

UNIVERSITY OF CALIFORNIA

Los Angeles

The role of modality and register in imitation by adults and children.

A dissertation submitted in partial satisfaction of the
requirements of the degree Doctor of Philosophy
in Linguistics

by

Nancy Ann Ward

2013

© Copyright by

Nancy Ann Ward

2013

ABSTRACT OF THE DISSERTATION

The role of modality and register in imitation by adults and children.

by

Nancy Ann Ward

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2013

Professor Megha Sundara, Chair

Research has shown that both adults and children will imitate acoustic properties of the speech around them. In fact, studies on adults have shown that this convergence occurs even when the subject simply sees, but does not hear, the interlocutor. Not only does visual speech elicit imitation on its own, but also imitation is greater for audiovisual speech than for auditory-only speech. However, these studies on audiovisual imitation have not looked at which properties of the speech are better imitated with the addition of the visual cues.

In this dissertation, I compare imitation in the auditory and audiovisual modalities to determine if audiovisual presentation (a) enhances the uptake of specific acoustic-phonetic cues, such as vowel formants, or (b) non-criterial information (f_0 and duration). For this I examine how closely children and adults imitate the productions of English-like and foreign vowels presented auditorily vs. audiovisually. Additionally, I attempt to determine how different speaking registers (adult-directed speech and child-directed speech) can aid in imitation.

Adult participants in this study showed greater imitation in the audio-visual modality, as has been shown previously. The increase in imitation was shown globally, across all different types of measures (acoustic-phonetic cues and non-criterial measures such as f0 and duration). The child-directed register as well facilitated adult imitation in a global manner. In contrast, child participants showed equivocal findings of increased imitation in the audiovisual modality and child-directed register on non-criterial measures such as f0 and duration. However, for the acoustic-phonetic cues, they showed more uniform increases of imitation in the audiovisual modality and child-directed register. The results of these experiments add to our understanding of how the visual modality is relevant in imitation, language development, and second-language learning.

The dissertation of Nancy Ann Ward is approved.

Sun Ah Jun

Patricia Keating

Lawrence Rosenblum

Kie Zuraw

Megha Sundara, Committee Chair

University of California, Los Angeles

2013

This dissertation is dedicated to my late father, John Ward, who will always be my inspiration
for every good thing I do in my life.

TABLE OF CONTENTS

Chapter 1: Introduction.....	1
1.0 Outline of the dissertation.....	2
1.1 Why study visual cues?.....	4
1.1.1 Visual cues in speech perception.....	4
1.1.2 McGurk Effect.....	7
1.1.3 Constraints on visual influences on speech perception.....	8
1.1.4 Autism and the use of the visual cues in speech.....	15
1.2 Feature-specific uses of the visual cues.....	16
 Chapter 2: Imitation.....	 20
2.1 Imitation in speech.....	23
2.2 Factors affecting imitation of familiar sounds in the auditory domain.....	25
2.3 How is imitation determined?.....	30
 Chapter 3: Adult-directed speech vs. child-directed speech.....	 33
3.1 What are the acoustic differences between adult- and child-directed speech?.....	34
3.2 What are the visual differences between adult- and child-directed speech?.....	35
3.3 What benefits does child-directed speech provide?.....	37
3.4 Child-directed speech vs. clear speech, or speech to foreigners.....	39
 Chapter 4: Relation of previous work to current study.....	 41

Chapter 5: Imitation experiments with adults	46
Chapter 6: Imitation experiments with children	175
Chapter 7: Discussion	281
7.1 Review of research questions	281
7.2 General summary of results	283
7.3 Implications for integration of visual cues in speech	286
7.4 Implications for research on language acquisition	287
7.5 Implications for research on imitation (not related to register or modality)	289
7.6 Future directions for the current research	290
Appendix	294
References	312

LIST OF FIGURES

Figure 5.1.	The speaker who produced the experimental stimuli.	51
Figure 5.2.	Vowel plots of F2 X F1 of adult- and child-directed stimuli.	57
Figure 5.3.	Vowel plots of F3 X F2 of adult- and child-directed stimuli.	57
Figure 5.4.	Schematic of the experimental design. In both the initial pre-exposure phase and post-exposure test phase, subjects saw a picture of clouds on the screen.	62
Figure 5.5.	Equations used in calculating convergence across each phonetic variable.	65
Figure 5.6.	Equations used in calculating the Euclidean distances for the pre-exposure and post-exposure productions, using all three formants, or just the first two. Also, the equation used for calculating the difference in distance.	66
Figure 5.7.	Comparison of pre-test and post-exposure convergence in each gender by modality for adult subjects in the duration measurement.	75
Figure 5.8.	Comparison of pre-test and post-exposure phonetic distance in each vowel by register for adult subjects in the duration measurement.	77
Figure 5.9.	Comparison of pre-test and post-exposure convergence in each register for adult subjects in the F0 measurement.	80
Figure 5.10.	Scatterplot of each subjects AQ and the difference in pre-test and post-exposure f0 measurement for speech register.	81

Figure 5.11.	Comparison of pre-test and post-exposure phonetic distance in each vowel by register for adult subjects in the Euclidean Distance (F1+F2+F3) measurement.	84
Figure 5.12.	Scatterplot of each subjects AQ and the difference in pre-test and post-exposure Euclidean Distance (F1 + F2 + F3) measurement for each register.	85
Figure 5.13.	Comparison of pre-test and post-exposure convergence in each modality for adult subjects in the F1 measurement.	90
Figure 5.14.	Scatterplot of each subjects AQ and the difference in pre-test and post-exposure F1 measurement for each modality.	91
Figure 5.15.	Comparison of pre-test and post-exposure convergence in each vowel by modality for adult subjects in the F2 measurement.	92
Figure 5.16.	Comparison of pre-test and post-exposure phonetic distance in each vowel by register for adult subjects in the F2 measurement.	94
Figure 5.17.	Scatterplot of each subjects AQ and the difference in pre-test and post-exposure F2 measurement for each register.	95
Figure 5.18.	Scatterplot of each subjects AQ and the difference in pre-test and post-exposure F2 measurement for each vowel.	95
Figure 5.19.	Mean duration of vowels produced by subjects in in the pretest and post-exposure phases, compared with the mean duration of the stimuli, separated by vowel, modality, and register.	98

Figure 5.20.	Mean f0 of vowels produced by male subjects in in the pretest and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel, modality, and register.	99
Figure 5.21.	Mean f0 of vowels produced by female subjects in in the pretest and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel, modality, and register.	99
Figure 5.22.	Formant plots of vowels produced by subjects in in the pretest and post-exposure phases, compared with the mean formant values of the stimuli, separated by vowel, modality, and register.	101-102
2Figure 5.23.	Results for convergence according to register for each measure in the pre-test/post-exposure comparison for adult participants.	104
Figure 5.24.	Results for convergence according to modality for each measure in the pre-test/post-exposure comparison for adult participants.	106
Figure 5.25.	Results for convergence in f0 by male participants by carrier type and modality of exposure in the pre-exposure/post-exposure comparison.	115
Figure 5.26.	Scatterplot of each subjects AQ and the difference in pre-test and post-exposure f0 measurement for each register.	116
Figure 5.27.	Results for overall convergence for male participants on the Euclidean Distance (F1+F2+F3) measure for English-like and foreign vowels by modality of exposure in the pre-exposure/post-exposure comparison.	118
Figure 5.28.	Results for overall convergence for male participants on the F1 measure by register for English-like and foreign vowels by modality of exposure in the pre-exposure/post-exposure comparison.	123

Figure 5.29.	Comparison of pre-exposure and post-exposure phonetic distance in each register by modality for female subjects in the duration measurement.	131
Figure 5.30.	Comparison of pre-exposure and post-exposure phonetic distance in each modality by context type for female subjects in the duration measurement.	132
Figure 5.31.	Comparison of pre-exposure and post-exposure phonetic distance in each register by modality for female subjects in the Euclidean distance convergence (F1/F2/F3) measurement.	138
Figure 5.32.	Comparison of pre-exposure and post-exposure phonetic distance in each modality by vowel type for female subjects in the Euclidean distance convergence (F1/F2) measurement.	139
Figure 5.33.	Comparison of pre-exposure and post-exposure phonetic distance for each type of vowel by register for female subjects in the F1 measurement.	143
Figure 5.34.	Comparison of pre-exposure and post-exposure phonetic distance for each type of vowel by modality for female subjects in the F1 measurement.	144
Figure 5.35.	Comparison of pre-exposure and post-exposure phonetic distance for each type of vowel by for female subjects in the F2 measurement.	147

Figure 5.36.	Mean duration of vowels produced by subjects in in the pre-exposure and post-exposure phases, compared with the mean duration of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.	152-153
Figure 5.37.	Mean f0 of vowels produced by male subjects in in the pre-exposure and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.	153-154
Figure 5.38.	Mean f0 of vowels produced by female subjects in in the pre-exposure and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.	155-156
Figure 5.39.	Formant plots of vowels produced by subjects in in the pre-exposure and post-exposure phases, compared with the mean formant values of the stimuli, separated by vowel, modality, and register.	158-159
Figure 5.40.	Results for convergence according to register for each measure in the pre-exposure/post-exposure comparison for adult participants, separated by gender.	161
Figure 5.41.	Results for convergence according to modality for each measure in the pre-exposure/post-exposure comparison for adult participants, separated by gender.	164

Figure 6.1.	Comparison of pre-test and post-exposure phonetic distance in each vowel by register for child subjects in the duration measurement.	187
Figure 6.2.	Comparison of pre-test and post-exposure phonetic distance in each modality for child subjects in the Euclidean Distance (F1+F2+F3) measurement.	191
Figure 6.3.	Comparison of pre-test and post-exposure phonetic distance in each vowel for child subjects in the F2 measurement.	196
Figure 6.4.	Comparison of pre-test and post-exposure phonetic distance in each vowel by register for child subjects in the F3 measurement.	197
Figure 6.5.	Comparison of pre-test and post-exposure phonetic distance in each vowel by gender for child subjects in the F3 measurement.	197
Figure 6.6.	Comparison of pre-test and post-exposure phonetic distance in each vowel by modality for child subjects in each of the individual phonetic measurements.	199
Figure 6.7.	Mean duration of vowels produced by subjects in in the pretest and post-exposure phases, compared with the mean duration of the stimuli, separated by vowel, modality, and register.	201
Figure 6.8.	Mean f0 of vowels produced by male child subjects in in the pretest and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel, modality, and register.	202
Figure 6.9.	Mean f0 of vowels produced by female child subjects in in the pretest and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel, modality, and register.	202

Figure 6.10.	Formant plots of vowels produced by subjects in in the pretest and post-exposure phases, compared with the mean formant values of the stimuli, separated by vowel, modality, and register.	204-205
Figure 6.11.	Results for convergence according to register for each measure in the pre-test/post-exposure comparison for child participants.	207
Figure 6.12.	Results for convergence according to modality for each measure in the pre-test/post-exposure comparison for child participants.	209
Figure 6.13.	Comparison of pre-exposure and post-exposure phonetic distance in each vowel type by register for male child subjects in the duration measurement.	217
Figure 6.14.	Results for convergence in f0 by male child participants by register of exposure in the pre-exposure/post-exposure comparison.	219
Figure 6.15.	Results for convergence in f0 by male child participants by vowel type and modality of exposure in the pre-exposure/post-exposure comparison.	220
Figure 6.16.	Results for convergence in Euclidean Distance (F1 + F2 +F3) by male child participants for each vowel type by register of exposure in the pre-exposure/post-exposure comparison.	223
Figure 6.17.	Results for convergence in F1 by male child participants to each vowel type by modality of exposure in the pre-exposure/post-exposure comparison.	228

Figure 6.18.	Results for convergence in F3 by male child participants to each vowel type by modality and register of exposure in the pre-exposure/post-exposure comparison.	230
Figure 6.19.	Results for convergence in duration by female child participants to each vowel type by modality of exposure in the pre-exposure/post-exposure comparison.	235
Figure 6.20.	Results for convergence in duration by female child participants to each carrier type by register of exposure in the pre-exposure/post-exposure comparison.	237
Figure 6.21.	Results for convergence in duration by female child participants to each vowel type by register of exposure in the pre-exposure/post-exposure comparison.	238
Figure 6.22.	Results for convergence in f0 by female child participants to each vowel type by modality of exposure in the pre-exposure/post-exposure comparison.	239
Figure 6.23.	Comparison of pre-exposure and post-exposure phonetic distance in each register for female child subjects in the Euclidean distance convergence (F1/F2/F3) measurement.	243
Figure 6.24.	Comparison of pre-exposure and post-exposure phonetic distance in each vowel type by register for female child subjects in the Euclidean distance convergence (F1/F2) measurement.	245
Figure 6.25.	Comparison of pre-exposure and post-exposure phonetic distance for each type of vowel for female child subjects in the F1 measurement.	248

Figure 6.26.	Comparison of pre-exposure and post-exposure phonetic distance for each type of vowel by register for female subjects in the F2 measurement.	251
Figure 6.27.	Results for convergence in F3 by female child participants to each carrier type by modality of exposure in the pre-exposure/post-exposure comparison.	253
Figure 6.28.	Results for convergence in F3 by female child participants to each vowel type by register of exposure in the pre-exposure/post-exposure comparison.	254
Figure 6.29.	Mean duration of vowels produced by child subjects in in the pre-exposure and post-exposure phases, compared with the mean duration of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.	260
Figure 6.30.	Mean f0 of vowels produced by male child subjects in in the pre-exposure and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.	262
Figure 6.31.	Mean f0 of vowels produced by female child subjects in in the pre-exposure and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.	264

Figure 6.32.	Formant plots of vowels produced by subjects in in the pre-exposure and post-exposure phases, compared with the mean formant values of the stimuli, separated by vowel, modality, and register.	266-267
Figure 6.33.	Results for convergence according to register for each measure in the pre-exposure/post-exposure comparison for child participants.	270
Figure 6.34.	Results for convergence according to modality for each measure in the pre-exposure/post-exposure comparison for child participants.	272

LIST OF TABLES

Table 5.1.	Wordlist of target stimuli for the experiment.	54
Table 5.2.	Filler words used as stimuli for the experiment.	55
Table 5.3.	Comparison of the present study's adult-directed formant values with past studies' formant values for the vowels used as stimuli in the present experiment.	56
Table 5.4.	Mean vowel measurements of all stimuli. Values are given in Hz.	58
Table 5.5.	Pretest wordlist items.	60
Table 5.6.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions.	74
Table 5.7.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for duration.	75
Table 5.8.	Summary of results from the mixed effects model for duration for adult participants in pretest tokens compared to post-exposure production tokens.	76
Table 5.9.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for f0.	78

Table 5.10.	Summary of results from the mixed effects model for f0 for adult participants in pretest tokens compared to post-exposure production tokens.	79
Table 5.11.	Summary of results from the mixed effects model for Euclidean Distance (F1 + F2 + F3) for adult participants in pretest tokens compared to post-exposure production tokens.	83
Table 5.12.	Summary of results from the mixed effects model for Euclidean Distance (F1 + F2) for adult participants in pretest tokens compared to post-exposure production tokens.	86
Table 5.13.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for F1.	88
Table 5.14.	Summary of results from the mixed effects model for F1 for adult participants in pretest tokens compared to post-exposure production tokens.	89
Table 5.15.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for F2.	92
Table 5.16.	Summary of results from the mixed effects model for F2 for adult participants in pretest tokens compared to post-exposure production tokens.	93

Table 5.17.	Results from overall convergence analyses, looking at whether there were significant findings of convergence across all measures in the comparison of pre-exposure and post-exposure productions.	111
Table 5.18.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for duration in male participants.	112
Table 5.19.	Summary of results from the mixed effects model for duration in pre-exposure compared to post-exposure productions for male participants.	113
Table 5.20.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for f_0 in male participants, and overall.	114
Table 5.21.	Summary of results from the mixed effects model for F_0 in pre-exposure compared to post-exposure productions for male participants.	114
Table 5.22.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance ($F_1 + F_2 + F_3$) in male participants, and overall.	118
Table 5.23.	Summary of results from the mixed effects model for Euclidean Distance ($F_1+F_2+F_3$) in pre-exposure compared to post-exposure productions for male participants.	119
Table 5.24.	Results from analyses looking at whether there was overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F_1+F_2) in male participants, and overall.	120

Table 5.25.	Summary of results from the mixed effects model for Euclidean Distance (F1+F2) in pre-exposure compared to post-exposure productions for male participants.	120
Table 5.26.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F1 in male participants, and overall.	121
Table 5.27.	Summary of results from the mixed effects model for F1 in pre-exposure compared to post-exposure productions for male participants.	122
Table 5.28.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F2 in male participants, and overall.	124
Table 5.29.	Summary of results from the mixed effects model for F2 in pre-exposure compared to post-exposure productions for male participants.	125
Table 5.30.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for duration in female participants, and overall.	129
Table 5.31.	Summary of results from the mixed effects model for duration in pre-exposure compared to post-exposure productions for female participants.	130
Table 5.32.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for f0 in female participants, and overall.	133

Table 5.33.	Summary of results from the mixed effects model for F0 in pre-exposure compared to post-exposure productions for female participants.	134
Table 5.34.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1 + F2 + F3) in female participants, and overall.	136
Table 5.35.	Summary of results from the mixed effects model for Euclidean Distance (F1+F2+F3) in pre-exposure compared to post-exposure productions for female participants.	137
Table 5.36.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1 + F2) in female participants, and overall.	139
Table 5.37.	Summary of results from the mixed effects model for Euclidean Distance (F1+F2) in pre-exposure compared to post-exposure productions for female participants.	140
Table 5.38.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F1 in female participants, and overall.	142
Table 5.39.	Summary of results from the mixed effects model for F1 in pre-exposure compared to post-exposure productions for female participants.	142

