanish speakers of English, with stimulus variability playing a crucial role in the amount of spectral changes in the learners’ production of the fullset of vowels. Yet, the findings for SDs between target vowels indicate that a 10-hour HVPT may not be sufficient to promote native-like production or sufficiently high accuracy at the vowel production level. When the effects of the two HVPT methods were compared, the results confirmed that the 10 sessions of ID HVPT were not sufficient to achieve large spectral changes in the vowel system of L2 learners. On the contrary, the ART HVPT led to more robust improvement in vowel production observed in a larger spectral distances between vowels of distinct vowel subsets. The findings in vowel production demonstrated that the ART training–including a big dose of repetition and immediate feedback based on increasing the degree of self-awareness– led to a larger amount of changes in vowel articulation involving quality and temporal changes. The type of feedback administered in the ART training, which enabled subjects to attend to quality differences between NS and

NNS productions, could be used within the FL classroom. . Although there is no evidence that a reweighting of cues occurred after training, the findings revealed a conscious attempt of trainees towards more native-like productions of the target vowels. ART trainees probably had to strive to produce acoustically distinct vowel quality target for L2 vowels similar to NS targets, which explained a great deal less spectral overlap after 10 training sessions compared to ID trainees. Repetition and immediate feedback based on self-correction, self-awareness and motor practice seem to explain larger number of spectral changes in vowel articulation after ART training and could indeed be incorporated into the FL teaching methodology

To conclude, the current results in the perception and production domains advocate for incorporating HVPT methods into the learning routines in the FL classroom routines –besides the communicatively-valid and task-based materials and –in order to achieve more accurate pronunciation (Mora & Levkina, 2017). This may suggest that brief phonetic training at the segmental level will be highly beneficial to learners in FL instructional settings and useful to improve teaching programs and methods.

References

Akahane-Yamada, R. (1996). Learning non-native speech contrasts: What laboratory training studies tell us. *Journal of the Acoustical Society of America*,100 (4, Pt. 2), 2728.

Akahane-Yamada, R., McDermott, E., Adachi, T., Kawahara, H., and Pruitt, J.-S. (1998). Computer-based second language production training by using spectrographic representation and HMM-based speech recognition scores, *Proc. Interspeech 5*, 1–4

Aliaga-García, C. (2007). The role of phonetic training in L2 speech learning. In proceedings of phonetics teaching & learning conference, *PTLC 2007* (pp.1-5). London, UK: PTLC.

Aliaga-García, C. (2009). Effects of audiovisual auditory and articulatory training on second- language (L2) vowel category learning. *The Journal of the Acoustical Society of America, 125(4),* 2765-2765. doi:10.1121/1.4784692

Aliaga-García, C. (2010). Measuring perceptual cue weighting after training: A comparison of auditory vs. articulatory training methods. In *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 12-18). Poznań, Poland: Adam Mickiewicz University press.

Aliaga-García, C. (2016). Vowel learning: Perceptual versus articulatory training. *Paper presented at the 8th international conference of language acquisition.* University of the Balearic Islands, Palma de Mallorca, Spain

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 Recent research in second language sound perception and production. phonetics/phonology: Perception and production, 2-31.

Aliaga-García, C., Mora, J. C., & Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics, 47, 1-14*. doi:10.2478/psicl-2011-0002

Alivuotila, L., Hakokari, J., Savela, J., Happonen, R. P., & Aaltonen, O. 2007. Perception and imitation of Finnish open vowels among children, naïve adults, and trained phoneticians. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 361-364), Saarbrucken, Germany.

Álvarez González, J. A. (1980). Vocalismo español y vocalismo inglés [Spanish and English vowels]. Doctoral thesis. Universidad Computense de Madrid.

Bamiou, D.-E., Musiek, F. E., and Luxon, L. M. (2003). The insula (Island of Reil) and its role in auditory processing. *Brain Res. Rev.* 42, 143–154.

Banks B., Gowen E., Munro K. J., Adank P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *J. Acoust. Soc. Am.* 137 2015–2024. 10.1121/1.4916265

Barry, W. J. (1989). Perception and production of English vowels by German learners: Instrumental-phonetic support in language teaching. *Phonetica: International Journal of Speech Science, 46(4),* 155–168.

Berga, M., Cots, J. M., Escobar C., Figueras, N. & Gómez, P. 2008. *Estudi sobre les mesures adequades per aconseguir una millor integració i presència de la llengua anglesa en l'activitat acadèmica del sistema universitari català.* DURSI, Generalitat de Catalunya.

Best C.T., McRoberts G.W., Sithole N.M. (1988) Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *J Exp Psychol 14:* 345–360.

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 171-204). Timonium, MD: York Press.

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

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 13-34). Amsterdam: John Benjamins.

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

Bever, T. (1981). Normal acquisition processes explain the critical period for language learning. In K. Diller Newbury House (Ed.), *Individual Differences in Language Learning Aptitude,* Rowley, MA, pp. 176– 198.

Birdsong, D. (1999). Introduction: Whys and why nots of the Critical Period Hypothesis. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis.* Mahwah, 1-22.

Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.6.05) [Computer program]. Retrieved June 2, 2007, from http://www.praat.org/.

Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics, 11(3),* 303–328.

Bohn, O., & Flege, J. E. (1993). Perceptual switching in Spanish/English bilinguals: Evidence for universal factors in stop voicing judgments. *Journal of Phonetics, 21,* 267-290.

Bohn, O., & Flege, J. E. (1997). Perception and production of a new vowel category by adult Second Language Speech, second language learners. 3-74. doi:10.1515/9783110882933.53

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

Bohn, O.-S., Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition 14.* 131-158.

Bongaerts, T., van Summeren, C., Planken, B. & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language, *Studies in Second Language Acquisition, 19,* 447-465.

Borrell, A. (1990). Perception et (re)production dans l'apprentissage des langues étrangères. *RPA 95-96-97*, 107-114 .

Bosch L., Costa A., Sebastián-Gallés N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology, 12,* 189–221.

Bosch, L., & Sebastian-Galles, N. (2003a). Simultaneous bilingualism and the perception of a languagespecific vowel contrast in the first year of life. *Language and Speech, 46,* 217–243.