Table 5.40.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F2 in female participants, and overall.	145
Table 5.41.	Summary of results from the mixed effects model for F2 in pre-exposure compared to post-exposure productions for female participants.	146
Table 5.42.	Output for the factor of whether a word was generalized.	166
Table 6.1.	Results from subgroup analyses for child data, looking at whether there was any convergence in the comparison of pre-test and post-exposure productions.	184
Table 6.2.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for duration in child subjects.	185
Table 6.3.	Summary of results from the mixed effects model for duration for child participants in pretest tokens compared to post-exposure production tokens.	186
Table 6.4.	Summary of results from the mixed effects model for f0 for child participants in pretest tokens compared to post-exposure production tokens.	188
Table 6.5.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for Euclidean Distance (F1+F2+F3) in child subjects.	189

Table 6.6.	Summary of results from the mixed effects model for Euclidean Distance (F1+F2+F3) for child participants in pretest tokens compared to post-exposure production tokens.	190
Table 6.7.	Summary of results from the mixed effects model for Euclidean Distance (F1+F2) for child participants in pretest tokens compared to post-exposure production tokens.	192
Table 6.8.	Summary of results from the mixed effects model for F1 for child participants in pretest tokens compared to post-exposure production tokens.	194
Table 6.9.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for F2 in child subjects.	194
Table 6.10.	Summary of results from the mixed effects model for F2 for child participants in pretest tokens compared to post-exposure production tokens.	195
Table 6.11.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for F3 in child subjects.	196
Table 6.12.	Summary of results from the mixed effects model for F3 for child participants in pretest tokens compared to post-exposure production tokens.	198

Table 6.13.	Results from subgroup analyses for child subjects, looking at whether were significant findings of convergence across all measures in the comparison of pre-exposure and post-exposure productions.	213
Table 6.14.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for duration in male child subjects.	215
Table 6.15.	Summary of results from the mixed effects model for duration in pre- exposure compared to post-exposure productions for male participants.	216
Table 6.16.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for f0 in male child subjects.	217
Table 6.17.	Summary of results from the mixed effects model for F0 in pre- exposure compared to post-exposure productions for male participants.	218
Table 6.18.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1+F2+F3) in male child subjects.	222
Table 6.19.	Summary of results from the mixed effects model for Euclidean Distance (F1+F2+F3) in pre-exposure compared to post-exposure productions for male child participants.	222
Table 6.20.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1+F2) in male child subjects.	224

Table 6.21.	Summary of results from the mixed effects model for Euclidean Distance (F1+F2) in pre-exposure compared to post-exposure productions for male participants.	224
Table 6.22.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F1 in male child subjects.	226
Table 6.23.	Summary of results from the mixed effects model for F1 in pre-exposure compared to post-exposure productions for male participants.	227
Table 6.24.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F2 in male child subjects.	228
Table 6.25.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F3 in male child subjects.	229
Table 6.26.	Summary of results from the mixed effects model for F3 in pre-exposure compared to post-exposure productions for male participants.	231
Table 6.27.	Summary of register and modality findings by vowel type in each measure for male child participants in the pre-exposure/post-exposure comparison.	233
Table 6.28.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for duration in female child subjects.	235

Table 6.29.	Summary of results from the mixed effects model for duration in pre-exposure compared to post-exposure productions for female participants.	236
Table 6.30.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for f0 in female child subjects.	238
Table 6.31.	Summary of results from the mixed effects model for F0 in pre-exposure compared to post-exposure productions for female participants.	240
Table 6.32.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1 + F2 + F3) in female child subjects.	241
Table 6.33.	Summary of results from the mixed effects model for Euclidean Distance (F1+F2+F3) in pre-exposure compared to post-exposure productions for female participants.	242
Table 6.34.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1 + F2) in female child subjects.	244
Table 6.35.	Summary of results from the mixed effects model for Euclidean Distance (F1+F2) in pre-exposure compared to post-exposure productions for female participants.	244

Table 6.36.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F1 in female child subjects.	247
Table 6.37.	Summary of results from the mixed effects model for F1 in pre-exposure compared to post-exposure productions for female participants.	248
Table 6.38.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F2 in female child subjects.	249
Table 6.39.	Summary of results from the mixed effects model for F2 in pre-exposure compared to post-exposure productions for female participants.	250
Table 6.40.	Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F3 in female child subjects.	252
Table 6.41.	Summary of results from the mixed effects model for F3 in pre-exposure compared to post-exposure productions for female participants.	252
Table 6.42.	Summary of register and modality findings by vowel type in each measure for female children in the pre-/post-exposure comparison.	257
Table 6.43.	Summary of register and modality findings by vowel type in each measure for female children in the pre- /post-exposure comparison.	258

ACKNOWLEDGMENTS

First off, I would like to thank my advisor and co-author, Megha Sundara, for all the time and effort she put in to making me a better researcher and writer. I truly believe that I have made strides in my ability to succeed in this domain since coming to UCLA, and that this is directly due to Megha's influence. To put it succinctly, working with her has made me better.

I also owe gratitude to my other committee members (Patricia Keating, Sun Ah Jun, Kie Zuraw, and Lawrence Rosenblum) for their input on the development of this thesis. I appreciate the time and energy they expended in helping me make this thesis the best it could be.

I would also like to thank the UCLA Language Lab managers and research assistants for their help with this project (especially Chad Vicenik, Robyn Orfitelli, Shivani Bhakta, Hadley Vogt, Jessica Angulo, and Freshta Ayuby). I owe a huge debt of gratitude to all of the families who participated in this project. Also, special thanks to Marc Garellek and Adam Chong for helping me with VoiceSauce from afar.

I am grateful for the funding received through the NSF Dissertation Improvement Grant (Award Number 1226300), which helped allow this research to move forward.

Finally, I am indebted to my family and friends for all the support I have received in grad school. I would especially like to thank my husband, Matthew Finifter, not only for the large amount of hours he spent programming my experimental software used in this experiment, but for supporting me, loving me, being nice to me, and finally, for agreeing that I could get a puppy to "help me finish my dissertation".

VITA

Education:

2011 M.A. in Linguistics, University of California, Los Angeles.

Adviser: Megha Sundara; Committee: Patricia Keating, Nina Hyams

2008 B.A. with high distinction and honors in Linguistics, University of California, Berkeley.

Grants and awards:

2012 National Science Foundation Doctoral Dissertation Improvement Grant

2010 Acoustical Society of America – 2nd Pan-American/Iberian Meeting on Acoustics in
Cancun, MX

Best Student Paper Award: Speech Communication

“Development of native language preference in the visual modality.”

2009 Acoustical Society of America – 158th Meeting in San Antonio, TX

Best Student Paper Award: Speech Communication – Second Prize

“Consequences of short-term language exposure in infancy on babbling.”

2008 UC Berkeley Linguistics Departmental Citation

Publications in preparation:

Sundara, M., Ward, N., Conboy, B., & Kuhl, P. (in prep) Listening to Spanish for 5 hours alters
infant babbling in response to Spanish (but not English).

Ward, N. & Sundara, M. (in prep). The contribution of talking faces to the development of native
language preference.

Research talks:

- 2011 Is infants' native language preference in the visual modality guided by speech rhythm?
160th meeting of the Acoustical Society of America.
- 2011 Effects of linguistic environment on prosody of infant babbling.
Generative Approaches to Language Acquisition 2011.

Poster presentations:

- 2012 The role of visual cues in imitating an unfamiliar sound.
International Symposium on Imitation and Convergence in Speech 2012.
- 2012 The development and basis of visual language preference in infancy.
International Conference on Infant Studies 2012.
- 2011 Effects of bilingual exposure in infancy on babbling.
Summer Heritage Language Research Institute.
- 2010 Development of native language preference in the visual modality.
2nd Pan-American/Iberian Meeting on Acoustics.
- 2010 Short-term exposure to a second language produces language-specific effects in babbling. (accepted but not presented)
Boston University Conference in Language Development 35.
- 2009 Consequences of short-term language exposure in infancy on babbling.
158th meeting of the Acoustical Society of America.
- 2008 A study of vowel duration as a cue for underlying voicing of intervocalic alveolar flaps.
Summer Meeting of the Linguistics Society of America.

Chapter 1

Introduction

One of the issues central to second language learning is - how does a person learn to produce a sound in a second language? This dissertation research will investigate if speakers can learn to imitate a new sound, focusing on aspects of the input that are salient for the learner, and how different speaking styles may maximize a learner's uptake of the input.

The main focus of this research is to investigate how imitation is affected by the modality of the cues presented. Visual cues have emerged in the literature as extremely relevant for speech perception, but little research addresses how exposure to the visual modality might help with speech production. Most recently, the visual modality has been shown to facilitate the imitation of speech sounds (Diaz & Rosenblum, 2011). In this dissertation, we are trying to evaluate *how* visual cues facilitate imitation. In other words, what about the visual cues allows for better imitation? We begin with the premise that features of speech sounds may be more or less visually salient. For example, lip rounding is more easily observable than vowel front / backness or vowel height in the visual modality. If certain features are perceived to be more salient in a particular modality, does access to that modality allow imitation of that particular feature?

Experiments in this dissertation are designed to answer the following questions - Does the addition of visual cues to speech help only with imitation of visually salient aspects of speech sounds? Or is there simply an overall gain in the imitation of all aspects of speech sounds, perhaps by increasing overall attention when visual cues are presented? The main focus of this dissertation research is to evaluate these questions by looking at imitation of English-like and

foreign vowel sounds, to allow for a better understanding of what aspects of the input are being imitated. A secondary goal of this dissertation is to investigate how imitation may be affected by the age of the subject and the speaking register in which the speech (auditory or audiovisual) is presented.

1.0 Outline of the dissertation

This first chapter serves to explain why visual cues are important to consider in the context of perceiving speech. In it, we show that visual cues play an important role in speech perception: they improve the perception of a number of linguistic features - stress, tone, segmental identity - and they can fundamentally alter perception. In Chapter 2, we focus more specifically on imitation, in order to get the context for why visual cues could be important in imitative processes. Chapter 2 begins with a review of the literature on imitation, its role in learning, and its social nature. Chapter 2 continues with a summary of the literature regarding observations of speech imitation in adults and children, in both the auditory and visual domains, establishing that imitation is an innate behavior, shown at the earliest stages of development, and can be used as a valuable learning tool. Following that, Chapter 3 compares auditory and audiovisual differences in adult-directed and child-directed speech, in order to provide an understanding of the registers that will be used in the experiments in this dissertation and how they differ. Chapter 4 finishes the background review by relating the literature presented in Chapters 1-3 to the experiments conducted in this dissertation.

Chapter 5 will describe the first experiment conducted for this research. The aim of this experiment is to determine how adults use auditory and visual cues in imitation. In the first part

of this experiment, adult subjects were asked to shadow a model talker's production of a set of words. Next, the subjects received a period of either (a) auditory or (b) audiovisual exposure to the talker producing a subset of the words. Finally, subjects were asked to shadow the model talker's productions of the word list again.

Imitation was determined by examining how similar the subjects' productions are to the model talker's before and after the exposure period. There are two conditions to this experiment: one condition in which the subjects hear or hear and see adult-directed speech, and another in which subjects hear or hear and see child-directed speech. The purpose of including child-directed speech was two-fold: (1) to compare performance on a "learning" register compared to an adult register, and (2) to provide results that can be compared to child data collected in Chapter 6. We used a measure of Euclidean distance to determine the distance between the subjects' pre- and post-exposure productions and the productions of the model talker, following the methodology of Babel (2009). Chapter 5 will have a description of the data collection process for these experiments, as well as statistical analysis of the data, concluding with a discussion of the findings.

Chapter 6 will describe the second experiment conducted for this research. The aim of this experiment is to determine how children use auditory and visual cues in imitation. The methodology of this experiment is identical to the experiment in Chapter 5, substituting four-to six-year-old child subjects for the adult subjects. It is important to study the behavior of children on this same experiment because of reports of their decreased reliance on the visual cues to speech described at the end of section 1.1, and because it is suggested that children learn language in a different way than adults (Penfield & Roberts, 1959, Lenneberg, 1967) and seem to be better able to learn new languages before puberty (Flege, 1987).

In Chapter 7, I will conclude with a summary of the findings and results presented in Chapters 5 and 6 separately, and I will compare the results from adult subjects presented in Chapter 5 and the child subjects in Chapter 6 on a number of different phonetic measures. Chapter 7 will also include a summary of the remaining questions posed by this set of experiments, as well as a discussion of the theoretical implications of the findings. Finally, the implications of these results for theories of speech perception and imitation will be discussed.

1.1 Why study visual cues?

1.1.1 Visual cues in speech perception

Although most of the speech perception literature focuses on processing auditory speech cues, visual cues have been shown to play a significant role in speech perception. Visual cues do not just include the look of a speaker (which can cue gender, race, age, etc.) but also the facial movements correlated with speech sounds and rhythm. Visual speech cues are so informative that, even alone, they can cue language identification in both adults (Soto-Faraco, Navarra, Weikum, & Vouloumanos, 2007, Ronquest, Levi, & Pisoni, 2010) and infants (Weikum, Vouloumanos, Navarra, Soto-Faraco, Sebastian-Galles, & Werker, 2007, Ward & Sundara, submitted). In order to establish that these cues may be relevant in learning to produce a sound, as will be examined in this dissertation, it is important to consider that these cues are also very important for speech perception. This subsection will present the results of studies that show that visual cues can aid in speech perception tasks.

Sumby and Pollack (1954) first established that the visual modality improves perception when speech is presented in noise. In this study, participants were presented with bimodal –both auditory and visual - speech in noise and they were able to identify the words at a lower signal-to-noise ratio than when presented with just auditory cues. Thus, in difficult listening conditions, such as when presented in a noisy environment, visual cues can improve perception.

Visual cues can also help improve perception of a second or unfamiliar language. They have been shown to aid an adult when learning to perceive a second language, learning new and unfamiliar speech sounds. Learners have improved perception of a second-language if they are presented with visual as well as auditory cues (Reisberg, McLean, & Goldfield, 1987). The addition of visual cues also helps with acquiring phonemic distinctions perceptually; L2 Catalan speakers (native Spanish speakers) showed improved recognition of the phoneme distinction /e/ - /ɛ/ in Catalan when they were presented with visual cues in addition to the auditory cues (Navarra & Soto-Faraco, 2007). Japanese and Korean subjects trained to learn the /r/-/l/ distinction in American English showed improved perception and production of these sounds if they were trained using audiovisual stimuli rather than just audio stimuli (Hardison, 2003). Additional research has confirmed that audiovisual perception training has been shown to help with speech production of second-language sounds in cases in which “*visual cues to the phonemic contrast are sufficiently salient*” (Hazan, Sennema, Iba, & Faulkner, 2005). These studies on second-language acquisition clearly illustrate the facilitatory effect of audiovisual exposure on the perception and production of unfamiliar sounds. The majority of these studies focus on how audiovisual cues aid in perception, but there are a few also looking at production advantages.

Not only can the visual cues help with speech perception, but they can also help with other processing tasks. For example, children and adults showed better memory for sentences that were shown audio-visually rather than just auditorily (Thompson, Driscoll, & Markson, 1998), and adults show better perception of a speaker they have recently lipread from (Rosenblum, Miller, & Sanchez, 2008).

The ability to use visual cues for speech perception emerges very early in development. Infants are able to match the auditory stimulus they are hearing to one of two people they are seeing (Kuhl & Meltzoff, 1982). This ability can be shown in infants as young as 2-months (Patterson & Werker, 1999; Patterson & Werker, 2002). Infants at these young ages are able to extract more than just whether the timing of heard and seen speech is aligned; they are also able to match the gender of a heard and seen voice (Patterson & Werker, 2003). Thus, from a very young age, infants can use the visual cues in speech in guiding their attention towards the appropriate speaker.

Recently it has also been established that visual cues alone can give the perceiver a lot of information about speech. Both adults and infants have also been shown to have the capacity to discriminate languages based on visual cues alone (Soto-Faraco et al., 2007, Weikum et al., 2007). These abilities are shown in infants as young as 4-months (Weikum et al., 2007). Additionally, it has been shown that not only can infants discriminate their native language and an unfamiliar language visually, but they prefer to look at a face speaking their native language, proving that they can identify their native language from these visual cues (Ward & Sundara, 2010, 2012). Additionally of interest is the vast literature on lipreading or speechreading, which shows the vast amount of segmental and prosodic information available in visual-only stimuli (see Summerfield, 1992, for a review of much of the literature concerning the advantages in

perception available through speechreading). Thus visual cues give the listener access to a lot of information, information that is valuable on its own, or in conjunction with the auditory cues.

1.1.2 McGurk Effect

Many studies have evaluated the advantageous nature of the visual cues in speech perception. However, one of the most significant findings in the literature highlighting the role of visual cues showcases that not only do visual speech cues facilitate perception; they can also *alter* what a person perceives. This knowledge changes the visual cues from being something “helpful” to something fundamentally integrated within the speech perception system. In their pioneering study, McGurk and MacDonald (1976) played listeners an auditory instance of ‘ba’ synched to a visual presentation of ‘ga’. They found that listeners reported hearing ‘da’, a merged percept. Interestingly, in the reverse condition (auditory ‘ga’ and visual ‘ba’), listeners did not report fused responses and instead reported hearing ‘bagba’ or ‘gaba’, also a merged percept, but a combinatory response rather than fused as in the first condition. McGurk and MacDonald attributed this to the saliency of the visual cues for lip closure in ‘ba’. Since the lip cues in the visual articulation of ‘ba’ are very obvious to subjects, they were unable to ignore them entirely and rely only on what they were hearing. Note that listeners were able to correctly identify stimuli as ‘ba’ or ‘ga’ when just presented with just the auditory cues.

The fused McGurk effect response has been replicated successfully under a variety of conditions. In addition to using consonants for stimuli, integrated percepts like in the original McGurk study have also been observed when the test stimuli are vowel mismatches (Green & Gerdeman, 1995; Summerfield & McGrath, 1984; Traunmüller & Öhrström, 2007) or voicing

mismatches (Green & Miller, 1985). Thus, a mismatch produces an integrated percept consistently regardless of the specific sound type or feature analyzed.

This “McGurk effect” is so robust that it persists even with knowledge of the illusion or with incompatible (ex. male face and female voice), distorted (ex. reconfigured facial features), or degraded (ex. point-light displays) visual cues (Green, Kuhl, & Meltzoff, 1991, Hietanen, Manninen, Sams, & Surakka, 2001, Rosenblum & Saldana, 1996). These results demonstrate that the visual cues are integrated at a basic level of speech processing, and listeners still report an integrated percept even if there exist fundamental, and recognizable incompatibilities between the seen and the heard stimulus. There are even reports of “natural” McGurk effects when there are degraded audiovisual stimuli (presented in noise): /r/ and /w/ show confusion, with /r/ being perceived as /w/ in cases of lower signal to noise ratios, in cases in which the auditory component of /r/ is not very salient (Nielsen, 2004, Jiang, 2003). The McGurk effect demonstrates that visual cues don’t just aid in perception, but they can affect perception at a fundamental level.

1.1.3 Constraints on visual influences on speech perception

While audiovisual integration is widely accepted as a fundamental part of the speech perception process, researchers disagree on when the visual cues are incorporated into the speech percept. There are two main opinions in the literature about this topic (described and compared in detail in Rosenblum, 2008). The first is considered the amodal, or modality neutral account, and researchers who subscribe to this account believe that the two modalities are never separate

in processing, and that all processing is independent of modality (Rosenblum, 2005). Thus, integration takes place automatically and immediately.

An amodal integration account is supported by findings such as the McGurk effect, in which integration effects are seen even if the perceivers are asked to focus only on the auditory signal. Additional support for the immediate incorporation of the visual cues is from research using brain imaging techniques; for e.g, in a study looking at how neuro-typical adults process unimodal - auditory or visual - compared to bimodal speech presentation, presence of visual speech cues delayed speech processing (compared to auditory only processing), and this delay in activity occurred as early as 11 ms after stimulation (Musacchia, Sams, Nicol, & Kraus, 2006). This delay in processing occurred whether the visual cues were matching or mismatching (Musacchia et al. 2006). This integration is so early, that it is implausible to believe the modalities are processed separately.

The other account for bimodal speech integration, which we will call the late integration account, posits that the auditory and visual modalities are initially processed separately, and that integration comes at a later point. Thus, integration is neither automatic nor immediate (Bernstein, Auer, & Takayanagi, 2004, Massaro, 1998). While there are variations within this view as to when the auditory and visual cues are integrated, under this view integrated processing that appears to occur earlier is due to other factors, such as top-down information (see Bernstein et al. 2004 for a discussion of these other factors). Support for this account comes from research suggesting that lexical or semantic factors can influence audio-visual integration; for example, whether a McGurk stimulus is a word or not can affect the proportion of time that it is integrated (Brancazio, 2004). If lexical or semantic cues can affect integration, then processing could be argued to be later. Regardless of the view of multimodal integration one subscribes to, it

is agreed that multimodal integration, in one way or another, has the ability to affect perception at a basic and fundamental level.

Although speech perception is largely affected by the visual modality, there are limitations to the influence of visual cues on speech perception; visual cue integration can be affected by a person's history and experience with the use of the visual cues in speech perception. For example, the amount to which speakers may be affected by the McGurk effect seems to be dependent on language experience. Japanese-speaking adults have been shown to report less integrated percepts than English-speaking adults (Sekiyama & Tohkura, 1993). This has been attributed to cultural differences in the politeness of looking at faces while an interlocutor is speaking.

At first, these results seem to be incompatible with the idea that visual cues are closely integrated with auditory cues in speech processing. However, the situation is not that different from that of a visually impaired person learning to rely more on the auditory cues, or a hearing impaired speaker learning to rely more on the visual cues (see Woodhouse, Hickson, & Dodd, 2009, for a review of literature on visual cues in hearing and hearing-impaired individuals). Perceivers often weight cues differently based on their experience (Francis, Kaganovich, & Driscoll-Huber, 2008, Gandour, 1983, Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert, 2003). In fact, Japanese speakers behave comparably to non-Japanese listeners in a non-McGurk visual speech task, namely a speech-recognition-in-noise task (Sekiyama & Tohkura, 1991). Massaro, Tsuzaki, Cohen, Gesi, & Heridia (2003) also find no differences in audiovisual processing in Japanese, English, and Spanish. Neilson (2004), in contrast, finds a lack of McGurk effect in Japanese listeners but attributes it to the segmental inventory of Japanese, and not to the cultural differences in visual cue use.. She suggests that

since Japanese has segments that vary less in their visual distinctiveness, and thus, Japanese listeners learn to attribute less weight to the visual cues in processing (Nielsen, 2004). This is supported by research looking at AV influences on the perception of English consonants. Although the perception of consonants with less salient acoustic cues improves with the addition of visual cues, consonants without salient visual cues showed only a minor or negative increase in intelligibility when the visual cues were added (Nielsen, 2004). The influence of specific visemes over others in audiovisual perception is supported by the literature on how *specific* visible segments can enhance the comprehension of auditory speech (Rosen & Corcoran, 1982, among others) and on perceptual confusability between visemes (Binnie, Montgomery, & Jackson 1974, Walden, Prosek, Montgomery, Scherr, & Jones, 1977, Jiang, 2003).

Language experience also influences the product of audiovisual integration; French listeners report a different merged percept than English listeners due to differences in the phoneme inventories of the two languages (Werker, McGurk, & Frost, 1992). Even within a single language, there can be individual variation in the degree of bimodal integration. One study on the McGurk effect found a clear divide in the use of visual cues by their subjects; some relied more heavily on the visual cues and were more biased towards responding with the *visual* stimulus in cases of a mismatch, the rest relied more on the auditory cues and were more biased towards responding with the *auditory* stimulus in cases of a mismatch (Traunmüller & Öhrström 2007). Language related impairments, such as autism, can also affect the reliance on the visual cues to speech (to be discussed in section 1.1.4 below). In sum, language experience as well as individual differences can alter the extent to which people rely on auditory and visual cues.

Reliance on the visual cues to speech can be permanently altered in the absence of exposure during development. For example, the integration of visual cues in the speech

perception system is largely affected by any deficits during development, such as a hearing impairment. Schorr et al., (2005) studied the McGurk effect in children who received cochlear implants (all deaf from birth). They found that 92% of children with cochlear implants, when presented with mismatched stimuli, reported “hearing” what was visually presented rather than a fused stimulus (Schorr, Fox, van Wassenhove, & Knudsen, 2005). However, there were some subjects who did fuse the stimuli consistently, similar to typically developing children. Looking closer into these findings, the researchers found that children who received the implant after 2.5 years of age, rarely showed the McGurk effect. Schorr et al., (2005) suggest that this implies a “critical age” for being able to use the visual cues in speech processing.

Considering that visual cues are utilized in infancy for a number of different speech related tasks, one of the most surprising findings related to the limitation in the influence of the visual cues is the finding that pre-adolescent children are less influenced by the visual cues to speech. In the original McGurk & MacDonald (1976) study, the authors also studied audiovisual integration in children using the same experimental design. They found that children were less likely to form a merged percept. The children’s performance in the McGurk task depended on age: 81% of 3-4 year old children reported a fused percept with auditory ‘ba’ and visual ‘ga’, whereas only 64% of 7-8 year olds did the same. Compare this to the 98% of adults who showed the effect. Children’s limited ability to demonstrate a McGurk effect has been replicated multiple times and with variations in the methodology (Massaro, 1984; Massaro, Thompson, & Laron, 1986; Desjardins, Rogers, & Werker, 1997). Research also shows that children are poor lipreaders (Massaro, 1987), further indicating that children rely less on visual cues in speech perception tasks.

Surprisingly, infants demonstrate audiovisual integration, even in tasks where the older children fail (Burnham & Dodd, 2004, Rosenblum, Schmuckler, & Johnson, 1997). Burnham & Dodd (2004) familiarized 4.5-month-old infants to a McGurk stimulus – auditory ‘ba’ and visual ‘ga’ and then tested their preference to look towards natural speech stimuli, ‘da’, ‘ba’, and ‘ga’. They found that the 4-month-olds looked towards the ‘da’ stimuli in the test phase, showing that, just like adult listeners, infants merge audio and visual information when presented with the McGurk stimulus (see also Kushnerenko, Teinonen, Volein, & Csibra, 2008 for ERP evidence for the McGurk effect in infants) . Studies showing influences of the visual cues on speech perception in infancy, suggest that the lack of visual cue integration in childhood is a result of a decline in sensitivity to these cues, not that children have yet to incorporate them into their perceptual systems.

In fact, researchers have argued that cue redundancy in multimodal stimulation provides many perceptual advantages in development (see Bahrick & Lickliter, 2000, for a review of these studies); bimodal cues aid infants in language-related tasks, garner infants’ attention, and are remembered better than cues that are either only salient, or only presented, in one modality. Also, once a specific property has been perceived “amodally”, then it is available in perception unimodally (Bahrick & Lickliter, 2000). Bahrick & Lickliter refer to this advantage of “redundant” multimodal cues as the intersensory redundancy hypothesis. While these studies make no claims about whether amodal perception is innate or learned, they emphasize the importance of early modality integration in development, and claim that this integration is well-established in infancy. As children get older, their decline in the use of the visual cues appears to be due to the re-weighting of their attention.

In fact, as children approach 8-9 years of age, they again begin to merge auditory and visual information more readily, with adult-like integration emerging when they reach 10-12 years (Hockley & Polka, 1994, Linden & Vroomen, 2008). For this reason, it has been suggested that multimodal integration might be a U-shaped developmental function, with a greater influence of visual cues before 4 years of age, and again, after 9 years (Jerger, Damian, Spence, Tye-Murray, & Abdi, 2009). In sum, the decline in the use of visual cues for young children has been well documented, and has implications for how children these ages might behave in other visual speech processing tasks, such as imitation.

One final factor that affects visual language influence is gender. While females and males show the same responses on traditional McGurk tasks, some McGurk task variations can reveal gender differences with regards to processing of visual cues. Irwin and colleagues designed a task in which they played brief duration (cut off at 100ms) or full duration McGurk-type monosyllabic stimuli (audiovisual matched 'ba', 'va', 'da', and 'ða', and mismatched visual 'va', 'da', and 'ða' with auditory 'ba'). They found that while males and females were comparable in responses to the mismatches of full duration stimuli, females were more biased to respond with the visual response when hearing the brief stimuli (Irwin, Whalen, & Fowler, 2006). There are also findings that suggest a gender difference in lip reading of sentences (Johnson, Hicks, Goldberg, & Myslobodsky, 1988, Watson, Qiu, Chamberlain, & Li, 1996). Infants also show gender differences in visual speech integration, but these differences do not pattern in a particular way to suggest that either gender has superior abilities, simply that they process tasks differently (Desjardins & Werker, 2004). The results on the influence of gender in the integration of the visual modality in speech perception suggest that there do exist differences, but the nature and influence of these differences remains unclear.

1.1.4 Autism and the use of the visual cues in speech

One of the main findings about the lack of the influence of the visual cues in speech perception is in people with autism. There are a number of findings that show that this disorder can affect different aspects of communication (Tager-Flusber, Paul, & Lord, 2005), especially as related to visual speech processing. Autistic individuals have trouble with cross-modal integration (Iarocci & McDonald, 2006) and integration in a McGurk task, reporting fewer fused responses than their non-autistic peers (de Gelder, Vroomen, & van der Heide, 1991, although also see Keane, Rosenthal, Chun, & Shams, 2010). The visual signal also contributes significantly less to an autistic individual's ability to recognize speech in noise (who are otherwise on par with normally-developed peers in an auditory-only condition), and they are known to be poor lip readers (Smith & Bennetto, 2007).

In general, those who are autistic not only have problems with visual speech tasks, but they also simply process faces differently. Autistic children have been shown to attend to different facial features than normally developing children (Langdell, 1978). They also have more trouble matching faces and voices (Boucher, Lewis, & Collis, 1998) and detecting temporal synchrony between a seen and heard stimulus (Bebko, Weiss, Demark, & Gomez, 2006). In sum, given a face and voice, the autistic individual will have a reduced ability to integrate auditory and visual information and rely more on what they are hearing rather than what they are seeing. This is in contrast to normal adults and children, who rely on both modalities.

If adults with autistic traits do have more trouble incorporating visual cues into perception, then it is possible that autistic traits in neurotypical adults could affect performance

on a task that uses visual language skills, as in the research study conducted in this dissertation. Recently, researchers have developed the Autism Quotient questionnaire (AQ), a gradient, non-clinical test that can identify the degree to which a person (autistic or non-autistic) exhibits autistic-like personality traits (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). The non-diagnostic questionnaire outputs scores on a scale from 1-50, with higher scores indicating that a person may have more autistic-like traits. Scores of 32 or higher are a good indication of autism (Baron-Cohen et al. 2001), and scores under 26 indicate little to no chance of the individual having autism. The AQ questionnaire has recently been used in a number of linguistic studies to show how scores affect linguistic performance in imitation tasks (Mielke, Nielsen, & Magloughlin, 2013), speech processing (Ota & Stewart, 2007, Lindell, Notice, & Withers, 2009, Stewart & Ota, 2008, Yu, Grove, Martinovic, & Sonderegger, 2011), and phonetic perception (Yu, 2010). Individual variability is shown in many audiovisual language tasks (for one example that discusses the variability in subjects, see Traunmüller & Öhrström, 2007), and, if autism and visual language skills are correlated as some literature suggests, it is possible that the relative AQ score might be one factor which could help explain this variability.

1.2 Feature-specific uses of the visual cues

The literature described so far in this chapter demonstrates an overall advantage for speech perception when visual cues are accessible, and how deficits in visual cue integration usually stem from lack of access to these cues. A key outstanding question concerns the nature of the specific advantages that visual cues might confer to the perceiver: are there particular cues in

the visual signal that provide these specific advantages and contribute more to the improvement in speech perception than others?

A few studies have looked into how specific visual cues signal linguistic information in speech perception, largely looking at correlations between acoustic cues and facial movements. In some cases, evidence has pointed at precise identifiable movements – specific visual cues – that can help a listener in speech perception tasks. For example, visual cues have been correlated with supra-segmental acoustic cues. Head and face movement has been shown to help Japanese subjects better identify syllables differing in pitch accents in an audiovisual speech in noise task (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004). Similarly, visual cues can also be used as a cue for the perception of lexical and phrasal stress in American English. Like in the study on Japanese perception, there are specific visual cues that appear broadly correlated with the perception of lexical and phrasal stress for American English, as well as certain specific movements, such as chin displacement, lip distance, and lip displacement that seem to help subjects perceive lexical stress visually (Scarborough, Keating, Mattys, Cho, & Alwan, 2009). Acoustic-visual cue correlations have also been shown with tone: subjects from tonal and non-tonal language backgrounds have been shown to be able to identify tone based on silent videos (Burnham, Lau, Tam, & Schoknecht, 2001), additionally providing evidence for the idea that supra-segmental characteristics can be associated with facial movements.

Like visual cues to supra-segmental features, there are also visual cues that correspond to segmental identity, referred to as *visemes* (Fisher, 1968). A viseme is the basic unique classification of the visual correspondents of sounds. Visemes can distinguish sounds that are similar acoustically, such as /f/ and /θ/, however, sounds that are acoustically distinct can have the same viseme classification, such as /b/ and /m/. Visemes and phonemes do not have a one-to-

one correspondence, and there are a number of sounds that are acoustically different, but visually similar, as well as ones that are visually similar but acoustically distinct (Owens & Blazek, 1985). The notion of a viseme is used extensively in studies on lipreading (Massaro & Cohen, 1990).

Incorporating this type of viseme information with the methodology of the McGurk effect, one study has looked at how mismatching auditory and visual cues are perceived. Traunmüller & Öhrström (2007) tested speech perception of rounded and unrounded vowels by Swedish listeners. They tested matching and mismatching audiovisual stimuli, with vowels varying in two dimensions: openness and rounding. They asked participants to identify which vowel they heard. They show that in perception of mismatches between auditory and visual vowel cues, participants used the cues that are the most salient characteristic of the sound, and their degree of visual integration depended on how visually salient the particular sound is, i.e. visual cues were more likely to bias a response for visually salient sounds than auditorily salient sounds. Specifically, for variations within the vowel quality in the open/close dimension, which is more salient auditorily, subjects relied more on the auditory cues, but for variations in rounding, a very visually salient feature, they used the visual cues more. For example, for the stimulus audio ‘geg’ (with a **mid**, front **unrounded** vowel) and visual ‘gyg’ (with a **high**, front **rounded** vowel), subjects identified the stimulus as ‘gøg’ (a **mid**, front **rounded** vowel), combining the feature of openness from the auditory dimension and rounding from the visual dimension. The contribution of auditory and visual cues to perceiving a sound was not equal and uniform across all sounds, but rather depended on what the properties were of the sound that was being perceived. Thus, bimodal speech integration depends on the features of particular sounds. As Traunmüller & Öhrström (2007) make clear, visual cues will dominate perception for visually

salient features. The name of this theory is the *Information Reliability Hypothesis* – “The perception of a feature is dominated by the modality that provides the more reliable information”. The Traunmüller & Öhrström (2007) study identifies the visual modality as more reliable for perception of the rounding feature as the acoustic differences were more minor (hence it dominates perception for contrasting rounding cues), and the auditory modality as more reliable for perceiving the openness feature, as the visual cues are less clear for this feature. This is similar to a previously suggested hypothesis by Welch & Warren (1980), which states that in general, in audiovisual perception, perception will be controlled by the modality more attuned to the stimulus being presented.