Bosch, L., & Sebastian-Galles, N. (2003b). Language experience and the perception of a voicing contrast in fricatives: Infant and adult data. *Proceedings of the 15th International Congress of the Phonetic Sciences,* Barcelona, 1987–1990.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61,* 977-985. *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech - 25*

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101(4),* 2299-2310. doi:10.1121/1.418276

Brière, E.J. (1966). Quantity before quality in second language composition. *Language Learning Research Club, University of Michigan.* 10.1111/j.1467-1770.1966.tb00816.x

Browman, C. and L. Goldstein (1986). Toward an articulatory phonology. *Phonology yearbook 3,* 219-252.

Browman, C. and L. Goldstein (1989). Articulatory gestures as phonological units. *Phonology 6,* 201-252.

Browman, C. P.,& Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston and M. E. Beckman (Eds.), Papers in *Laboratory Phonology I: Between the Grammar and the Physics of Speech. Cambridge,* U. K.: Cambridge University Press. (pp.341-376).

Browman, C. P.,& Goldstein, L. (1991). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston and M. E. Beckman (eds), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech.* Cambridge, U. K.: Cambridge University Press. (pp. 341–376).

Callan D. E., Tajima K., Callan A. M., Kubo R., Masaki S., Akahane-Yamada R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage 19* 113–124. 10.1016/S1053-8119(03)00020-X

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

Carlet, A. (2017). L2 perception and production of English consonants and vowels by Catalan speakers: the effects of attention and training task in a cross-training study. Unpublished Doctoral Dissertation. Barcelona: Universitat Autònoma de Barcelona.

Carlet, A., & Rato, A. (2015). Non-native perception of English voiceless stops. In E. Babatsouli & D. Ingram (Eds.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2015*. Chania, Greece: Institute of Monolingual and Bilingual Speech. pp. 57-67.

Carney, A.E., Wildin, G.P, Viemeister, N.F. (1977). Non-categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America. 62:* 961–970.

Catford, J. C., & Pisoni, D. B. (1970). Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal, 54(7),* 477. doi:10.2307/321767

Cebrian, J. (2002). Phonetic similarity, syllabification and phonotactic constraints in the acquisition of a second language contrast. PhD Dissertation. Toronto Working Papers in Linguistics Dissertation Series. University of Toronto, Toronto, Canada.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics, 34,* 371-387.

Cebrian, J. (2007) Old sounds in new contrasts: L2 production of the English tense-tax vowel distinction. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences.*Saarbrucken: University of Saarland, pp.1637-1640. ISBN: 978-3-9811535-1-4.

Cebrian, J. (2009). Effects of native language and amount of experience on crosslinguistic perception. *Journal of the Acoustical Society of America, Vol. 125, No. 4, Pt. 2*, April, p. 2775.

Cebrian, J., Mora, J.C. & Aliaga García, C. (June 2011) Comparing direct and indirect methods of evaluating crosslinguistic similarity. Paper presented at the *Phonetics and Phonology in Iberia conference (PaPI).* Organized by Universitat Rovira i Virgili, Tarragona, Spain.

Cebrian, J., Mora, J.C. & Aliaga García,C. (September 2010) Discrimination (Perceptual dissimilarity) of Catalan and English vowels by Catalan learners of English. Poster presented at the *6th International Conference on Langauge Acquisition (CIAL).* Organized by the University of Barcelona and Universitat Pompeu Fabra, Barcelona, Spain.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). Teaching pronunciation: A reference for teachers of English to speakers of other languages. (2nd ed). New York: Cambridge University Press.

Celce-Murcia, M., Brinton, D.M. & Goodwin, J.M. (1996). Teaching pronunciation. A reference for teachers of English to speakers of other languages. Cambridge, UK: Cambridge University Press.

Cerviño, E. and Mora, J.C. (2009) Cerviño, E. and Mora, J. C. (2009) Duration as a phonetic cue in the categorization of /iː/-/I/ and /s/-/z/ by Spanish/Catalan learners of English. *Journal of the Acoustical Society of America*, Vol. 125, No. 4, Pt. 2, April, p. 2764.

Cerviño-Povedano, E. & Mora, J. C. (2015). Spanish EFLs' identification of /i:-I/ and phonological short-term memory. *Proceedings of the 32nd International Conference of the Spanish Association of Applied Linguistics (AESLA): Language Industries and Social Change. Procedia - Social and Behavioral Sciences 173*, 18-23.

Chan, C. P. H. (2001). The perception (and production) of English word-initial consonants by native speakers of Cantonese. *Hong Kong Journal of Applied Linguistics, 6.1,* 26-44.

Chan, Y.Y. (2004). EyeSpeak V2.0 software specifications development. MEM thesis, University of Canterbury: NewZeleand.

Chistovich L, Fant G, de Serpao-Leitao A, Tjernlund P. (1966). Mimicking of synthetic vowels. Quart. Prog. Status Rep., *Speech Transmission Lab., Roy. Inst. Technol., Stockholm.* No. 2:1–18.

Coe, N. (2001). Speakers of Spanish and Catalan. In M. Swan & B. Smith (Eds.), Learner English: A teacher's guide to interference and other problems (pp. 90-112). Cambridge: Cambridge University Press.

Cox, F. & Palethorpe, S. (2001) The changing face of Australian vowels. In D. Blair & P. Collins (Eds.), *Language and Identity in Australia.* Amsterdam: John Benjamins.

Cruttenden, A. (2001). Gimson's Pronunciation of English, 6th ed. London: Arnold Publishers.

CSd'A (2005). *La situació de la llengua anglesa al batxillerat 2000-2004*. Barcelona, Departament d'Ensenyament, Consell Superior d'Avaluació del Sistwema Educatiu (Informes d'Avaluació, n. 7).

CSd'A (2006). *El coneixement de llengües a Catalunya*. Barcelona, Departament d'Ensenyament, Consell Superior d'Avaluació del Sistema Educatiu (Quaderns d'Avaluació, n. 6).

CSd'A (2008). *Resultats de la llengua anglesa de l'alumnat de 4t d'ESO de Catalunya*. Barcelona, Departament d'Ensenyament, Consell Superior d'Avaluació del Sistwema Educatiu (Informes d'Avaluació, n. 12).

Darcy, I., Ewert, D. & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation "toolbox." In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference,* (pp. 93-108). Ames, IA: Iowa State University

Darcy, I., Mora, J.C., Daidone, D.(2016). The role of inhibitory control in Second Language Phonological Procesing. Language Learning DOI; 10.1111/lang.12161.