The results of the study by Traunmüller and Öhrström (2007) allow us to make predictions in the current study for when subjects may use the visual cues in an imitation task, which combines both perception and production. Based on the information reliability hypothesis, subjects are likely to use visual cues when they are the salient feature (contain the more reliable information) of a sound, and they will not use these cues when the sound is more salient in the auditory dimension. To show a specific correlation between presentation of the visual cues and imitating particular sound features would be a novel finding, since the prior research on the contribution of the visual cues has largely focused simply on this overall advantage and has not looked at *how* these cues contribute specifically to perception and production processes. If we can show that visual cues are helpful in imitating a *new* sound, that would have strong implications for how to learn a second language, namely that visual cues could facilitate learning.

Chapter 2

Imitation

Imitation is a powerful implicit learning mechanism available in early infancy. Humans are social creatures that learn from observing others. From the earliest stages of life, imitation is an automatic behavior that is demonstrated in human neonates. Forty-two-minute-old newborn infants display imitative facial configurations (Meltzoff & Moore, 1983, 1989). Once infants are just a few hours old, they imitate tongue and lip gestures made by an adult (Meltzoff & Moore, 1997, 1983, Kugiumutzakis, 1999). The imitated gestures shown by infants include a variety of different facial movements: tongue and lip protrusions, side-to-side lip movements, straight protrusions, mouth openings (Meltzoff & Moore, 1977, 1994, 1997). While gestural imitation is seen in other species, certain imitative behaviors shown in human infants, such as imitations of certain lip gestures, are exclusive to humans and have been attributed to advanced learning strategies in humans (Meltzoff & Williams, 2010).

Imitation is a naturally social behavior. The nature of imitation is that there must be some connection not just between what is seen and what is done, but also between a doer and a seer. This social nature of imitation is often demonstrated in infancy. For example, infants show closer degrees (Nielsen, 2006, Brugger, Larivier, Mumme, & Bushnell, 2007, Gergerly, Bekkering, & Kiraly, 2002) and larger amounts of imitative behavior towards an adult who is attempting to socially engage them, rather than an adult who is merely completing the action in the presence of the infant (Brugger et al. 2007). Infants also demonstrate that they are more socially engaged if a person interacting with them is also imitating them; they demonstrate longer looking times

towards the adult and smile more frequently (Meltzoff, 2007). Infants even demonstrate physiological reactions to imitation: infants show different heart rate patterns for imitative acts compared to self-initialized acts (Nagy & Molnar, 1994, 2004).

Imitative behavior is not simply to foster relationships and social identity; while there are many other learning strategies that could be employed by the developing infant, such as trial and error (an unsystematic problem-solving approach in which the learner would try repeated different attempts to solve the problem) and independent invention (an approach of simply attempting different strategies to solve a problem with no knowledge of how to actually solve the problem) (Meltzoff & Williamson, 2010), imitation has been shown to be a viable learning mechanism. Infants demonstrate that imitation is not just repetition, but that they understand the motivation and goals behind an incomplete observed action: infants watching an adult attempt to pull apart a toy unsuccessfully will perform that action successfully (Meltzoff, 1995, 2007). At 18-months, children are able to differentiate between accidental and purposeful acts (if an adult marks completion of a task with an accidental exclamation, such as “whoops”, the child is less likely to imitate the adult than if they mark completion of the task), demonstrating their understanding of the goal of the person they are interacting with. Additionally, infants imitate from memory, proving that imitation is not simply direct replication; the acts are learned and committed to memory and can be reenacted in a separate situation (see Meltzoff & Williamson, 2010, for an extended review of the literature on infant imitation and memory). Infants do not limit their imitation to behaviors produced by adults, but they can also learn from their peers. Infants have been shown to imitate “knowledgeable peers” who have proven able to produce a reaction from a toy novel that was presented to the infants (Hanna & Meltzoff, 1993). Thus,

imitation can be used for learning and understanding behaviors in infancy, and is much more complex than mere repetition.

Imitation as a strategy for learning and social identification is not limited to infancy. Adults often demonstrate implicit imitative behaviors for the same purposes. Like infants, it is logical that imitative behaviors could also be used by adults for the purpose of learning. Think of a language-learning situation, and how verbal imitation of a native speaker could facilitate learning the new sounds or the new language, as it is a well-proven learning strategy (Meltzoff & Williamson, 2013, Meltzoff, Kuhl, Movellan, & Sejnowski, 2009, Repacholi, Meltzoff, & Olsen, 2008, Williamson, Meltzoff, & Markman, 2008, Meltzoff, 2007, Schaal, 1999). Imitation in adults, also like in infants, seems to be very socially motivated. For example, subtle gestural imitation can facilitate bonding and relationship forming through social identification (Chartrand & Bargh, 1996). There also seems to be something special about the face, evident in the vast amount of findings related to automatic imitation of facial expressions (see Meltzoff & Moore, 1997, for a discussion of this literature). In a study on gestural imitation in conversing adults, facial expressions were the most easily recognized forms of imitation, compared to foot movements or face rubbing (Chartrand & Bargh, 1996). This study also evaluated the social relationships developed over the course of the conversation, and found that greater degrees of facial imitation resulted in higher ratings for the likeability of the interlocutor and the amount of imitation shown by a subject was related to empathetic characteristics. Additionally in this study, subjects not only felt more positive feelings towards their interlocutor, but they also felt the interaction was smoother if there was a higher degree of imitation. Thus, imitation serves this social communicative function that works two-fold; it helps a person learn a particular action, and also helps them maintain social relationships.

Perhaps due to this social nature of imitation, autistic individuals have been demonstrated to show a lesser propensity for imitation. Characteristically, individuals with autism show less inclination towards social behaviors, and they show deficits in communicative abilities (American Psychiatric Association, 1994). Thus, perhaps unsurprisingly, autistic people show reduced inclination towards imitative behaviors compared to their non-autistic peers (see Williams, Whiten, & Singh, 2004, for a review).

In sum, imitative behavior seems to play a crucial role in both infant and adult interaction. Due to its role as a learning mechanism, and considering how it fosters social relationships, it is an important behavior to understand. In the following section, I will review one specific type of imitation: the implicit imitation of speech sounds.

2.1 Imitation in speech

Imitation of speech sounds has been extensively investigated. Not only is this type of imitation very prevalent, but also people are often aware of speech imitation. People can generally recall instances in which they have imitated a friend's word choice, accent style, or speech mannerism. Speech imitation occurs in every realm of linguistic study, from phonetics (studies on phonetic convergence cited in this section) to lexical word choice (Garrod & Doherty, 1994) and syntax (Pickering & Ferreria, 2008, Bock, 1986), and while sometimes it is obvious and people are conscious of when they exhibit imitation, there are other forms of imitation that are so subtle they would not be consciously recognized.

Speech is a ripe candidate for imitation. Every person's speech is unique due to the individual vocal anatomy and physiological differences between all humans. Yet, despite all of

the differences exhibited by speakers, there is also a lot of similarity. These similarities are demonstrated forms of social group membership. Accents and speaking styles can be attributed to a number of different social factors: gender, socioeconomic status, age, geographical region, sexual orientation, etc. These stylistic speech variations help with group identification and social membership. Imitation can be a powerful tool to be used for including or excluding others in-group membership.

Even infants demonstrate the ability to imitate speech sounds. However, they do not show this ability until after 12 weeks of age, much later than when they imitate facial gestures (Kuhl & Meltzoff, 1996). Before then, however, they do imitate facial correlates of speech; newborn infants are shown to imitate mouth movements correlated with /ma/ but they do not reliably imitate the sounds (Chen, Striano, & Rakoczy, 2004). As children grow, they move from simply imitating speech *sounds* to imitating at a broader level. In a study on babbling in 12-month-old infants, subjects demonstrated differences in babbling according to the language of their interlocutor; infants produced more multisyllabic utterance patterns when interacting with a speaker who also produces more multisyllabic utterance patterns, as long as they had experience in both languages (Ward, Sundara, Conboy, & Kuhl, 2009). The finding that infants readily imitate people while learning their first language suggests that this may also be an effective approach to learning that may be employed in second-language acquisition (see work by Meltzoff and colleagues, cited in this dissertation).

One of the most studied forms of subtle imitative speech behavior is a process referred to as phonetic convergence (also referred to as alignment, or accommodation). This process is when interlocutors alter the fine phonetic detail in their speech to sound more like the person they are interacting with, even without instructions to imitate. Phonetic convergence refers to a specific

type of imitation, namely implicit speech imitation, and is exhibited in just the fine, sub-phonemic detail in the speech signal. Phonetic convergence occurs even in asocial laboratory conditions; talkers in a laboratory immediately shadowing a heard voice converge to the heard speech (Goldinger, 1998, Pickering & Garrod, 2004, Shockley, Sabadini, & Fowler, 2004, Nielsen, 2008, 2011a). Although implicit, convergence can be socially motivated, and the amount of convergence (or divergence – when fine phonetic detail is altered to make speech more different than an interlocutor’s) can vary with the characteristics of the sound or the talker being imitated. These effects will be discussed in the following section.

2.2 Factors affecting implicit imitation of known sounds in the auditory domain

To determine the factors affecting imitation of new sounds in the audiovisual modality, as will be done in this dissertation, it is important to recognize the factors that have been shown to affect implicit imitation when speech is presented in the an interlocutor’s speaking style, and this section will review key factors shown to be relevant to this process.¹

Like other forms of imitation, phonetic convergence is often believed to be at least somewhat socially motivated.² Convergence must be a social activity due to the nature of imitation (imitation needs a person to observe and another one to be observed), but it is also

¹ In addition to the factors discussed below, which are speaker independent, there are also inherent speaker dependent variations, which could affect imitation, particularly the physiological factors such as the physical size and anatomy of the vocal organs dependent on height and gender (Johnson, 2006, Peterson & Barney, 1952), but we will focus solely on other factors, as we will not be addressing these factors in our study.

² There are varying theories in the literature on phonetic convergence about whether it is a controlled (socially motivated) or automatic process, but there are substantial and compelling claims that even if it is automatic and uncontrolled, social factors still have an effect on convergence (see Babel, 2009, for an extensive review of this debate).

demonstrated to be modulated by social attitudes or biases, and by characteristics of a speaker. For example, female participants are likely to converge more with males than other females, and converge a greater proportion of the time than male participants (Namy, Nygaard, & Sauertieg, 2002). Alongside gender, the speaker's race, role in the conversation, and apparent social prejudices can also have an impact on the degree to which an interlocutor converges in a conversation (Babel, 2009, 2010, Bourhis & Giles, 1977, Pardo, Jay, & Krauss, 2010). Additionally, the perceived attractiveness of a speaker can affect how much a subject will implicitly imitate that speaker (Babel, 2009). Thus, speakers have some control over this process, and it is a readily available strategy for maintaining social distances.

Phonetic imitation, even just in the auditory domain, does not occur to an equal degree across all types of sounds: the degree of imitation can be affected by the qualities of the sounds being imitated. For example, exposure to a only few words starting with voiceless stop consonants with modified voice onset time (VOT) allowed for implicit imitation of this modified VOT for other consonants as well as for new words (Nielsen, 2008, 2011a). This imitation generalized to a featural level. Specifically, exposure to modified voice onset time for two of the three voiceless stop consonants in English (/p/ and /t/) facilitated imitation for the other voiceless stop consonant (/k/), demonstrating that speakers were imitating the feature, not just a pronunciation of a particular sound (Nielsen, 2008, 2011a). This research also showed that VOT in voiceless stop consonants is imitated more closely by English speakers when it is lengthened, rather than reduced. This can be attributed to the fact that when the voice onset time for English voiceless stop consonants is reduced, they become acoustically more similar to voiced stops (Nielsen, 2008, 2011a). Imitation is modulated by this distance between sounds, and occurs only when the resulting sound is not perceptually confusable with another sound. In sum, imitation

appears to be affected by phoneme categories, and can be generalized to other words and similar sounds.

Similarly, the degree of implicit imitation in vowels has been shown to depend on a number of factors specific not just to the phonetic inventory of the language but also the status of the vowel within the language, i.e., how stable pronunciations are across dialects. Vowels that have a larger accepted pronunciation range - with larger phonetic categories that are in less cramped vowel space - are imitated more closely (Babel, 2009, 2010). In a study looking at how social prejudices affect convergence, Babel (2010) showed that vowels with stigmatized pronunciations across dialects were imitated less, or even showed divergence, perhaps potentially to maintain a degree of distance from the interlocutor. However, for vowels that showed variations across the two dialects in question, but were not stigmatized, convergence occurred between the speakers. In sum, implicit vowel imitation varies according to (a) whether the pronunciation of a vowel is socially meaningful (stigmatized or non-stigmatized dialect variations) and is also likely to be affected by (b) whether convergence would place that vowel's pronunciation too close to another vowel's pronunciation (analogous to VOT not being imitated as closely when the resulting VOT would be too close to category boundaries).

Imitation does not just occur between sociolinguistically varying dialects but it also occurs between native and non-native dialects. Interlocutor language distance has been shown to affect the degree of phonetic convergence, but results have been equivocal regarding whether maximizing the degree of difference between interlocutors allows for more or less convergence. In two studies, Kim and colleagues analyzed convergence between speakers of native and non-native dialects. In the first study looking at convergence within a conversation between (a) native English speakers, (b) Korean speakers of different regional dialects, and (c) native and non-

native English speakers, they found that listeners perceived speakers of the same dialect to demonstrate greater convergence than speakers of two different dialects, or speakers with different native languages (Kim, Horton, & Bradlow, 2011). In another study, however, Kim found that native English-speaking subjects implicitly imitated the duration of the non-native speech more closely than the duration of native speech, although she admits that methodological factors (lack of controlling properly for their stimuli, specifically related to second mention reduction) could be affecting the results (Kim, 2011). This, combined with the fact that in the original Kim et al. study one pair of speakers per group was investigated shows that although there seems to be some effect of the degree of the distance between the dialects of the interlocutors, the exact nature of this effect remains unclear.

Convergence has been observed in children as well as adults. Children in conversational interaction implicitly imitate turn-taking pauses as well as speaking rate (Street & Cappella, 1989; Eaton & Bernstein Ratner, to appear). Using a repetition task, convergence on duration measures has been shown in the speech of children as young as 4-years (Ryalls & Pisoni, 1997). Additionally, convergence on phonological measures in speech imitation has recently been shown for 3-4 year olds: children imitated phonological consonant reduction and speech timing measures in heard speech without explicit directions to do so (Eaton & Bernstein Ratner, to appear). Thus far, only one study has looked at phonetic feature imitation in children. Recently, Nielsen (2011b) showed that 9-10 year olds exposed to artificially lengthened VOT values for /p/ adjusted their VOT pronunciation, converging with the talker. They not only generalized these new VOT values to other words with /p/, but they also extended this imitation to words with /k/, showing feature-level generalization. Convergence in children in this study was found to be at least to the same degree as in studies with adults, if not greater (no statistical comparisons were

made). Perhaps this appearance of a greater level of convergence is due to the fact that Nielsen used child-directed speech with these children; possibly that register either made the speech more imitable or more interesting. Another possibility suggested by Nielsen is that children may have been more easily influenced by speech heard in the experimental paradigm given that they have less accumulated exposure over their lifespan. In other words, they have less well-formed phonetic representations due to less input. This would be consistent with the observation that children are learning to adjust their pronunciations. Therefore, children may be better imitators than adults. To tease apart these two hypotheses – whether children are better imitators or whether child-directed speech facilitates greater levels of convergence – we need to (a) compare convergence by adults and children on the same register, and (b) test children on imitation of adult-directed speech.

All of the above-described research on phonetic convergence focused solely on the auditory presentation of speech. In the previous chapter we established that visual cues are important to speech perception and recent research shows that speakers seem to be sensitive to more than just the auditory cues when they are imitating a person in a convergence task. Subjects shown a silent video of a model talker in a repetition task converged with the talker's articulations, even in the absence of auditory cues (Miller, Sanchez, & Rosenblum, 2010, Sanchez, Miller, & Rosenblum, 2010). Recently, it has been shown that visual speech not only elicits convergence on its own, but that convergence was determined to be greater for audiovisual speech than for auditory-only speech; talkers who had visual access to an interlocutor converged in conversation more with the interlocutor than talkers who could only hear their interlocutor (Dias & Rosenblum, 2011). Together, these two studies show that the speech information in the visual modality can facilitate imitative behavior in adults. However,

both of these studies use perceptual measures to establish imitation. In the absence of measures of the specific advantage provided by visual cues (what acoustic or articulatory aspects of the speech are affected by visual cue presentation), it is difficult to determine the exact contribution of the visual modality to the convergence process. In the section below will this methodology is contrasted with other ways of measuring convergence.

2.3 How is imitation determined?

There are two main experimental methods to measure imitation. Findings of phonetic convergence or imitation are evident both perceptually and acoustically. In a number of studies convergence is demonstrated by presenting results from an imitation task to a new set of listeners and asking them to identify similarities in the speech of the model talker and the experimental subject (following the model of Goldinger, 1998). In these AXB tasks, a new group of subjects is presented with the model talker's articulations (X) as well as samples of the experimental subject both before and after exposure to the model talker (A and B). The new subjects are asked to identify which articulation (A or B) is more similar to the model talker's. These tasks focus on simply showing that imitation / convergence occurs and do not evaluate what characteristics are specifically being imitated.

Convergence results have also been measured acoustically. Acoustic measurements have shown that there can be imitation at the level of phonetic features, in the production of VOT (Shockley et al., 2004, Nielsen, 2008, 2011a, 2011b) and vowel formants (Babel, 2009, 2010), as well as in timing measures which look at broader prosodic and conversational effects, such as in pause and word durations as well as f_0 (Natale, 1975a, 1975b, Kim, 2011, Babel & Bulatov,

2011, as well as work by Gregory & colleagues, 1982, 1993, 1996, 1997, 2001, with adults; Street & Cappella, 1989, Ryalls & Pisoni, 1997, Eaton & Bernstein Ratner, to appear, with children). For the VOT studies, the specific VOT values in the subject's pre- and post-exposure productions and the model talker productions are compared. The studies on vowels compare formant proximity using Euclidean distance measures (Babel, 2009). In this kind of an analysis, all vowel formants are converted to the Bark scale and the Euclidean distance between the subjects' formants (which correlate with the openness and height dimensions of vowel sounds) and the model talker's, before and after exposure to that model talker, is compared. This type of analysis suggests (although it does not establish empirically) that at least some of the differences that subjects perceive in AXB tasks are measurable in the acoustic signal, and also provides a way to index degree of convergence on particular dimensions of the model talker's speech.

Each type of analysis has its advantages and its disadvantages. In a study on the acoustics of imitation, the researchers are limited because they must have a specific acoustic feature in mind, and then evaluate whether there is imitation of that feature. But perhaps subjects instead imitated a different feature of the speech. In an AXB task using listener judgments, global imitation – imitation of all the acoustic characteristics of the speech signal - is evaluated, but there is no way to pick out the specific characteristics that listeners are identifying as being imitated. For the purposes of this study, it is more important to look at the acoustics of imitation because we are interested in which specific vowel features are imitated in the auditory and visual modalities.

Thus far, only one study has compared the two measures of convergences. Babel and Bulatov (2011) looked at f0 imitation in a word shadowing task. They compared acoustic measurements of f0 imitation with listener judgment results using the same data set. They found

significant imitation in both the acoustic measures and in the perceptual measures. While the results showed the same pattern of data, the acoustic measurements and AXB data were not significantly correlated, and showed slightly different patterning, which can be attributed to f_0 being the only variable of analysis for the acoustic measures. The listener judgment (AXB) measure takes into account the entire signal, not just the particular variable measured (i.e., f_0); however, due to acoustic manipulations and the design of their stimuli, they believed that f_0 was likely the only target for imitation. It is unclear how a comparison would look when more acoustic measurements were taken into account.

For the purposes of our study, we are particularly interested in featural imitation of the vowel, not simply whether or not there is imitation. For this reason, a perceptual study will not suit our purposes, because there is no way to analyze exactly what is guiding a listener's rating. We are interested instead in which particular acoustic features are imitated, in order to assess the contribution of visually salient and non-salient information in the speech signal.

Chapter 3

Adult-directed speech vs. child-directed speech

Recall in the previous chapter on imitation, we discussed a finding by Nielsen that children imitated voice onset time of stop consonants more closely than adults (Nielsen, 2011b). One of the possible explanations for this was the use of a different speaking style, or register: child-directed speech rather than adult-directed speech. Perhaps this register, as a learning register, is more amenable to imitation? Speech to children tends to not only increase attention (Schachner & Hannon, 2011), but it also provides perceptual advantages (Thiessen & Saffran, 2005, Singh, Nestor, Parikh, & Yull, 2009) by maximizing certain acoustic characteristics of the speech signal. In this chapter, we will provide an overview of some of the major differences between adult- and child-directed speech in order to provide a background for why child-directed speech may be a more imitable register.

Note that the comparison of adult-directed speech with infant-or child-directed speech is an entire field of study. Therefore, in this chapter, we will not be reviewing all of the literature on this topic, rather just a few main findings that will be relevant to our study. We will limit our discussion of differences between adult- and child-directed speech to the phonetic characteristics (there are also syntactic, semantic, lexical, and phonological differences). In the first subsection below, we will discuss differences in how adult- and child-directed speech sound, and we will look at a few studies looking at acoustic differences in speech in the two register. In the next subsection, we will review one study, a study on visual differences between adult- and child-directed speech, which will be relevant to our experiments, as we will be looking at imitation

responses to speech in the auditory and audiovisual modalities. In the following subsections, we will then discuss why child-directed speech is important, and then finally, how it differs from clear speech, or speech to foreigners. We will end with a review of the predictions for our experiments based on the literature presented in this chapter.

3.1 What are the acoustic differences between adult- and child-directed speech?

The phonetic differences between adult- and child-directed speech in English are well established in the literature. The most prominent characteristic (as well as the most well-known and well-studied) is the intonation or prosody of child-directed speech. Child-directed speech has a different pitch pattern than adult-directed speech, and is known by its higher mean f_0 , and larger range of f_0 values and f_0 variability (Remick, 1976, Garnica, 1977, Stern, Spieker, Barnet, & MacKain, 1983, Jacobson, Boersma, Fields, & Olson, 1983, Fernald, 1984, Grieser & Kuhl, 1988, Fernald, Taeschner, Dunn, Papousek, de Boysson-Bardies, & Fukui, 1989, among many others). In fact, the pitch of child-directed speech is such a strong feature of this register, that speakers will be faithful to the contours even when it affects the phonemic use of pitch (Grieser & Kuhl, 1988). The second most well-known feature of child-directed speech is in its timing. Child-directed speech shows an increase in the amount of time in a pause, as well as the amount of pauses within an utterance (Stern et al. 1983, Fernald, 1984, Fernald et al. 1989). As a final measure of timing, similar to VOT in that it is in the segmental realm, child-directed speech also shows longer duration in vowels (Ferguson, 1964, Garnica, 1977, Sachs, 1977, Snow, 1977).

In addition to the prosodic differences and the differences in speech timing, there is also a well-known difference between child- and adult-directed speech in regards to vowel production.

In child-directed speech, there is an increase in the acoustic distance between vowels, resulting in an expanded vowel space with more extreme formant values (Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, & Ryskina, 1997). In other words, child-directed speech spans a greater region in the vowel space than adult-directed speech, providing clearer unambiguous speech samples for listeners. These findings are seen cross-linguistically.

In summary, child-directed speech tends to have a greater pitch range, a higher pitch, and is slower in duration compared to adult-directed speech (Fernald & Simon, 1984; Swanson, Leonard, & Gandour, 1992; among many others). Child- and adult-directed speech also differ segmentally; vowels tend to differ acoustically to a greater degree in child-directed speech compared to adult-directed speech (Kuhl et al. 1997). These modifications change the acoustics of the input to children, and it has been suggested that this serves, not just to make the speech clear and distinct, but also to draw attention.

3.2 What are the visual differences between adult- and child-directed speech?

There has been little research comparing the visual cues in child- and adult-directed speech. However, we can draw inferences about these differences based on two lines of research. First, in adult-directed speech, a number of studies have looked at what visual speech cues can be associated with prosody and supra-segmental characteristics. Specifically, pitch cues, which are more likely to differ across adult- and child-directed speech, are moderately correlated with eyebrow movement (Cave, Guaitella, Bertrand, Santi, Harlay, & Espresser, 1996) and head movement (Yehia, Kuratate, & Vatikiotis-Bateson, 2002). However, adults were better at perceiving pitch accent when they had access to the bottom half of the face than when they had

access to the top, indicating that although eyebrows and head movements are correlated with prosody, they are not necessary to detect it (Lansing & McConkie, 1999). In the lower half of the face, lip opening, chin lowering and dynamic movements associated with the chin have been implicated as visual correlates of lexical and phrasal stress (Scarborough et al. 2009).

Considering the features of child-directed speech, we can compare these features to what we know about the connection between visual cues and articulator movements in order to readily abstract some information about how visual cues in child-directed speech may be different from adult-directed speech.

Given the prosodic differences between adult- and child-directed speech, these differences in visual correlates of prosody are likely to be amplified in child-directed speech. Along the same lines, recall that child-directed speech has a slower rate than adult-directed speech. This slow rate of speech should allow for a longer duration of access to the visual cues, such as a greater duration of jaw opening, thereby enhancing visual speech cues.

There is one study that specifically looks at visual differences between adult- and child-directed speech. In this study by Green and colleagues, they looked at facial differences in child- and adult-directed speech. They found that lip movements differed in vowel production between the two registers, with more exaggerated movements in child-directed speech (Green, Nip, Wilson, Mefferd, & Yunusova, 2010). They asked mothers to produce speech to adult interlocutors and their own infants, and analyzed the visual (and acoustic) characteristics of their speech. When they analyzed the specific articulatory movements that were exaggerated in child-directed speech, they found differences in vertical aperture of the mouth, or a larger jaw opening in child-directed speech, but there were no significant differences in vowel rounding or lip spreading. This is particularly interesting when you consider that rounding is already a visually

salient cue even in adult-directed speech. These adults maximized a visual cue in child-directed speech that is not usually salient in adult-directed speech, viz., the vertical aperture of the mouth, but for cues that were already salient visually, viz., lip rounding, they made no change. When comparing the visual differences with acoustic differences, the authors found little evidence that the exaggerated facial movements increased acoustic distance between the vowels. Rather, they suggested that child-directed speech simply had greater motion, perhaps exploiting infants' preference to look at moving rather than still faces (see Adamson & Frick, 2003, for a review of this literature). This study brings up an interesting prediction: if visually salient features are perceived better in the audiovisual register, then in child-directed speech, which makes features that are not normally visually salient more visually salient, will these newly visually salient features also be better perceived in the audiovisual register? In the current study, by using imitation to test this question, we will be examining not just whether these features will be better perceived with combined exposure to child-directed speech and the audiovisual modality, but whether this added advantage in perception can also translate to useful advances in production.

3.3 What benefits does child-directed speech provide?

Although child-directed speech is not a necessary part of language acquisition, as evidenced by those who learn language even without exposure to child-directed speech (Pinker, 1994), child-directed speech is thought to serve several functions. The most widely discussed function of child-directed speech is that it invokes more attention than adult-directed speech (Fernald, 1982, 1985, Werker & McLeod, 1989, Werker, Pegg, & McLeod, 1994). Infants prefer to listen to (Cooper & Aslin, 1990, Fernald, 1985) and look at (Werker & McLeod, 1989,

Werker et al., 1994) child-directed speech compared to adult-directed speech. Increased attentiveness to child-directed speech is thought to be related to its more positive emotional affect (Fernald, 1993, Werker et al. 1994, Trainor & Desjardins, 2002, Singh, Morgan, & Best, 2002). Research on the connections between vocal characteristics and emotional responses show a correlation and mapping between high pitch and positive emotional responses; higher pitch is said to signal sociability and non-aggressive temperament (Scherer, 1986, Morton, 1977). It is well understood that, whether there is a direct correlation or not, the child-directed register attracts and maintains attention better than adult-directed speech, and that it instills a more positive affective response than adult-directed speech.

However, this is not the only function attributed to child-directed speech.³ The child-directed register maximizes differences in the speech signal, and is thus, likely to provide clearer input (Bernstein Ratner, 1986, Fisher & Tokura, 1996). For example, the prosodic structure and durational cues evident in child-directed speech help with word-learning tasks, by making the word boundaries easier to detect (Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989, Thiessen, Hill, & Saffran, 2005) and remember (Singh et al. 2002). The hyperarticulation of vowels in mothers' speech, resulting in a maximized vowel space in child-directed speech is correlated with improved phonetic discrimination abilities of their infants (Kuhl et al., 1997, Liu, Kuhl, & Tsao, 2003). An analysis of the effects of pitch in child-directed speech showed that the contoured pitch of child-directed speech improved vowel discrimination (Trainor & Desjardins, 2002).

³ Child-directed speech is thought to benefit language learning (Bernstein Ratner, 1986, Fernald et al., 1989, Kuhl et al. 1997), encourage more positive emotions (Fernald, 1993, Werker et al, 1994), as well as facilitate performance in perceptual tasks (Trainor & Desjardins, 2002, among others).

3.4 Child-directed speech vs. clear speech, or speech to foreigners

Another commonly used learning register used in second-language is referred to as “clear speech” or “speech to foreigners”. Clear speech is relevant to more than just learning a sound or language; it is also relevant when speaking to listeners who may not perceive speech as easily, such as blind, deaf, or elderly interlocutors. This register shows many similarities to child-directed speech, such as frequent pauses, limited vocabulary, a tendency for repetition, as well as a simpler syntactic structure (Freed, 1981). This register also shows a number of differences from child-directed speech. Some examples is that these register show differences in amplitude, with higher loudness for speech for foreigners than to child-directed speech (Garnica, 1977), and in the added word plays and diminutives in child-directed speech (Ferguson, 1977).

Child-directed speech and speech to foreigners do have a number of similarities but there are also observable phonetic differences between these two registers. Biersack and colleagues compared child-directed, adult-directed, and foreigner-directed speech (with all registers produced by all subjects). They found that child-directed speech showed specific pitch characteristics that were not shared by either of the other registers; in child-directed speech, there were larger ranges of f_0 variation, and higher maximum f_0 values (Biersack, Kempe, & Knapton, 2005). They also noted similarities between child-directed speech and foreigner-directed speech in that in both registers, the rate of speech was slower than for adult-directed speech (although it was more significantly slower for foreigner-directed speech). However, the manner in which rate was slowed was different between the two registers: speech to children had longer segment durations, and speech to foreigners had longer pauses between words (Biersack et al. 2005). In

another study comparing these three registers, but this time in British English, in addition to prosodic measurements, there was a comparison of vowel hyperarticulation. The results showed that both the child-directed and the foreigner-directed register allowed for vowel hyperarticulation, yet once again, only in child-directed speech were there differences from adult-directed prosody (Uther, Knoll, & Burnham, 2007). One additional aspect of this study was that they also looked at whether each register resulted in positive emotive-affect responses, and they found that this was evident in child-directed speech, but not foreigner-directed speech. Child-directed speech elicited more positive emotion feelings from subjects than foreigner-directed speech, possibly due to its prosody (Uther et al. 2007).

In summary, speech to foreigners is similar to child-directed speech in that it shows increased acoustic differences in vowels and a slower speaking style, but does not show increased pitch differences like child-directed speech (Uther et al. 2007, Biersack et al. 2005). Child-directed speech also results in a more pleasant emotional response than speech to foreigners (Uther et al. 2007). The learning register designed for children is similar to the learning register for adults, with differences in pitch, affective response, and some timing measures, but similar in overall timing as well as in maximized vowel articulations. It is important to draw the connection between the two learning register in order to put into context adult imitation of child-directed speech.

Chapter 4

Relation of previous work to current study

The central goal of this study is to determine what speakers are imitating when they receive audiovisual compared to auditory exposure, as well as to determine what the source of an audiovisual advantage in imitation consists of. Does the addition of the visual cues give an overall, global advantage? Or does the advantage of adding the visual modality manifest only on visually salient features?

Recall the Traunmüller & Öhrström (2007) study discussed in the Chapter 1 which looked at perception of Swedish round and unrounded vowels by Swedish adults. Subjects used the cues that were most informative and reliable in perception when the auditory and visual cues mismatched. For example, speakers perceived rounding better if they had the visual cues to rounding, and they perceived vowel openness better if they had the auditory cues to openness (even if they were also presented with information in the other modality which contrasted this information).

In the current study, we tested whether this “information reliability hypothesis” can account for the facilitative effects of audiovisual exposure on imitation. For this purpose, we used French vowels differing in both rounding as well as openness (similar to the Swedish vowels used in Traunmüller & Öhrström, 2007). English has none of the front-rounded vowels in its phonetic inventory but it does have vowels varying in openness as well as vowels varying in rounding. Thus, the features of the unfamiliar front rounded vowels are familiar to English speakers, but the particular combination of the features is foreign. Based on the information

reliability hypothesis, if subjects get audiovisual rather than auditory exposure they are expected to converge on rounding features better, with the result that imitation should be maximal for acoustic correlates of rounding (the third formant) in the audiovisual condition. In contrast, there should be no benefit from audiovisual exposure when imitating differences in the open/close dimension.

Besides testing adults, we also test 4- to 6-year-olds' imitation, given reports in the literature that children may show great degrees of imitation compared to adults (Nielsen, 2011b). If children are more likely to imitate implicitly, we expected greater degrees of imitation for children than adults across all conditions. However, if English-speaking children are less able to integrate visual information than adults as shown in the literature (McGurk & MacDonald, 1976), then we expected adults, but not children to show a facilitatory effect of audiovisual exposure.

The second goal of the study was to determine how speaking register might impact imitation. Child-directed speech provides two advantages; because of its positive affect it attracts more attention in younger listeners; the distinctiveness of some phonetic and articulatory features are also enhanced in child-directed speech. If child-directed speech simply draws the child's attention more to the input, based on sociolinguistic research showing that subjects are more likely to imitate if they have positive feelings towards their interlocutor (Babel, 2010, Bourhis & Giles, 1977), children (and potentially adults, due to a positive affective response) should show greater imitation of speech produced in the child-directed register. The only previous study looking at imitation in the child-directed register found large amounts of imitation on voice onset time comparative to imitation in the adult-directed register (Nielsen, 2011b). However, she was also using child subjects for her study, so it is impossible to determine whether the greater

amount of imitation was due to the use of the child-directed register or the child subjects. We also tested adult and children's imitation of child-directed speech to distinguish between the two accounts.

Together, with all of these goals, we want to explore the role of imitation as a learning mechanism. We designed our experiment to test not just whether subjects would imitate in a particular register or modality, but whether they would imitate non-native pronunciations or non-native speech sounds. With this manipulation, we sought to determine how exposure in the audiovisual modality and the child-directed register (compared to the auditory modality and adult-directed register) would aid in acquiring a new sound in production. For this reason, we used stimuli produced by a French native speaker.