De Bot, K. (1992). A Bilingual Production Model: Levelt's 'Speaking' Model Adapted. *Applied Linguistics, 13,* 1-24.

de Jong, K.J., Y.C. Hao, and H. Park (2009). Evidence for Featural Units in the Acquisition of Speech Production Skills: Linguistic Structure in Foreign Accent. *Journal of Phonetics,* 37: 357 - 373.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *SSLA, 22,* 499 – 533.

Dell G.S. (1986). A Spreading activation theory of retrieval and sentence production. *Psychological Review. 93:* 283-321.

Derwing, T. M. & Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-based Perspectives for L2 teaching and research.* Amsterdam: John Benjamins.

Dörnyei, Z., & Scott, M.L. (1997). Review article. Communication strategies in a second language: Definition and Taxonomies, *Language learning, 47(1),* 173-210.

Dörnyei, Z., Kormos, J. (1998). Problem-solving mechanisms in L2 communication. A Psycholinguistic Pespective. *SSLA, 20:* 349-385.

Eckman, Fred. 2004. From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition 26(4).* 513-49.

Edman, T. R. (1980).*Learning of intra-phonemic discrimination for several synthetic speech continua.* Doctoral dissertation (University of Minnesota, University Microfilms No. 81-09, 419

Eliades, S. J., and Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 453, 1102–1107.

Elman, Jeffrey L., Diehl, Randy L., & Buchwald, Susan E. (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America 62,* 971-974.

Elsendoor, B. A. G. (1984). Tolerance of duratioal properties in British English vowels. Dissertation RU Utrecht.

Escudero P. (2006). Second language phonology: the role of perception. In Pennington M., (Ed.). *Phonology in Context.* New York: Palgrave Macmillan; 109–134

Escudero P., Benders T., Lipski S. (2009a). Native, non-native and L2 perceptual cue weighting for Dutch vowels: the case of Dutch, German and Spanish listeners. *J. Phon.* 37 452–465. 10.1016/j.wocn.2009.07.006

Escudero, P. (2000). Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels in Spanish speakers. Unpublished M. Sc. thesis, University of Edinburgh.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26(04),* 551-585. doi:10.1017/s0272263104040021 275

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and Journal of Phonetics, 36 asymmetric lexical access. (2), 345-360. doi:10.1016/j.wocn.2007.11.002

Flege, J. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. Journal of Phonetics 15. 47-65 Flege, J. 1991. Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America 89*. 395-411.

Flege, J. (1989). Differences in inventory size affect the location but not the precision of tongue positioning in vowel production. *Language and Speech 32:* 123-147.

Flege, J. (1999). Age of learning and second-language speech. In: Birdsong, D. (ed.). 101-132. Flege, J. 2002. *Interactions between the native and second-language phonetic systems.* In: Burmeister, P. et al. (eds.). 217-244.

Flege, J. (2003). Assessing constraints on second-language segmental production and perception. In Meyer, N. Schiller (eds.). 319- 355. Flege, J. 2007. Language contact in bilingualism: phonetic system interactions. In: Cole, J., Hualde, J. (eds.). 353-382. Flege, J. 2009. *Give input a chance! In*: Piske, T., Young-Scholten,M. (eds.). 175-190.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America, 89(1),* 395-411. doi:10.1121/1.400473

Flege, J. E. (1995a). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics, 16,* 425-442. doi:10.1017/S0142716400066029

Flege, J. E. (1998). Second-language learning: The role of subject and phonetic variables. In *Proceedings of Speech Technology in Language Learning Conference* (pp.1-8). Marholmen, Sweden: ESCA.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. *Phonetics and Phonology in Language Comprehension and Production,6,* 319-355. doi:10.1515/9783110895094.319

Flege, J. E., & Davidian, R. D. (1984). Transfer and developmental processes in adult foreign Applied Psycholinguistics, 5 language speech production. (04), 323. doi:10.1017/s014271640000521x

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics, 15,* 67-83.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America, 76(3),* 708-721. doi:10.1121/1.391257

Flege, J. E., & Mackay, I. R. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26(01),* 1-34. doi:10.1017/s0272263104026117 277

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26,* 1-34.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*, 437-470.

Flege, J. E., Frieda, E. M., Walley, A. C., & Randazza, L. A. (1998). Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition, 20,* 155-187.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *The Journal of the Acoustical Society of America, 95(6),* 3623-3641. doi:10.1121/1.409931

Flege, J. E., Munro, M. J., & Mackay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97(5),* 3125-3134. doi:10.1121/1.413041

Flege, J. E., Schmidt, A. M, & Wharton, G. (1996). Age of learning affects rate-depending processing of stops in a second language. *Phonetica, 53,* 143-161.

Flege, J., Birdsong, D., Bialystok, E., Mack, M., Sung, H., Tsukada, K. (2005). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics 33*. 263-290.

Flege, J., Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23,* 527-552.

Flege, J., MacKay, I. (2004). Perceiving vowels in a second language. Studies in second language acquisition 26. 1-34. Flege, J., Munro, M., MacKay, I. 1995. Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America 97*. 3125-3134.

Flege, J., Yeni-Komshian, G., Liu, S. 1999. Age constraints on second language learning. *Journal of Memory and Language 41*. 78- 104.

Flege, J., Schirru, C. & MacKay, I. (2003) Interaction between the native and second language phonetic subsystems. *Speech Comm., 40,* 467-491.

Flege, J.E., & Schmidt, A. M. (1995). Native speakers of Spanish show rate-dependent processing of English stop consonants. *Phonetica, 52,* 90-111.

Flege, J.E., Yeni-Komshian, G.H., and Liu, S. (1999). 'Age Constraints on Second-Language Acquisition'. *Journal of Memory and Language 41,* pp. 78-104.

Fletcher S.G. & Hasegawa A. (1983). Speech modification by a deaf child through dynamic orometric modeling and feedback. *Journal of Speech and Hearing Disorders 48,* 178-185.

Fox, M. M., & Maeda, K. (1999). Categorization of American English vowels by Japanese speakers. *Proceedings of the International Conference of Phonetic Sciences* (pp. 1437-1440). San Francisco.

Fox, R. A., Flege, J. E., and Munro, M. J. (1995). 'The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *J. Acoust. Soc. Am. 97,* 2540–2551

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 349–366.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62,* 1668–1680.