The French vowels used as stimuli in this experiment fall into two categories – ones that are English-like, but foreign in pronunciation, (/i/ and /u/) and ones that are foreign to English-speakers (/y/, /ø/, and /œ/). Having English-like and foreign sounds allowed us to determine how the extent of the benefit derived from audiovisual modality and child-directed register might be modulated by learners' experience with that sound.

Additionally, the use of /i/ and /u/ as vowel stimuli allow us to replicate a previous finding. Acoustic investigations show that American English and Parisian French /i/ have comparable first and second formant values (F1 and F2); however, Parisian French has a lower F2 for the vowel /u/, indicating that the vowel has a more fronted articulation in American English (Strange, Weber, Levy, Shafiro, Hisagi, & Nishi, 2007). Not only do /i/ and /u/ differ in their acoustic similarity across French and English, even within English there are sociolinguistic differences between these two vowels. The front unrounded vowel /i/ is relatively stable in dialects of American English (Babel, 2009), and so its range of pronunciation is not subject to

“meaningful variation”. This is in contrast to the vowel /u/ which is often fronted in California English (Babel, 2009), resulting in a higher second formant (Clopper & Pisoni, 2004, Clopper, Pisoni, & de Jong, 2005, Clopper & Pierrehumbert, 2008), but not New York English (Strange et al., 2007). Thus, /u/ has more socially meaningful variation across dialects of English (Babel, 2010). Given Trudgill’s (1981) predictions that vowels that vary due to sociolinguistic factors are subject to greater imitation because listeners are more accustomed to variation of these vowels, we expected to replicate previous findings of greater imitation in socially meaningful vowels, i.e., in the imitation of /u/ compared to /i/.

As a final component to our study, it is recognized that subject and speaker differences can affect imitation (see Babel, 2009 for a review). For example, it is well-established in the literature that gender can affect imitation of speech sounds; female subjects generally show greater convergence than male subjects, and male talkers are converged to more than female talkers (Namy et al. 2002, Pardo et al. 2006, 2010, Babel, 2009). Gender can also affect audiovisual speech integration (Irwin et al., 2006, Johnson et al., 1988, Watson et al., 1996, Desjardins & Werker, 2004). For this reason, we varied the gender of our subjects (but look at gender as a factor in our results), and used a single gender for a talker. A lesser-known source of variation in linguistic research is the relative level (within neurotypical adults) of autistic-like traits a subject has. The Autism Quotient questionnaire (AQ), previously discussed in Chapter 1, establishes a gradient degree of how much a person exhibits autistic-like personality qualities (Baron-Cohen et al., 2001) and scores are shown to affect performance on an imitation task (Mielke et al., 2013). This, combined with the literature on autism and audiovisual integration and imitation that has thus far been discussed suggest that score on the AQ questionnaire could

account for some of the variability in imitation across subjects. Thus, we included both gender and AQ score as variables in our analyses.

Chapter 5

Effects of modality and register on imitation by adults

Adults imitate the speech of the people around them. Research on imitation has largely been studied only with reference to the auditory cues of speech. More recent research shows that adults are sensitive to more than just the auditory cues when they are imitating a person implicitly. Adults shown a video of a model talker converged with the talker's articulations, even in the absence of auditory cues (Miller et al. 2010, Sanchez et al. 2010). Not only does visual speech elicit convergence on its own, but also convergence is greater for audiovisual speech than for auditory-only speech (Dias & Rosenblum, 2011). Thus, visual cues can be highly influential in the convergence process. The primary goal of the present experiment is to investigate the nature of the contribution of visual speech cues to implicit imitation of foreign vowel sounds.

A secondary question is to determine whether visual cues help with imitation of particular acoustic cues. Studies on integration of auditory and visual speech cues in perception tasks show that the uptake of specific features is modality specific (Traunmüller & Öhrström, 2007). In audiovisual integration tasks, speakers used the cues that were most informative and reliable (clearly able to distinguish between two features) in perception when the auditory and visual cues mismatched. For example, lip rounding, which visual cues could distinguish sounds clearly and reliably, was better exploited in the visual modality, whereas for openness, which auditory cues could distinguish sounds clearly and reliably, the visual modality was ignored and the auditory cues were used in perception. In this study, we look at whether specific acoustic qualities are imitated better with auditory or audiovisual exposure to a speaker. Are the sound

features that are clearly contrastive in visual modality also *imitated* to a greater extent with the addition of the visual cues?

Visual cues are not the only aspect of the speech signal that can affect implicit imitation. In a study on imitation by children, children imitated the voice onset time of their interlocutor more closely than adults in a similar task (Nielsen, 2011b). One of the possible explanations for this was the use of a child-directed speaking style, or register, rather than adult-directed speech. Infant- or child-directed speech differs in a number of ways from adult-directed speech: prosodically, with a greater pitch range, a higher mean pitch, and slower duration (e.g., Fernald & Simon 1984, Swanson et al., 1992), and segmentally, with a more expanded vowel space (Kuhl et al., 1997). There are also important visual differences between adult- and child-directed speech; the slower rate of child-directed speech allows for greater jaw opening, thereby enhancing visual speech cues to vowel height, and exaggerated prosody can produce differences in eyebrow, chin, and head movements. Additionally, lip movements have been shown to differ in vowel production between infant- and adult-directed speech, with more exaggerated movements in child-directed speech (Green et al. 2010). Lip opening in child-directed speech is maximized, a cue known to be less visually salient in adult-directed speech. Visual cues that were already salient in adult-directed speech, such as lip rounding, however, are not increased to any significant degree in child-directed speech. In the present experiment, we will compare imitation to vowels that vary on openness and rounding. By comparing imitation in infant/child- and adult-directed speech we hope to pinpoint the influence of speech register on convergence (if any) and examine whether speaking register affects imitation of fine phonetic measures differently in the auditory and audiovisual modalities.

Phonetic imitation, even just in the auditory domain, does not occur to an equal degree across all types of sounds: the degree of imitation can be affected by the qualities of the sounds being imitated. For example, voice onset time in voiceless stop consonants is imitated to a greater extent by English speakers when lengthened, rather than when it is reduced. This can be attributed to the fact that in English, voiceless stop consonants with reduced voice onset time become acoustically more similar to voiced stops and thus, more perceptually confusable (Nielsen 2008, 2011a). Additionally relevant is which cues are actually imitated in phonetic convergence; recent work has suggested that the majority of imitation found in natural imitation contexts might be found in global measures, such as duration and f_0 and although there is imitation shown in phonetic characteristics, it might not be as prominent (Mitterer, 2013). This is contrary to studies looking at formant imitation in vowels (Babel, 2009, Pardo et al. 2010).

The degree of imitation in vowels not only depends on factors specific to the phonetic inventory of the language, but also on the status of the vowels within the language, i.e., how stable pronunciations are across dialects. Vowels that have a larger accepted pronunciation range (larger phonetic category with less cramped vowel space) are imitated more closely (Babel 2009).

Finally, speakers generalize imitation behaviors to sounds not heard during exposure. Exposure to modified voice onset time for one of the three voiceless stop consonants in English facilitated imitation for the other two voiceless stop consonants, demonstrating that speakers were imitating a feature, not just a pronunciation of a particular sound (Nielsen 2008, 2011a).

While imitation is known to be an important learning mechanism in infancy and childhood, little to no research has evaluated how imitation could contribute to learning a new sound in adulthood. The last research question to be addressed in this chapter is how imitation

differs for types of sounds (English-like and foreign). In order to establish how imitation differs between English-like and foreign sounds, in this experiment imitation of both types of sounds will be tested. Imitation is shown to be better for sounds with a larger accepted pronunciation range (Babel 2009, 2010), but what about sounds that are not in the subject's language? Are these sounds imitated better because they are judged to be a "new" sound, and therefore not constrained by the phonetic system of the native language?

Overall, the present experiment seeks to investigate *how* visual cues aid in imitation: is there a global advantage to the visual cues, or do they give an advantage only for visually salient sounds? What acoustic characteristics show evidence of a visual advantage? Additionally, as we know that the addition of the visual cues does allow for closer degrees of imitation, we are also interested in determining what other factors affect imitation in the auditory or audiovisual modalities (either alone or in conjunction with the addition of the visual cues). For example does speaking register, with the many acoustic differences as well as differences in attracting attention, affect the degree of imitation? Does the quality of the sound affect auditory and audiovisual imitation? In this study, we will be looking at imitation of foreign vowel sounds, either foreign vowels sharing the same phonological representation as English, or truly foreign vowels which do not exist in the English inventory.

In this experiment, we test monolingual English-speaking adults on auditory and audiovisual imitation of English-like and foreign sounds by comparing performance on a convergence task. We compare performance on this task by subjects exposed to an adult-directed speaking register and a child-directed speaking register.

Methods

Subjects

The subjects were monolingual English-speaking adults who confirmed in a language questionnaire that they had no extensive exposure to a language other than English, and no familiarity with French. Subjects were all UCLA undergraduate students who received course credit for their participation. Group A-A (auditory exposure, adult-directed register) consists of 20 subjects (females = 11). Thirteen of these subjects filled out a personality questionnaire, the Autistic Spectrum Quotient (AQ). The mean AQ score in Group A-A was 15.4 (range – 4:28; lower scores indicate a lower degree of autistic traits). Group AV-A (audiovisual exposure, adult-directed register) consists of 18 subjects (females = 12). All subjects in Group AV-A took the AQ questionnaire (mean score = 14.4, range 7:33). Group A-C (auditory exposure, child-directed register) consists of 17 subjects (females = 13). All subjects in Group A-C took the AQ questionnaire (mean score = 15.2, range 11:25). Group AV-C (audiovisual exposure, child-directed register) consists of 18 subjects (females = 11). All subjects in Group AV-C took the AQ questionnaire (mean score = 14.2, range 6:25).⁴

Speaker

The stimuli were produced by a male native speaker of French and English. A male speaker was used because previous studies on convergence report more convergence *to* a male

⁴ Additional participants were recorded but not used for the experiment due to poor recording quality or they were not native English speakers. (n = 46)

than a female talker (Namy et al. 2002, Pardo, 2006, Pardo et al. 2010, Babel, 2009).⁵ A screenshot of the audiovisual stimuli recording showing the speaker is below in Figure 5.1.



Figure 5.1. The speaker who produced the experimental stimuli.

The speaker who produced the stimuli was also phonetically trained, and a teaching assistant for introductory linguistics classes, which made him aware of the targets for the vowels. His instructions were to produce the words with French pronunciation. The speaker was born in a French-speaking region of Canada, but grew up in France and considers himself a speaker of

⁵ Phonetic convergence is shown more by female subjects, however, the tendency is for all subjects to converge more to a male model talker (Namy et al. 2002).

Standard French. The speaker is a French-English bilingual who learned English growing up in school.

Stimuli

The digital audio-visual recordings were made in a soundproof booth using a Sony digital HD handy cam (model HDR-HC7) and a Sony microphone (model ECM-MS907). The audio track and the video component of the recordings were separated using iMovie in order to create the auditory-only stimuli. All recordings were made with the speaker's face at a distance of approximately 3 feet from the video camera, with the microphone approximately 6 inches from the speaker's mouth, out of view of the video camera.

Two sets of recordings were made, in order to control for consistency across recording sessions. In the first set, the speaker recorded the stimuli with an adult-directed speaking style. The speaker was instructed to read the stimuli in a manner as if speaking to another adult. Next, the speaker recorded the stimuli with a child-directed speaking style. He was instructed to read the stimuli in a manner as if speaking to a child. Additionally, for the second set of recordings, a number of toys were brought into the room to facilitate a child-friendly setting. For both sets of recordings, the subject repeated each target word three times, and the clearest and most natural sounding stimulus was selected for the experiment.

The stimuli were modeled after Traunmüller & Öhrström (2007). The Swedish nonsense syllables that they used in their study were used in this study, alongside other nonsense words with the same CVC structure, and a set of filler words. All 75 target words were monosyllabic in order to avoid possible differences in vowel quality due to stress placement. The target sounds

being imitated were the set of front rounded French vowels varying in height /y/, /œ/, and /ø/⁶ as well as the vowels /i/ and /u/ which occur in both French and English. English has none of the front-rounded vowels in its phonetic inventory but it does have vowels varying in openness as well as vowels varying in rounding. Thus, the features of the unfamiliar front rounded vowels are familiar to English speakers, but the particular combination of the features is foreign. As mentioned in Chapter 2, while the vowels /i/ and /u/ are familiar to speakers of American English, the specific pronunciation of the French versions of these vowels is different than the American English pronunciation. Strange et al. (2007) describes the difference in vowel quality between Parisian French and New York English as lesser for /i/ than for /u/, and the main difference in the vowels in the two languages is for F2 of /u/. American English /i/ was fronter than Parisian /i/, but Parisian /i/ was higher, but both differences were slight compared to how much fronter American English /u/ was compared to Parisian French /u/ (Parisian French /u/ was also higher than American English /u/). Also as described in Chapter 2, the two English-like vowels /i/ and /u/ differ in their accepted pronunciation range and stability within the dialect of American English (/u/ shows more variation), and so these two vowels may show different levels of convergence.

All target vowels were embedded in CVC or CV contexts, where the initial consonant was from the following set [g, k, t, d, h]⁷. The final consonant, when present, was either /g/ or /k/. We chose these particular consonants because they provided the least amount of overt visual cues and therefore did not obscure visual information about the vowels, especially in the most

⁶ The speaker for the present experiment had a clear phonemic distinction between /œ/, and /ø/ in his French.

⁷ We included the consonant /h/ even though it is not a consonant in French, and so for some of the non-words, the speaker was presented with non-French consonants and French vowels, but since he was phonetically trained and a speaker of English, he was able to produce them.

relevant articulators, the lips. These words were modeled after Traunmüller & Öhrström (2007), but also included the alveolar consonants /t/ and /d/ to expand the range of contexts. The complete set of words used as target words in the experiment is listed in the table below.

Carrier	/y/	/ø/	/œ/	/u/	/i/
gVg	gyg	gøg	gœg	gug	gig
gVk	gyk	gøk	gœk	guk	gik
gV	gy	gø	gœ	gu	gi
kVg	kyg	køg	kœg	kug	kig
kVk	kyk	køk	kœk	kuk	kik
kV	ky	kø	kœ	ku	ki
hVg	hyg	høg	hœg	hug	hig
hVk	hyk	høk	hœk	huk	hik
hV	hy	hø	hœ	hu	hi
tVg	tyg	tøg	tœg	tug	tig
tVk	tyk	tøk	tœk	tuk	tik
tV	ty	tø	tœ	tu	ti
dVg	dyg	døg	dœg	dug	dig
dVk	dyk	døk	dœk	duk	dik
dV	dy	dø	dœ	du	di

Table 5.1. Wordlist of target stimuli for the experiment. The cells with the gray shading indicate the words that were used in the subset for the exposure phase.

Due to the need to use consonants that were not visually salient (as to not obscure visual cues produced in the vowels), the wordlist does contain five words that are actually English words. These words are ‘geek’ /gik/, ‘key’ /ki/, ‘he’ /hi/, ‘tee/T’ /ti/, and ‘D’ /di/. When evaluating the results of this study, we did not find this factor to influence the results, so it will not be discussed further.

In addition to the target words, there were an additional 10 words used as fillers, included to mask the purpose of the experiment. These words all contained bilabial voiceless stop

consonants as the onsets, and also contained the vowels used as English-like targets for the experiment, in addition to the vowels /ɑ/ and /e/. These fillers were also not analyzed.

pak	pag	pik	pig	pug
pip	peg	pek	pit	puk

Table 5.2. Filler words used as stimuli for the experiment.

Acoustic properties of the stimuli

The model talker's vowel productions are very similar to published reports on the acoustics of French vowels (Kim & Lee, 2001, Gendrot & Adda-Decker, 2005; Calliope, 1989). Table 5.3 below shows average formant values shown in past studies compared to the values seen for the stimuli in the current study; we show values for males and females, as in many cases our speaker's productions were intermediate between the two means. Thus, the speaker produced French vowels accurately.

Study	Vowel	F1	F2	F3
<i>present study</i>	i	338	2222	2929
Calliope, 1989 (males only)		300	2050	
Gendrot & Adda-Decker, 2005 (males)		310	2005	2784
Gendrot & Adda-Decker, 2005 (females)		348	2365	3130
Kim & Lee, 2001 (males and females)		250	2260	
<i>present study</i>	æ	507	1579	2212
Calliope, 1989 (males only)		500	1450	
Gendrot & Adda-Decker, 2005 (males)		400	1445	2440
Gendrot & Adda-Decker, 2005 (females)		436	1643	2715
Kim & Lee, 2001 (males and females)		510	1460	
<i>present study</i>	u	359	982	2220
Calliope, 1989 (males only)		350	850	
Gendrot & Adda-Decker, 2005 (males)		371	1105	2470
Gendrot & Adda-Decker, 2005 (females)		404	1153	2742
Kim & Lee, 2001 (males and females)		250	820	
<i>present study</i>	ø	457	1576	2196
Calliope, 1989 (males only)		400	1450	
Gendrot & Adda-Decker, 2005 (males)		384	1474	2405
Gendrot & Adda-Decker, 2005 (females)		420	1693	2687
Kim & Lee, 2001 (males and females)		370	1420	
<i>present study</i>	y	352	1982	2276
Calliope, 1989 (males only)		300	1800	
Gendrot & Adda-Decker, 2005 (males)		336	2005	2784
Gendrot & Adda-Decker, 2005 (females)		371	2063	2745
Kim & Lee, 2001 (males and females)		250	1810	

Table 5.3. Comparison of the present study’s adult-directed formant values with past studies’ formant values for the vowels used as stimuli in the present experiment. Blank cells indicate no measurement for that formant in the study.

The model talker’s mean vowel productions in each register are shown below in Figure 5.2 and 5.3 and the averages are summarized in Table 5.4. The methodology for the measurement of formants is discussed below in the “Analysis and Coding” section.