Fridland, Valerie. (2008). Regional differences in perceiving vowel tokens on Southerness, education and pleasantness ratings. *Language Variation and Change* 20:67–83

Frieda E.M., Nozawa T. (2007). You are what you eat phonetically: The effect of linguistic experience on the perception of foreign vowels. In: Bohn O-S, Munro MJ, editors. *Language Experience in Second Language Learning.* Amsterdam: John Benjamins. 79–96.

Fullana, N. (2006). The development of English (FL) perception and production skills: Starting age and exposure effects. In C. Muñoz (Ed.), Age *and the rate of foreign language learning* (pp. 41–64). Clevedon, UK: Multilingual Matters.

Fullana, N., Mora, J. C. (2009) Production and perception of voicing contrasts in English word-final obstruents: Assessing the effects of experience and starting age. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.) *Recent Research in Second Language Phonetics/Phonology: Perception and Production.* (pp. 97-117). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Fullana, N., MacKay, I. R. (2003). Production of English sounds by EFL learners: The case of /i/ and /ɪ/. In *Proceedings of the 15th international congress of phonetic sciences, ICPhS* (pp. 1525-1528). Barcelona, Spain: ICPhS organizing comittee.

Fullana, N.. Muñoz, C. (1999). The Development of Auditory Discrimination Skills in EFL Learners of Different Ages. Presentació comunicació, XXIII AEDEAN (Asociación Española de Estudios Anglo-Norteamericanos)Conference.

Fullana, N.; Muñoz, C. (2000). A J.L. Chamosa, Proceedings of the XXIII AEDEAN International Conference. León: Servicio de Publicaciones de la Universidad de León. CD-ROM., León, ESPANYA

García Lecumberri, M. L., & Cenoz, J. (1998). Influencia de la duración en la adquisición de las vocales inglesas. In I. Vázquez, & I. Guillén (Eds.), *Perspectivas pragmáticas en lingüística aplicada* (pp. 201-207). Zaragoza: Textos de Filología 6, Asociación Española de Lingüística Aplicada (AESLA).

Garcia-Lecumberri, M., Gallardo del Puerto, F. (2003). English FL sounds in school learner of different ages. In M. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the Acquisition of English as a Foreign Language*. Clevedon: Multilingual Matters.

Goldenberb. (2008). Teaching English Language Learners What the Research Does—and Does Not—Say. *American Education*, (pp 8 - 44). AFT publications and reports.

Goldenberg, C. (2008). Teaching English language learners. What the research does −and does not− say. *American Educator*, summer, 8-23, 42-44.

Golfinopoulos, E., Tourville, J. A., and Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage* 52, 862–874.

Gordon E, Maclagan M, Hay J, Campbell L, Trudgill P. (2004). New Zealand English: Its Origins and Evolution. Cambridge University Press; Cambridge, UK.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "I" and "r."*Neuropsychologia*,9, 317–323.

Gracco, V. L., and Lofqvist, A. (1994). Speech motor coordination and control: evidence from lip, jaw and laryngeal movements. *J. Neurosci.* 14, 6585–6597.

Green, K.P., Kuhl, P.K., Meltzoff, A. N., Stevens, E.B. (1991). Integrating speech information across speakers, gender, and sensory modality: Female faces and male voices in the McGurk effect.*Perception & Psychophysics*,50, 524–536.

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.* 96, 280–301.

Guion, S.G. & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. Munro (Eds.) *Language Experience in Second Language Speech Learning*. Amsterdam: John Benjamins, 57-77.

Hardison, D. M. (1999). Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk effect. *Language Learning, 49,* 213-283. doi:10.1111/0023-8333.49.s1.7

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24(04),* 495-522. doi:10.1017/s0142716403000250

Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America, 125(1),* 469-479. doi:10.1121/1.3021295

Hawkins, S., & Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association 35, 183* – 199.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *Journal of Phonetics, 28,* 377-396. I

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., Chung, H., (2006). The use of visual cues in the perception of nonnative consonant contrasts. J. Acoust. Soc. Am. 119, 1740–1751.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication, 47(3),* 360-378. doi:10.1016/j.specom.2005.04.007

Heeren, W.F.L. & Schouten, M.E.H. (2008). Perceptual development of phoneme contrasts: how sensitivity changes along acoustic dimensions that contrast phoneme categories. *Journal of the Acoustical Society of America* 124: 2291–2302. doi: 10.1121/1.2967472.

Hietanen, J.K., Manninen, P., Sams, M., Surakka, V. (2001). Does audiovisual speech perception use information about facial configuration? *European Journal of Cognitive Psychology*, Vuosikerta 13, Nro 3, Sivut 395-407. DOI: 10.1080/09541440126006

Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (2000). Some effects of duration on vowel recognition. *J. Acoust. Soc. Am. 108,* 3013–3022.

Hirata, Y., Whitehurst, E., and Cullings, E. (2007). Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. J. *Acoust. Soc. Am. 121,* 3837–3845.

Højen, A., Flege, J. E. (2006). Second-language vowel perception. *J. Acoust. Soc. Am., Vol. 119,* No. 5.

Holt LL, Lotto AJ. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America. 119:*3059–3071.

Howard, I. S., & Messum, P. R. (2007). A computational model of infant speech development. Moscow: Moscow Linguistics University In Proceedings of SpeCom XII (pp. 756–765).

Indefrey P., Levelt W.J. (2004). The spatial and temporal signatures of word production components. *Cognition 92,* 101–144.

Ingram, J. C., & Park, S. G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of phonetics, 25(3),* 343–370

Ioup, G. (2005). Age in second language development. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 419-435). Mahwah, NJ: Lawrence Erlbaum.

Isaacs, T., & Trofimovich, P. (2011). Phonological memory, attention control, and musical ability: effects of individual differences on rater judgments of second language speech. *Applied Psycholinguistics, 32(1),* 113- 140. DOI: 10.1017/S0142716410000317

Iverson, P., & Evans, B. G. (2007). Auditory training of English vowels for first-language speakers of Spanish and German. *Proceedings of the 16th International Congress of Phonetic Sciences.* Saarbrücken, Germany. (pp. 1625-1628).