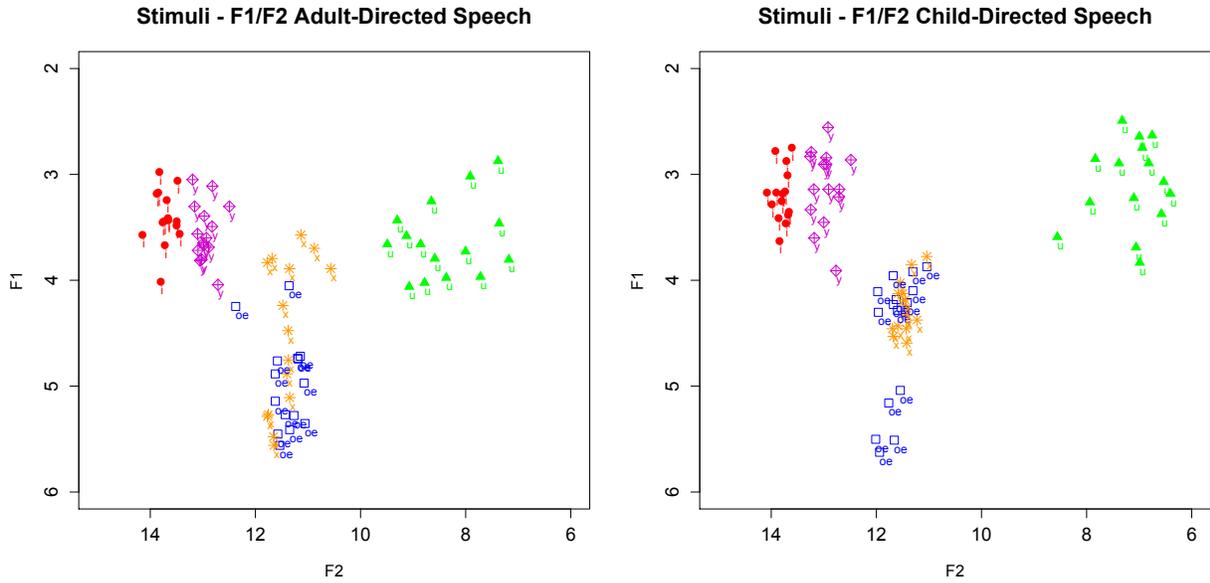


Figure 5.2. Vowel plots of F2 X F1 of adult- and child-directed stimuli. Adult-directed stimuli are on the left side, child-directed on the right. In these plots, we use ‘oe’ to symbolize /œ/ and ‘x’ to symbolize /ø/. Scale is in **Bark**.

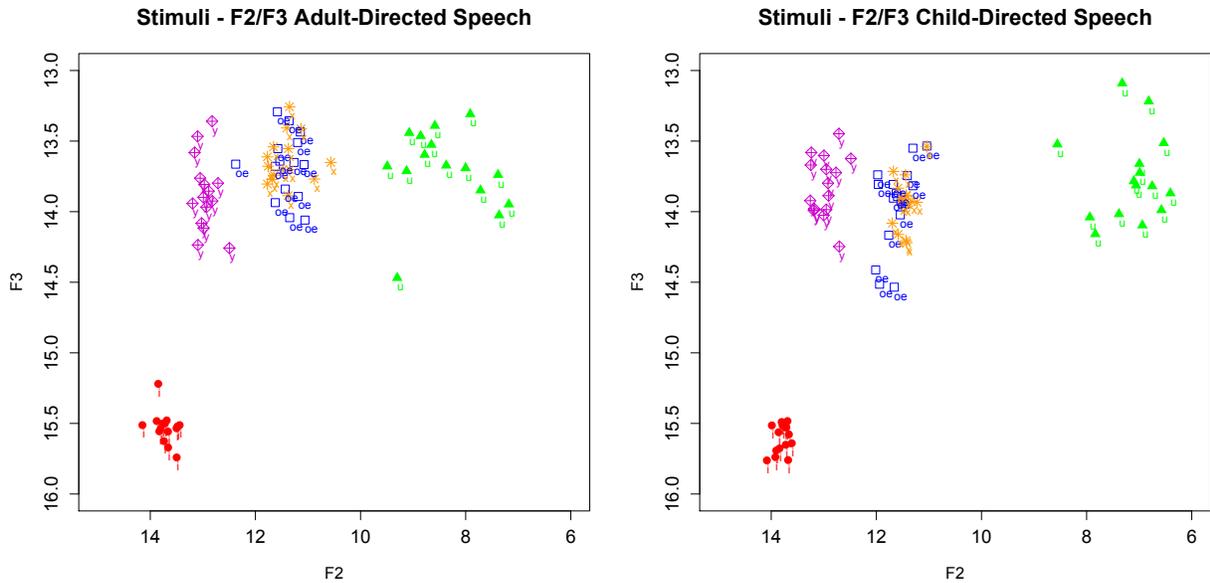


Figure 5.3. Vowel plots of F3 X F2 of adult- and child-directed stimuli. Adult-directed stimuli are on the left side, child-directed on the right. In these plots, we use ‘oe’ to symbolize /œ/ and ‘x’ to symbolize /ø/. Scale is in **Bark**.

Vowel	Register	Duration	f0	F1	F2	F3
i	ADS	173	184	338	2222	2929
	CDS	627	224	316	2251	2964
œ	ADS	212	167	507	1579	2212
	CDS	603	208	458	1629	2306
u	ADS	186	185	359	982	2220
	CDS	605	227	307	789	2238
ø	ADS	167	175	457	1576	2196
	CDS	615	215	426	1585	2295
y	ADS	157	184	352	1982	2276
	CDS	688	224	308	1988	2256

Table 5.4. Mean vowel measurements of all stimuli (range is evident in the plots above). Formant and f0 values are given in Hertz, duration in milliseconds.

Two-tailed t-test comparisons between the adult- and child-directed stimuli, looking at each formant, f0, and duration as separate variables, revealed significant differences between the adult directed stimuli and the child directed stimuli. Duration, f0⁸, and F1, and F3 were significantly different between the two sets of stimuli [Duration: $t(74) = -16.51, p < 0.001$; f0: $t(74) = -15.77, p < 0.001$; F1: $t(74) = 5.73, p < 0.001$; F3: $t(74) = -3.86, p < 0.001$], but F2 did not significantly vary across speech registers [F2: $t(74) = 1.28, p = 0.20$].⁹ The duration and f0 differences are expected because speech to children is known to be slower and have higher pitch (Fernald & Simon, 1984). However, the child-directed speech showed lower mean formant values for F1 (mean F1 ADS = 403 Hz/4.0 Bark; mean F1 CDS = 362 Hz/3.6 Bark)¹⁰, and slightly higher formant values for F3 (mean F3 ADS = 2366 Hz/14.1 Bark; mean F3 CDS =

⁸ The speaker had a rather high overall f0 for both adult- and child-directed speech.

⁹ These statistics were computed on the vowel formant values in Hz, not Bark, but t-tests on the Bark values showed the same pattern.

¹⁰ It looks in the plot as though F1 is higher for CDS than ADS, but it is just that F1 shows a larger range for CDS (249:584) than ADS (285:576).

2412 Hz/14.2 Bark). This is consistent with findings that speech to infants/children has an expanded vowel space (Kuhl et al., 1997).

Procedure

The task used in this experiment was a *modified* version of the implicit imitation paradigm (Goldinger 1998; Nielsen, 2008, 2011a). Since we were testing imitation of foreign vowels, we could not directly copy a methodology previously used in implicit imitation tasks (which includes word reading to establish the initial baseline pronunciation). Since subjects were unfamiliar with the foreign vowels, they could not be asked to read them. The procedure consisted of four phases: a pretest, an initial pre-exposure phase, an exposure phase, and a post-exposure test phase.

Pretest: Prior to the experiment, the subjects read a list of words designed to elicit the subjects' natural pronunciation of the English-like vowel /i/ and /u/ (we could not elicit natural pronunciations of the foreign vowels because the subjects have no established pronunciations of these vowels). Subjects saw a list of the words with pictures corresponding to the words and were asked to read them three times. This provided us with each subject's pronunciations before hearing the model talker ever utter the vowels. These words were designed to be words that a child would be able to identify from an accompanying picture (designed for the experiments with children, which will be presented in the following chapter). These words are listed below in Table 5.5.

Initial Consonant	Word with /u/	Word with /i/
p	pooh	“P” (the letter)
t	two	“T” (the letter)
d	dude	“D” (the letter)
k	coop	key
g	goose	geese

Table 5.5. Pretest wordlist items.

Pre-exposure Phase: Following the pretest reading, in the initial pre-exposure phase, subjects heard a production of each of the stimuli (no visual stimuli) and were instructed to simply repeat¹¹ the word that they heard. This initial pre-exposure phase included all target words and was subject controlled, lasting about 4-5 minutes. The productions were either all in adult-directed speech, or all in child-directed speech, depending on the condition that the subjects were assigned to.

Exposure Phase: Following the initial pre-exposure phase, the subjects underwent an exposure phase. During this phase, all subjects were exposed to three repetitions of a *subset* of the target words introduced in the pre-exposure phase, in order to be able to test for generalizations in the test phase. Half the subjects received *auditory exposure*, the other half received *audiovisual exposure*.

¹¹ Many of the alignment studies vary in the instructions that they give to participants, including, for example, “say” (Pardo, 2010), “identify the word you hear” (Shockley et al. 2004, Nielsen, 2011) and “repeat” (Nye & Fowler, 2003). We used “repeat” because of our use of non-words and foreign sounds. We thought to use “identify” would suggest that the word should be selected from the speakers’ lexicon, and “say” might suggest the same to the children. Also, our compared pre-exposure/post-exposure productions had the same instruction, and so any effects we see in imitation are cannot be due to the instructions themselves.

- In the *auditory exposure* condition, subjects saw a static image of the talker¹² and heard him producing the subset of the stimuli three times.
- In the *audiovisual exposure* condition, subjects saw the speaker audiovisually producing the subset of the stimuli three times.

In the exposure phase for the adult-directed speech, subjects heard a word every 2250 ms. In order to accommodate the longer durations of the child-directed speech stimuli, in the exposure phase for the child-directed speech, a word was presented every 2750 ms. The exposure phase lasted about 5-7 minutes, depending on whether the subject took breaks.

Post-Exposure Phase: Upon completion of the exposure phase, subjects participated in a post-exposure test phase. This post-exposure phase was identical to the initial pre-exposure phase. Subjects were instructed to repeat the word they just heard. Subjects heard all target words in this phase, including those not shown during in the exposure phase, in order to determine if subjects generalized to words not presented in the exposure phase. Total testing time was about 15-20 minutes. See Figure 5.4 below for a summary the experimental design (following the pretest).

¹² A static image of the talker was included in order to mask any effect in the audiovisual exposure condition due to the speaker's appearance, or due to the differences in social situations.

Phase	Auditory exposure		Audiovisual exposure	
	<i>See</i>	<i>Hear</i>	<i>See</i>	<i>Hear</i>
<i>Initial pre-exposure phase</i>		STIMULI ITEMS (all)		STIMULI ITEMS (all)
<i>Exposure phase</i>	SPEAKER STATIC FACE 	STIMULI ITEMS (subset)	VIDEO OF SPEAKER ARTICULATING	STIMULI ITEMS (subset)
<i>Post-exposure test phase</i>		STIMULI ITEMS (all)		STIMULI ITEMS (all)

Figure 5.4. Schematic of the experimental design. In both the initial pre-exposure phase and post-exposure test phase, subjects saw a picture of clouds on the screen.

Testing took place in a sound proof booth in the UCLA Language Acquisition Lab. Subjects wore a lapel microphone and their productions in the pretest and both the pre-exposure phase and the post-exposure test phase were recorded audiovisually using a Sony digital HD handy cam (model HDR-HC7) and a Sony microphone (model ECM-MS907). The microphone was connected wirelessly to a desktop computer where the recordings were made using ProTools software (sampling rate of 44.1kHz, 16-bit resolution). Subjects were videotaped for the purpose of conducting an analysis of the degree of lip protrusion they exhibited throughout the study, but that analysis will not be included in this dissertation.

The experiment was presented to subjects using software specifically written for this study. Subjects controlled stimulus presentation during the pre-exposure and test phase through a laptop keyboard. The exposure phase was not subject controlled; rather, exposure stimuli were played in three parts, each including a third of the randomized exposure words. The exposure phase was broken into three parts in order to make it analogous to the child experiment presented in the following chapter, because the children needed more breaks than the adults. Additionally to make it analogous to the child experiment, there was a numerical indicator at the bottom of the screen indicating how far along in the experiment subjects were at any given trial. The stimuli in both the auditory and audiovisual conditions were randomized to control for any potential effects of order.

Analysis and coding

In order to assess how the factors in the present experiment affect convergence, we made acoustic measurements of the productions of our subjects. For our analysis, we measured the first, second, and third formants of the vowels in the pretest, pre-exposure, and post-exposure productions. The first formant (F1) is often correlated with vowel height (tongue height), whereas the second formant (F2) is correlated with vowel backness (tongue backness). A lower value for F1 is correlated with a higher tongue position, and a lower value for F2 is correlated with a more retracted tongue position. Lip rounding is suggested to correlate with F3 values, with rounded lip positions signaling a lower value for F3 (Stevens & House, 1955, Fant, 1959, 1960, 1983). In addition to measuring the formant values, we also looked at convergence on

vowel duration and fundamental frequency, to determine global, vocal tract-independent, ways in which the subjects might have imitated the model talker.

In the audio recordings from each subject, the vowels were first segmented and labeled in Praat; onset was taken to be where the first and second formant frequencies first became apparent in the spectrogram, and offset was where these formants clearly ended. Once these labels were created, acoustic analysis was run in VoiceSauce (Shue, Keating, & Vicenik, 2011). VoiceSauce measured a value every millisecond for the acoustic measures for the fundamental frequency and formants. We extracted the duration of the vowel, and the mean values for f_0 , F1, F2, and F3 for the middle third of the vowel. We used the Snack Sound Toolkit (Sjölander, 2004) to calculate formant frequencies and the STRAIGHT algorithm (Kawahara, Masuda-Katsuse, & de Cheveigné, 1999) to calculate f_0 values. To confirm that the formant frequencies were accurate, a second algorithm, Praat (Boersma & Weenink, 2013) was used to measure the formant frequencies. After the measurements were taken, the two sets of numbers were compared, and cases in which the formant value generated by the two algorithms was greater than 300 Hz were inspected by the author. In cases where it was clear, through comparison of the vowel and the values from the two algorithms, that the formant was not being tracked properly (ex. third formant was tracked in place of second formant), the secondary Praat algorithm value was used instead.

Once all measurements were adjusted, formant measurements were converted to the Bark scale, which approximates auditory distance (Traunmüller, 1990). Measurements were converted with the formula from Traunmuller (1990) in Excel. We then completed two sets of analyses. In the first set of analyses, we looked at difference in convergence within the two phases of the experiment. We calculated phonetic distances between the model talker's productions and the

subjects' pre-exposure (*pre-exposure difference*) and post-exposure productions (*post-exposure difference*) for each acoustic dependent variable, then computed *convergence* by subtracting the absolute value of the distance from the stimulus in the post-exposure production from the absolute value of the distance from the stimulus in the pre-exposure production, giving us our *convergence* measurement, following the methodology of previous studies (Babel, 2009, 2010, 2012). See Figure 5.5 for the formulas used in these calculations.¹³

$$\text{PRE-EXPOSURE DIFFERENCE}(Variable) = \text{Stimulus Word}_{Variable} - \text{Subject Word (Pre-Exposure)}_{Variable}$$

$$\text{POST-EXPOSURE DIFFERENCE}(Variable) = \text{Stimulus Word}_{Variable} - \text{Subject Word (Post-Exposure)}_{Variable}$$

$$\text{CONVERGENCE}(Variable) = |\text{PRE-EXPOSURE DIFFERENCE}(Variable)| - |\text{POST-EXPOSURE DIFFERENCE}(Variable)|$$

Figure 5.5. Equations used in calculating convergence across each phonetic variable. Variable refers to any of: duration, F0, F1, F2, or F3, individually.

The convergence analysis determined imitation on each acoustic cue individually, in order to assess whether imitation was better on a particular acoustic cue (ex. whether F3 imitation was better after auditory or audiovisual exposure) to make claims about vowel feature imitation.

We also wanted a sense of overall convergence, so we compared the degree of difference between the subject's pre-exposure and post-exposure productions by computing the Euclidean distance between the model talker's and the subjects' vowel pronunciations in each phase. We

¹³ We did not do any normalization (other than scaling to the Bark scale). While this is consistent with other studies, we could have chosen to normalize according to the subjects' vocal tracts by recording target vowels, and looking at convergence within their individual vowel spaces. This would get at perhaps a more accurate measure of convergence.

made one Euclidean distance calculation using just the first and second formant in order to make our data analogous to past studies (Babel, 2009), but we also computed a Euclidean distance measurement taking into account all three formants, since all three were relevant for our vowels. The Euclidean distance formulas (Figure 5.6) take into account the formant measurements together and provide one “degree of distance” number for the analysis, in order to assess overall convergence. For the analysis of duration and F0, we used the formulas in Figure 5.5 above; however, duration and f0 measures were not incorporated into the Euclidean distance calculations.

$$\text{ORIGINAL DISTANCE (ALL)} = \sqrt{(\text{Pre-Exposure Difference (F1)})^2 + (\text{Pre-Exposure Difference (F2)})^2 + (\text{Pre-Exposure Difference (F3)})^2}$$

$$\text{FINAL DISTANCE (ALL)} = \sqrt{(\text{Post-Exposure Difference (F1)})^2 + (\text{Post-Exposure Difference (F2)})^2 + (\text{Post-Exposure Difference (F3)})^2}$$

$$\text{ORIGINAL DISTANCE (F1 \& F2 ONLY)} = \sqrt{(\text{Pre-Exposure Difference (F1)})^2 + (\text{Pre-Exposure Difference (F2)})^2}$$

$$\text{FINAL DISTANCE (F1 \& F2 ONLY)} = \sqrt{(\text{Post-Exposure Difference (F1)})^2 + (\text{Post-Exposure Difference (F2)})^2}$$

$$\text{DIFFERENCE IN DISTANCE} = \text{ORIGINAL DISTANCE} - \text{FINAL DISTANCE}$$

Figure 5.6. Equations used in calculating the Euclidean distances for the pre-exposure and post-exposure productions, using all three formants, or just the first two. Also, the equation used for calculating the *difference in distance*.