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America, 126(2),* 866-877. doi:10.1121/1.314819

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America, 118,* 3267-3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for Cognition, 87 non-native phonemes. (1), B47-B57. doi:10.1016/s0010-0277(02)00198-1

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English *Applied Psycholinguistics, 33 vowels. (01),* 145-160. doi:10.1017/s0142716411000300

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: English /ð/-/θ/ contrast by francophones. *Perception & Acquisition of the Psychophysics, 40(4),* 205-215. doi:10.3758/bf03211500

Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology and morphology. *Journal of Experimental Child Psychology, 77,* 30-60.

Jongman, A., & Wade, T. (2007). Acoustic variability and perceptual learning: The case of non-native accented speech. In O.-S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 135-150). Amsterdam: John Benjamins.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America, 138(2),* 817-832. doi:10.1121/1.4926561

Keating, P. (2004). Analysis of identification data. [Online resource]. Retrieved September 18, 2007, from http://www.linguistics.ucl.edu/faciliti/facilities/statistics.ident.htm]. *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech - 26*

Kewley-Port, D., and Atal, B. (1989), "Perceptual differences between vowels located in a limited phonetic space," *J. Acoust. Soc. Am. 85,* 1726–1740.

Kluge, D. C., Rauber, A. S., Reis, M. S., & Bion, R.A.H. (2007). The relationship between perception and production of English nasal codas by Brazilian learners of English. In *Proceedings of Interspeech 2007.* Antwerp, Belgium. pp. 2297-2300.

Koerich, R. D. (2006). Perception and Production of vowel paragoge by Brazilian EFL students. In: B. O. Baptista, M. A. Watkins (eds.), *English with a Latin Beat: Studies in Portuguese/Spanish – English Interphonology* (pp. 91-104). Amsterdam: John Benjamins.

Kormos, Judit . (1999). Monitoring and self-repair in L2. *Language Learning 49:* 303–42.

Koundarova. M. V., Francis, A. L.(2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Acoust. Soc. Am.* 1246.

Kuhl P.K., Conboy B.T., Coffey-Corina S., Padden D., Rivera-Gaxiola M., Nelson T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences. 363:*979–1000.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50,* 93-107.

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics, 21,* 125-139.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97,* 11850-11857.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). *Cross-language analysis of phonetic units in language addressed to infants. Science, 277,* 684-686.

Kuhl, P. K., Meltzoff, A. N. (1982). The Bimodal Perception of Speech in Infancy. *Science*, v218 n4577 p1138-40

Kuhl, P., & Iverson, P. (1995). Linguistic experience and the "Perceptual Magnet Effect". In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 121-154). Timonium, MD: York Press.

Kuhl, P.K., & Meltzoff, A.N. (1996). Infant vocalizations in response to 266 MELTZOFF speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America, 100,* 2425–2438.

Lacabex, E. G., & García-Lecumberri, M.L. (2010). Investigating training effects in the production of English weak forms by Spanish learners. In *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 137-43). Poznań, Poland: Adam Mickiewicz University press.

Lacabex, E. G., García-Lecumberri, M. L., & Cooke, M. (2008). Identification of the contrast full vowel-schwa: training effects and generalization to a new perceptual context. *Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies, (55).* 173-196. doi:10.5007/2175-8026.2008n55p173

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26(02),* 227-247. doi:10.1017/s0142716405050150

Lambacher, S. G., William L. Martens, Kazuaki Kakehi, Marasinghe, Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics* 26*(2):* 227-247.

Larson-Hall J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Lang. Res. 24*, 35–63 10.1177/0267658307082981

Leather, J. (1999). Second-language speech research: An introduction. *Language Learning, 49*(s1), 1-56. doi:10.1111/0023-8333.49.s1.1

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In Leather, Jonathan & Allan James (ed.). *New sounds 90,* 72-97. University of Amsterdam.

Lengeris, A. (2008). The effectiveness of auditory phonetic training on Greek native speakers‟s perception and production of southern British English vowels. In

Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics, ExLing 2008 (pp. 133-136). Athens: ISCA.

Lengeris, A. (2009). Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese. *Phonetica 66*, 169–187.

Lengeris, A., Hazan, V. (2010) The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek, *Journal of the Acoustical Society of America, 128(6),* 3757-3768.

Lenneberg, E. H. (1967). Biological foundations of language. New York: Wiley.

Lev-Ari, S., & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics,41(5),* 320-331. doi:10.1016/j.wocn.2013.06.002

Lev-Ari, S., & Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics, 47,* 36-46. doi:10.1016/j.wocn.2014.09.001

Levelt, W. J. M. (1989) *Speaking: from intention to articulation*. Cambridge, Massachusetts: The MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38.

Liberman, A. M., Harris, K. S., Hoffman, H.S., Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psychol., 54(5)*:358-68.

Lisker, L., Rossi, M. (1992). Auditory and visual cueing of the [rounding] feature of vowels. *Lang. Speech 35:* 391–417 (1992).

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94(3)*, 1242-1255. doi:10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America, 96(4),* 2076- 2087. doi:10.1121/1.410149

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In *Proceedings of the XIIIth International Congress of Phonetic Sciences , ICPhS* (Vol. 4, pp. 92-99). Stockholm, Sweden: Stockholm University.

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues (pp. 351-378).* Timonium, MD: York Press.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America, 89(2),* 874-886. doi:10.1121/1.1894649

Logan, J., & Pruitt, J. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 351-378). Timonium, MD:York Press.

Long, M. (1991). Focus on Form. Studies in Bilingualism Foreign Language Research in Cross-Cultural Perspective, 39-52. doi:10.1075/sibil.2.07lon

Lucea, J. M. (2004). The challenge of being an effective EFL teacher. *APAC Quarterly Magazine, 50,* 37- 43.

Ludlow, C. L., and Cikaoja, D. B. (1998). Is there a self-monitoring speech perception system? *J. Commun. Disord.* 31, 505–510.

Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Percept. Psychophys. 46,* 187–200.

Mackain B., Best., C., Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals.*Applied Psycholinguistics*,2, 369–390.

MacKay, I. R. A., Flege, J. E., Piske, T., & Schirru, C. (2001). Category restructuring during secondlanguage (L2) speech acquisition. *Journal of the Acoustical Society of America, 110,* 516-528.

MacKay, I.R.A., Meador, D., Flege, J.E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica 58:* 103–125.

Magnuson J.S., Yamada, R.A., Tohkura, Y., Pisoni, D.B., Lively, S.E., Bradlow, A.R. (1995). Proceedings of the 1995 Spring Meeting of the Acoustical Society of Japan. Tokyo: Acoustical Society of Japan; 1995. The role of talker variability in non-native phoneme training; pp. 393–394.

Markham, D. (1997). Phonetic imitation, accent and the learner. Lund University Press.

Martorell, J. (2006). English language teaching in our schools. A real challenge. *APAC Quarterly Magazine*, 58, pp. 38-45.

Massaro D. W., Cohen M. M. (1998). Visible speech and its potential value for speech training for hearing-impaired perceivers. in STiLL-Speech Technology in Language Learning (Marholmen), 171–174.

Matthews, J. (1997). The influence of pronunciation training on the perception of second language contrasts. In Leather, J. & James, A. (Eds.), *New Sounds 97: Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech* (pp. 223- 229). Klagenfurt: University of Klagenfurt

McAllister, R., Lubker, J. F. & Carlson, J. (1974). An EMG study of some characteristics of the Swedish rounded vowels*, Journal of Phonetics, 2,* 267–278.

McCandliss B.D., Fiez, J.A., Protopapas, A., Conway M., McClelland, J.L. (2002). Success and failure in teaching the [r]–[l] contrast to Japanese adults: Predictions of a Hebbian model of plasticity and stabilization in spoken language perception. Cognitive, Affective and Behavioral Neuroscience. 2:89–108.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2,* 89-108.

McDougall, K. and Nolan, F. (2007) Discrimination of speakers using the formant dynamics of /uː/ in British English. In J. Trouvain and W. J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences.* Saarbrücken: Universität des Saarlandes. 1825-1828.

McGurk, H., MacDonald, J. (1976). Hearing lips and seeing voices. *Nature 264,* 746–748 10.1038/264746a0

Messum, PR. (2007). The Role of Imitation in Learning to Pronounce, PhD Thesis. University of London.

Mora, J. C. (2007). Learning context effects on the acquisition of a second language phonology. In C. Pérez-Vidal, M. Juan-Garau, & A. Bel (Eds.), *A portrait of the young in the new multilingual Spain* (pp. 241-263). Clevedon: Multilingual Matters.

Mora, J. C. and Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism, 16(4)*, 484-500.

Mora, J. C., & Fullana, N. (2007). Production and perception of English /h9/-/H/ and /æ/-// in a formal setting: Investigating the effects of experience and starting age. *Proceedings of the 16th International Congress of Phonetic Sciences (*pp. 1613-1616). Saarbrücken, Germany.

Mora, J. C., Keidel, J. L. & Flege, J. E. (2015). Effects of Spanish use on the production of Catalan vowels by early Spanish-Catalan bilinguals. In Romero, J. & Riera, M. (Eds.) *The Phonetics-Phonology Interface: Representations and Methodologies*. Amsterdam: John Benjamins. 33-53.

Morrison, (2008). L1-Spanish speakers' acquisition of the /iː/-/ɪ/ contrast: Duration-based perception is not the English /i/–/ɪ/ initial developmental stage. Lang. Speech 54(4).

Morrison, G. S. (2002). Perception of English /i/ and /ɪ/ and Spanish Listeners: Longitudinal Results. *Proceedings of the North West Linguistics Conference 2002,* G. S. Morrison, and L. Zsoldos /eds.) Simon Fraser University Linguistics Graduate Student Association, Burnaby, BC, Canada), pp. 29–48.

Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition, 21,* 81–108.

Moyer, A. (2004) Age, Accent and Experience in Second Language Acquisition. An Integrated Approach to Critical Period Inquiry. Clevedon: Multilingual Matters

Muñoz C. (2006). The BAF Project: research on the effects of age on foreign language acquisition, in Age in L2 Acquisition and Teaching, eds Abello-Contesse C., Chacón-Beltrán R., López-Jiménez M. D., Torreblancaa-López M. M., editors. (Bern: Peter Lang;), 81–92

Munro, M. J. (1993). 'Production of English vowels by native speakers' of Arabic: Acoustic measurements and accentedness ratings. *Lang. Speech 36,* 39–66.

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

Munro, M. J., Flege, J. E., and MacKay, I. R. A. (1996). The effects of age of second-language learning on the production of English vowels. *Appl. Psycholing. 17,* 313–334.

Neufeld, G. G. (1988). Phonological asymmetry in second language learning and performance, *Language Learning, 38,* 531-559

Nieto-Castanon, A., Guenther, F. H., Perkell, J. S., and Curtin, H. D. (2005). A modeling investigation of articulatory variability and acoustic stability during American English/r/production. *J. Acoust. Soc. Am.* 117, 3196–3212.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech Language and Hearing Research, 50(6),* 1496-1509. doi:10.1044/1092-4388(2007/103)

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech Language and Hearing Research, 51(6),* 1480-1493. doi:10.1044/1092-4388 (2008/07-0109)

Nobre-Oliveira, D. (2007). Effects of perceptual training on the learning of English vowels in th non-native settings. In *Proceedings of the 5 International symposium on the acquisition of second language speech, New Sounds* (Vol. 5, pp. 382-389). Florianopolis: Federal University of Santa Catarina.

Öhrström, N., & Traunmüller, H. (2004). Audiovisual perception of Swedish vowels with and without conflicting cues. In *Proceedings, Fonetik 2004* (pp. 40-43). Stockholm: Stockholm University, Department of Linguistics.

Ortega-Llebaria, M., Faulkner, A., Hazan, Valerie (2001). Auditory-visual L2 speech perception: Effects of visual cues and acoustic-phonetic context for Spanish learners of English. In *AVSP-2001*, 149-154.

Osberger, M. & Kent, R. (1983). Effects of training on vowel production by deaf speakers. *Journal of the Acoustical Society of America, 73,* Supplement 1, S15(A).

Osberger, M. (1987). Training effects on vowel production by two profoundly hearing-impaired speakers. *Journal of Speech and Hearing Research, 30*, 241-251.

Pallier C., Bosch L., Sebastián-Gallés N. (1997). A limit on behavioural plasticity in speech perception. *Cognition 64,* B9–B17.

Pallier C., Colomé A., Sebastián-Gallés N. (2001). The influence of native-language phonology on lexical access: exemplar-based versus abstract lexical entries. *Psychol. Sci. 12,* 445–449.

Pallier C., Bosch L., Sebastián-Gallés N. (1997). A limit on behavioral plasticity in vowel acquisition. *Cognition, 64,* B9–B17.

Paradis, M. (2001) An Integrated Neurolinguistic Theory of Bilingualism (1976-2000). *LACUS Forum,* 27: 5-15.

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege, *Applied Linguistics, 11,* 73-89.

Pereira, Y. & Hazan, V. (2013). Impact of different training modes on the perception and production of English vowels by L2 learners. *Paper presented at New Sounds 2013,* Concordia University, Montreal, Canada

Pérez-Vidal, C., Juan-Garau, M. and Mora, J.C. (2011). The Effects of Formal Instruction and Study Abroad Contexts on Foreign Language Development: The SALA Project. In C. Sanz and R.P. Leow (Eds.) *Implicit and Explicit Language Learning. Conditions, Processes and Knowledge in SLA and Bilingualism.* Washington DC: Georgetown University Press, 115-128.

Piske, T. (2008). Phonetic awareness, phonetic sensitivity and the second language learner. In *Encyclopedia of language and education* (pp. 1912-1923). New York, US: Springer Publishing.

Piske, T., Flege, J., MacKay, I., Meador, D. 2002. The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica 59.* 49-71.

Piske, T., Mackay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics, 29(2),* 191-215. doi:10.1006/jpho.2001.0134

Piske, T., Young-Scholten, M. (eds.) 2009. Input matters in SLA. Bristol: Multilingual Matters.

Piske, Thorsten, Ian R. MacKay & James E. Flege (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics 29,* 191–215.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of Perception & Psyc hophysics, 13 consonants and vowels. (2), 253-260. doi:10.3758/bf03214136

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8(2),* 297- 314. doi:10.1037//0096-1523.8.2.297

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of non-native speech contrasts: Implications for theories of speech perception. In: Nusbaum H, Goodman J (Eds.), *Development of Speech Perception: The Transition from Recognizing Speech Sounds to Spoken Words* (pp. 121-166). Cambridge: MIT. 292

Pisoni, D., & Lively, S. (1995). Variability and invariance in speech perception. A new look at some old problems in perceptual learning. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 443-459). Timonium, MD: York Press.

Poulisse, N. (1997) Language Production in Bilinguals. In De Groot, A. M. B. and Kroll, J. F. (eds.) *Tutorials in Bilingualism : Psycholinguistic Perspectives* (pp. 201-224). New Jersey: Lawrence Erlbaum Associates.

Povel, J., Wansink , M. (1986). A computer-controlled vowel corrector for the hearing-impaired. *Journal of Speech and Hearing Research, Vol. 29*, pp. 99-105.

Praamstra, P., Hagoort, P., Maassen, B., & Crul, T. (1991). Word deafness and auditory cortical function: A case history and hypothesis. Brain, 114,1197-1225. doi:10.1093/brain/114.3.1197.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *Journal of the Acoustical Society of America, 119,* 1684-1696.

Raizada, R.D., Tsao, F.M., Liu, H.M., Kuhl, P.K. (2010) Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: prediction of individual differences. *Cereb Cortex 20:*1–12.

Rallo, L. 2005. Acquisition of a second language vowel system: The case of the Catalan learners of English. Unpublished PhD thesis. Universitat de Barcelona.

Rallo-Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of *Journal of Phonetics 40 (3), English vowels.,* 491-508. doi:10.1016/j.wocn.2012.01.001

Rauber, A. (2010). Acoustic characteristics of Brazilian English vowels: Perception and production results. Saarbrücken: Lambert Academic Publishing.

Rauber, A. S., Escudero, P., Bion, R. A. H., & Baptista, B. O. (2005). The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. *Paper presented at the Eurospeech-Interspeech 2005 9th European Conference on Speech Communication and Technology, Lisbon.*

Rauber, A., Rato, A., Kluge, D., Santos, G. (2011) TP-S (Version 1.0) [Application software]. Retrieved from http://www.worken.com.br/sistemas/tp-s/

Recasens, D., Farnetani, E., Fontdevilla, J., & Pallarès, M. D. (1993) An electropalatographic study of alveolar and palatal consonants in Catalan and Italian. *Language and Speech – Special Issue on Coarticulation, 36 (2/3),* 213- 234.

Repp B. H., Williams D. R. (1985). Categorical trends in vowel imitation: preliminary observations from a replication experiment. *Speech Commun. 4,* 105–120 10.1016/0167-6393(85)90039-1

Rochet, B. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 379-410). Timonium, MD: York Press.

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 22(2),* 318.

Rota G., Reiterer S. (2009). Cognitive aspects of pronunciation talent, in Dogil G., Reiterer S., (Eds.o). *Language Talent and Brain Activity,*. Berlin: Mouton de Gruyter. 67–96

Scovel, T. (1988). A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech. New York: Newbury House/Harper and Row

Scovel, T. (2000). A critical review of critical period research. *Annual Review of Applied Linguistics 20*. 213-223.

Scovel, T.(1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning, 19,* 245–253.

Sebastián-Gallés, N. , Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals.*Cognition*,72, 111–123.

Sebastian-Galles, N., Bosch, L. (2005). Phonology and bilingualism. In J.F. Kroll & A.M.B. de Groot (Eds.), *Handbook of bilingualism. Psycholinguistic approaches* (pp. 68–87) . Oxford: Oxford University Press.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52,* 240–255.

Segalowitz, N. (1997). Individual differences in second language acquisition. In A. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism* (pp. 85-112). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sheldon A. (1985). The relationship between production and perception of the /r/-/l/ contrast in Korean adults learning English: A reply to Borden, Gerber, and Milsark. *Lang Learn 35:* 107–113. doi:10.1111/j.1467-1770.1985.tb01018.x

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243-261.

Simmons A. J., Wise R. J. S., Leech R. (2011). Two tongues, one brain: imaging bilingual speech production. *Front. Psychol. 2:*166 10.3389/fpsyg.2011.00166

Sperbeck, M., Strange, W., Ito, K. (2005). Training Japanese L2 learners to perceive difficult American vowel contrasts. *Journal of the Acoustical Society of America, 117(4):* 2400.

Stadler, M., Richter, P. H., Pfaff, S., and Kruse, P. (1991). Attractors and perceptual field dynamics of homogenous stimulus areas. *Psychol. Res. 53,* 102–112.

Steinlen, A. K. (2005). The influence of consonantal context on native and non-native vowel production: A cross-language study. Tübingen: Gunter Narr.

Stockwell, R. P., Bowen, J. D. (1965). The sounds of English and Spanish. Chicago: University of Chicago Press.

Strange, W. (1995). Cross-language studies of speech perception: a historical review. In W. Strange (Ed.), *Speech perception and singuistic experience: Issues in cross language research* (pp. 3-45). Maryland: York Press.

Strange, W. (2007). Cross-language phonetic similarity of vowels: Theoretical and methodological issues". In O. Bohn & M. Munro (Eds.), Language experience in second language speech learning: In honor of James Emil Flege, (pp. 35-56). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r- l/ by Japanese adults learning English. *Perception & Psychophysics, 36(2),* 131-145. doi:10.3758/bf03202673

Tajima, K., Hiroaki, K., Rothwell, A., Akahane-Yamada, R., Munhall, K. G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *Journal of the Acoustical Society of America 123:* 397–413.

Tallal, P., Miller, S.L., Bedi, G., Wang, X., Nagarajan, S., Schreiner, C., Jenkins, W.M., Merzenich, M.M. (1998). Language comprehension in language-learning impaired children improved with acoustically modified speech, in M.E. Hertzig & E.A. Farber (Eds.) *Annual Progress in Child Psychiatry and Child Development 1997,* Brunner/Mazel, Inc., Bristol, PA.

Téllez, K. & Waxman H. (2006). A meta-synthesis of qualitative research on effective teaching practices for English language learners. In J. Norris & L. Ortega (eds.) *Synthesizing Research on Language Learning and Teaching.* Amsterdam: John Benjamins.

Torgersen E. (1997). Some phonological innovations in southeastern British English. University of Bergen: Unpublished MA Dissertation.

Tragant E. (2009). Trilingualism in Catalan schools and universities? Not yet. *APAC Quarterly Magazine*, 65, pp. 33-38.

Tragant E., Serrrano R., Miralpeix I., Navés T., Pahissa I., Serra N. & Gilabert R. (2010). *Estudi de sis instituts amb resultats destacables en la prova d'anglès de les PAU.* Universitat de Barcelona.

Tragant, E. (2009). Trilingualism in Catalan Schools and Universities? Not Yet. *APAC Quarterly Magazine, 65,* 33-38.

Tremblay, K., Kraus, N., Carrell, T.D., and McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *J. Acoust. Soc. Am.* 102:3762 -3773.

Trubetzkoy, N. (1939/1969). Principles of Phonology (C.A. Baltaxe, Trans.). Berkeley, CA: University of California Press.

Trubetzkoy, N.S. (1969). Principles of Phonology. (Translation of Grundzüge der Phonologie, 1939) Berkley. University of California Press.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2003). The perception and production of English /e/ and /æ/ by Korean children and adults living in North America. In *Proceedings of the fifteenth international congress of phonetic sciences* (pp. 1589–1592), Barcelona.

Vallabha G. K., Tuller B. (2004). Perceptuomotor bias in the imitation of steady-state vowels. *J. Acoust. Soc. Am. 116,* 1184–1197 10.1121/1.1764832

Vallabha, G. K., and Tuller, B. (2002), "Systematic errors in the formant analysis of steady-state vowels," *Speech Commun. 38,* 141–160.

van Hest, E. (1996). Self-repair in L1 ad L2 production. Tilburg, Netherlands: Tilburg University Press.

Walsh, T., Diller, K. (1978). Neurolinguitic Foundation to Methods of Teaching a Second Laguage. *International Review of Applied Linguistics 16:* 1-14.

Wang, A. J, Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America, 113(2):*1033–1043

Wang, X. (2002). Training Mandarin and Cantonese speakers to identify English vowel contrasts: Long-term retention and effects on production. Unpublished doctoral dissertation. Simon Fraser University.

Wang, X. (2008). Perceptual training for learning English vowels. Perception, production and long-term retention. Saarbrücken: VDM Verlag Dr. Müller.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System, 34*, 539-552

Watt, D., & Tillotson, J. (2001). *A spectrographic analysis of vowel fronting in Bradford English*. *English World-Wide*, *22*(2), 269-303.

Weiss, W. (1992). Perception and production in accent training. *Revue de Phonétique Appliquée, 102*, 68-92.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75,* 1866–187

Wildgen, W. (1990). Basic principles of self-organization in language. In H. Haken and M. Stadler (eds) (Springer-Verlag, Berlin), *Synergetics of Cognition*, pp. 429–452.

Wilson, S. M., Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage* 33, 316–325.

Wode, H. (1999). Incidental vocabulary acquisition in the foreign language classroom. *Studies in Second Language Acquisition, 21,* 243-258.

Wong, J. (2012). Training the Perception and Production of English /e/ and /æ/ of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training. *Proceedings of the 14th Australasian International Conference on Speech Science and Technology,* Sydney, Australia, 3-6.

Wong, J. (2013). Does proficiency matter? Effects of High Variability Phonetic Training on the Perception and Production of English Vowels by Cantonese ESL learners with high and low proficiency levels. *Paper presented at the 3rd International Conference on English Pronunciation: Issues & Practices (EPIP3),* University of Murcia, Spain.

Yamada, R. A. (1995). Age of acquisition of second language speech sounds: Perception of American English. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 305–320). Baltimore, MD: York.

Yamada, R., Tohkura, Y. I., Bradlow, A. R., & Pisoni, D. B. (1996). Does training in speech perception modify speech production? In *Proceedings of the fourth International Conference on spoken language processing, Vol. 2* (pp. 606-609). Philadelphia: IEEE publications.

Yamada, R.A., Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics*,52, 376–392.

Ylinen S., Uther M., Latvala A., Vepsäläinen S., Iverson P., Akahane-Yamada R., et al. (2009).Training the brain to weight speech cues differently: a study of Fnnish second-language users of english. *J. Cogn. Neurosci. 22,* 1319–1332 10.1162/jocn.2009.21272

Yoshikawa, Y., Koga, J., Asada, M., & Hosoda, K. (2003). A constructivist approach to infants' vowel acquisition through mother-infant interaction. *Connect. Sci. 2003; 15(4):* 245–258.

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to languagespecific auditory patterns. *NeuroImage, 26,* 703-720