For the Euclidean distance measurement, a positive difference reflects that the difference between the model talker’s production and the subjects’ production shrank between the pre-exposure and post-exposure phase, and subjects converged. A negative value reflects a finding of

divergence, or an increase in acoustic distance as the subject got more exposure to the model talker. This is contrary to what has been done in some other studies (Babel, 2009, which subtracts original distance from final distance and *negative* values indicate convergence), but we felt that intuitively, thinking of a negative difference as a positive result of convergence made less sense than the reverse.

The formant measurements were compared on a number of different factors. Between-subject factors included the condition (auditory or audiovisual), gender, and score on the Autism Quotient questionnaire. Within-subject factors included the type of vowel (English-like: /i/ and /u/, foreign: /y/, /ø/, and /œ/)¹⁴, the context (CV, CVC), the voicing and place of articulation of each of the surrounding consonants, and whether the particular word was included in the exposure session (recall, only a subset of the words were played here, in order to test for generalizations).

Two analyses of “convergence”

Note that both the pre-exposure and post-exposure productions are shadowed productions, so they are both likely targets for convergence to the model talker as neither are uninformed, natural pronunciations. In actuality, the comparison between the pre-exposure and post-exposure productions is a measure of *change* in convergence; this looks at whether convergence increases or decreases across a number of acoustic dimensions with increasing exposure to the model talker. Not finding a significant change in “convergence” in the pre-

¹⁴ Instead of English-like and foreign, we could have additionally separated /i/ and /u/ as we found they patterned differently in the pre-test/post-exposure comparison, to additionally classify them by feature (front unrounded, back rounded, front rounded). However, since the goal was to look at acquisition of sounds, we thought this would be the most relevant classification.

exposure/post-exposure comparison simply means there was no *change* in convergence after exposure to the model talker. This methodology was necessary due to our interest in whether subjects can imitate foreign sounds. For these sounds we were not able to get an actual baseline; subjects have no natural pronunciation of a foreign sound.

However, we were also interested in the basic question of whether there is convergence, and so for this purpose we analyzed pre-test recordings of the English-like vowels. These were looked at separately in our analysis; we compared the pre-test vowel productions with the post-exposure productions for the English-like vowels only. In this sense, this is the closer-to-true measure of convergence: whether the subject changed their production of a known sound after exposure to a model talker.¹⁵ Because pre-test items were different than the items produced by the model talker, and the number of tokens differed between the model talker and the subjects, for this portion of analysis, we compared the mean values for duration, f0, F1, F2, and F3. We slightly modified the equations above in Figure 5.5 to make these calculations, substituting the pre-test productions for the pre-exposure productions, and also comparing means rather than individual values. For this part of the analysis, we had the between subjects factors of condition (auditory or audiovisual), gender, and score on the Autism Quotient questionnaire. The only within-subject factor was the vowel being imitated (word was not included since we were looking at means).

¹⁵ However, pre-test words were read, and post-exposure tokens were shadowed, which meant that these were different types of tasks being compared here. Traditionally, imitation paradigms compare two instances of read pronunciations or shadowed pronunciations, but due to the nature of testing foreign vowels, this was not possible here.

Results

For the first subsection below, we will compare the pre-test results to the post-exposure task results, in order to analyze how far subjects' modified their natural pronunciations of the English-like vowels for the experiment.¹⁶ This is, in a sense, our only measure looking at absolute “convergence” because subjects already have an established pronunciation of the English-like vowels, and in this measure, we are testing how the subjects modify their natural pronunciation after exposure. In this section, we analyze how the experimental factors affected convergence on the English-like vowels, first answering the question “is there convergence?” and then moving on to an analysis of how the experimental factors affect this convergence. We begin with global measures of convergence, and then look at the more fine-tuned acoustic measurements.

In the second analysis section, we will evaluate how the exposure session affected implicit imitation between the two different sessions of the experiment for every experimental variable. Here, we are interested in looking at which factors (register, modality, carrier type) affected imitation between the two conditions. We will separate this data out by gender, looking at male and female performance separately (as has been done in Babel, 2009), to simplify our models as well as due to reports of difference in the extent of convergence between males and females (Namy et al. 2002, Pardo, 2006, Pardo et al. 2010, Babel, 2009). Like in the previous analysis section, we will first look at the question of “is there convergence?” before we move on to looking at which variations in the experiment affected subject performance.

¹⁶ In order to test how the subjects modified their natural pronunciations, we would have had to ask subjects to repeat a recording of the pre-test word list after the experiment was over

Following this, we have two brief subsections, the first looking at whether the patterns were generalized to new words. While this measure of generalization is not as robust as previous studies analyzing generalization in convergence, we do present a simple measure looking at this issue with vowels. Previous studies have looked at whether subjects generalized imitation of a feature to new sounds with that feature, or to new words. In our study, we simply look at whether subjects generalize to new words. Finally, in the last brief subsection, we will discuss whether the place of articulation of the onset consonant affected convergence. Recall that we included alveolar onset consonants, and we wanted to verify that these consonants, which could obscure some visual cues, did not significantly affect results (Traunmuller & Ohrstrom, 2007, only included velar consonants because they did want less-visually salient consonants).

In both analysis sections, we separate the variables into three subsections: global measures, aggregate phonetic measures, and individual phonetic measures. Global measures included duration and f_0 , aggregate phonetic measures included the two Euclidean Distance measurements, and individual phonetic measures included the three formants. We separated out the variables this way in order to conceptualize the analysis in terms of types of variables, and to reconcile our results with other studies looking at convergence, especially those using perceptual measures.

Pre-test vs. post-exposure comparison

In this section, we will compare the means of the pretest tokens to the means of the post-exposure tokens for the vowels /i/ and /u/. We could not compare individual tokens to each other,

because for the pre-test, we wanted the words to be recognizable to subjects to elicit correct pronunciations, while for the experiment, we wanted non-words. Since the tokens could not be matched at the word level, we are comparing the means for each vowel. Since there are varying predictions about convergence on the English-like vowels, we looked at /i/ and /u/ separately in this portion of the analysis. Two subjects did not complete the pretest and were excluded from this portion of the analysis (female, AQ = 15; female, AQ = 11).

Within each subsection of analysis, we report two separate analyses. For the first analysis, the variable for analysis was the difference between the model talker and the participant, in the pre-test and post-exposure exposure condition for each dependent variable [global measures of duration and F0, aggregate phonetic measures of Euclidean Distance (F1+F2+F3) and Euclidean Distance (F1+F2), and individual phonetic measures of F1, F2, and F3], and added a factor of Repetition (pre-test or post-exposure) to differentiate the two productions. We ran mixed effects models for each vowel separately with the fixed effect factor of Repetition, and the random effect of Subject for the entire data set (this allowed each subject to have a unique pronunciation), as well as within subsets of data based on Gender, Register, and Modality. A significant effect of repetition would mean that subjects either converged or diverged between their pre-test and post-exposure productions.

Following the analysis of whether there is convergence, in each subsection, we will also take a more in-depth look at how all of the experimental factors affect convergence. For this analysis, we do complex mixed effect models. The experimental design for this part of the analysis was a 2 (Gender: male or female) X 2 (Modality: audio or visual) X 2 (Register: child-directed or adult directed speech) X 2 (Vowel: /i/ or /u/) factorial design, with the AQ score included as a fixed covariate. We also included the random effect of Subject to allow each

subject to differ in his or her overall degree of convergence. For this portion, the dependent variable is the measure of *convergence* for each of the phonetic factors (the difference between the pretest difference and the post-exposure difference, see the coding section above for details). The 7 dependent variables were duration convergence, f0 convergence, F1 convergence, F2 convergence, F3 convergence, Euclidean Distance (F1+F2), and Euclidean Distance (F1+F2+F3). We will provide full models for each variable, including all main effects and two-way interactions, following the recommendations of Harrell (2001), Jaeger & Snider (2013), and Jaeger (2011).¹⁷ All modeling was done using the lme4 package in R (R Development Core Team, 2008). To determine the unique contribution of each variable, we compared models in a subset relationship using likelihood ratio tests. To confirm significance and to interpret pairwise comparisons, we subjected the data to a repeated measures analysis of variance, using Tukey's HSD test, also in R, for all factors except AQ. For AQ, we computed analysis of covariance on data subsets to further clarify the results, as both these measures were continuous.

Note that we could see an influence of a particular factor in the first analysis and not the second, or vice versa. While the first analysis looks at whether convergence is significantly different from zero, the second analysis takes into account relative degrees of convergence. Subjects may show differences in degree of convergence across conditions, with only one or even neither being significantly different from zero (ex. in cases where there was divergence in one condition, and zero convergence on the other). Then, convergence would not be significantly different from zero in either condition; however, the two conditions could be significantly different from each other. Alternatively, the degree of convergence across conditions might be

¹⁷ We report only two-way interactions because a full set of interactions would include over 80 different interactions, making the model too complex for analysis. Additionally, many of the complex 5-, 6-way interactions are difficult to interpret.

the same for a particular measure; however, due to within group variation, only one result might be significantly different from zero.

In model summary tables for each effect, we report the parameter estimate, the standard error, and two tests of significance: Wald's Z statistic, which tests whether coefficients are significantly different from zero, given the estimated standard error, as well as the χ^2 over the change in data likelihood, $\Delta(-2\lambda)$, associated with the removal of the factor or interaction from the final model. For the likelihood ratio test for the main factors, we tested with and without the factor *and its interactions*. Degrees of freedom are reported for these tests. Finally, we report pseudo R-squared as a measure of effect size for that variable. Significant main effects and interactions were identified based on the Wald's Z statistical tests.

Is there convergence between the pre-test and post-exposure productions?

Before we delved into the analysis, we wanted to ensure that subjects did show significant findings of convergence. In Table 5.6, we show the output for mixed effects models for the Repetition factor with all the data included in the analysis. Each row in the table represents the output for the factor of Repetition (pre-test or post-exposure) in separate mixed effect models for each factor for each vowel. The dependent variable is convergence on each measure (in ms. for duration, Hz for f0, and Bark for the formant measurements). As you can see, there was overall convergence on all experimental variables except F1 and F3. An inspection of the means reveal that the significant results are all results of convergence, not divergence.

The absences of overall convergence on F1 and F3 were qualitatively different – although analysis of subsets of data showed significant convergence on F1, no analysis with F3 as a

dependent variable showed a significant difference. So, we will not discuss F3 further for this pre-test/post-exposure comparison.

	Variable	Vowel	Estimate	Std. Error	t-value	p-value	Significant?
Global Measures	Duration	/i/	71.86	21.18	3.39	0.0009	*
		/u/	111.35	21.20	5.25	p < 0.0001	*
	f0	/i/	20.84	2.72	7.67	p < 0.0001	*
		/u/	19.40	2.98	6.51	p < 0.0001	*
Global Phonetic Measures	Euclidean F1+F2+F3	/i/	-0.18	0.04	-4.80	p < 0.0001	*
		/u/	-1.73	0.13	-12.97	p < 0.0001	*
	Euclidean F1+F2	/i/	-0.17	0.04	-4.78	p < 0.0001	*
		/u/	-1.64	0.14	-11.74	p < 0.0001	*
Individual Phonetic Measures	F1	/i/	-0.05	0.04	-1.12	0.2637	ns
		/u/	-0.08	0.05	-1.67	0.0980	ns
	F2	/i/	-0.15	0.04	-4.08	0.0001	*
		/u/	-2.09	0.15	-13.56	p < 0.0001	*
	F3	/i/	-0.03	0.03	-1.12	0.2650	ns
		/u/	0.02	0.05	0.39	0.6988	ns

Table 5.6. Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions.

We will now begin with analyses of each variable individually, in order to determine not just whether there was convergence, but what experimental factors affected convergence on each variable. In order to better understand trends in the data, we break up the experimental variables into three categories: global measures, aggregate phonetic measures, and individual phonetic measures.

Global Measures

Duration:

Recall from the previous section that we saw convergence overall for both /i/ and /u/. Looking at further subsets of data, we see convergence across both modalities, both registers, and both genders for /u/, but for /i/, we do not see convergence overall for male participants, and the auditory modality overall. Results for gender by modality are shown in the graph below.

		Overall	Register		Modality		Gender	
			Adult-Directed	Child-Directed	Auditory	Audiovisual	Males	Females
Duration	/i/	✓	✓	✓	✗	✓	✗	✓
	/u/	✓	✓	✓	✓	✓	✓	✓

Table 5.7. Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for duration. The full statistical output is shown in the appendix.

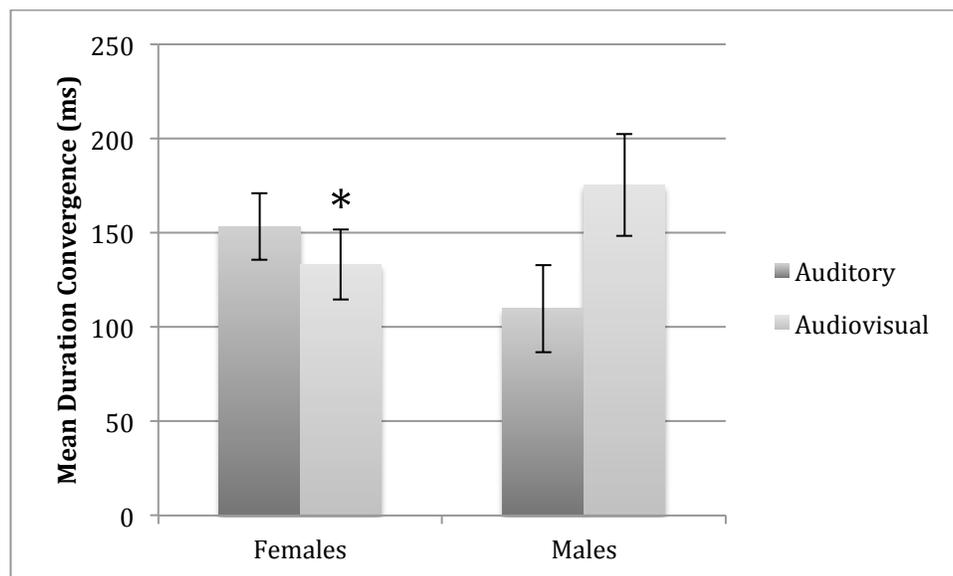


Figure 5.7. Comparison of pre-test and post-exposure convergence in each gender by modality for adult subjects on the duration measure. Significance stars represent significant convergence or divergence (difference from zero) with a *p*-value of 0.05 or smaller.

The final output for the mixed effects model including all factors and interactions for duration is presented in Table 5.8. Table 5.8 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction.

<i>Adult Pretest Duration</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Pseudo-R^2</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c^2</i>	<i>df</i>	<i>p</i>	
Register	85.96	63.48	1.35	0.178	110.57	5	<0.001	0.0320
Modality	68.16	62.63	1.09	0.279	7.81	5	0.167	0.0014
Gender	117.48	80.81	1.45	0.149	9.76	5	0.082	0.0024
Vowel	-28.91	17.73	-1.63	0.106	40.86	5	<0.001	0.0420
AQ	3.97	3.11	1.28	0.204	153.48	5	<0.001	0.0008
RegisterXModality	44.89	38.13	1.18	0.242	1.59	1	0.207	0.0000
RegisterXGender	34.77	39.99	0.87	0.387	0.85	1	0.357	0.0000
ModalityXGender	18.32	39.96	0.46	0.648	0.19	1	0.659	0.0001
RegisterXVowel	67.86	9.95	6.82	0.000*	37.24	1	<0.001	0.0382
ModalityXVowel	13.38	10.06	1.33	0.186	1.95	1	0.162	0.0013
GenderXVowel	-18.37	10.44	-1.76	0.081	3.34	1	0.068	0.0024
RegisterXAQ	2.83	3.58	0.79	0.432	0.69	1	0.407	0.0000
ModalityXAQ	-6.40	3.54	-1.81	0.073	3.78	1	0.052	-0.0001
GenderXAQ	-7.77	4.58	-1.69	0.093	3.32	1	0.069	-0.0001
VowelXAQ	-0.39	0.94	-0.42	0.677	0.25	1	0.617	0.0001

Table 5.8. Summary of results from the mixed effects model for duration for adult participants in pretest tokens compared to post-exposure production tokens.

Overall, there were no significant main effects, but there was a significant interaction of Register X Vowel (Figure 5.8). Post-hoc testing found significant pairwise differences between the two vowels across register (child /i/ was different from adult /i/ and /u/, $p < 0.0001$, and child /u/ was different from adult /i/ and /u/, $p < 0.0001$), but there were no significant differences between the two vowels in the same register (adult vowels: $p = 0.163$; child vowels: $p = 0.261$).

Note that we saw convergence for both registers, so this result is a difference in the amount of convergence. It is not surprising to find more imitation in the child-directed register with regards to imitation of duration because the child-directed speech stimuli showed a much larger original duration difference from the pre-test tokens. Thus, there was a large duration difference to which subjects could converge to in this register.

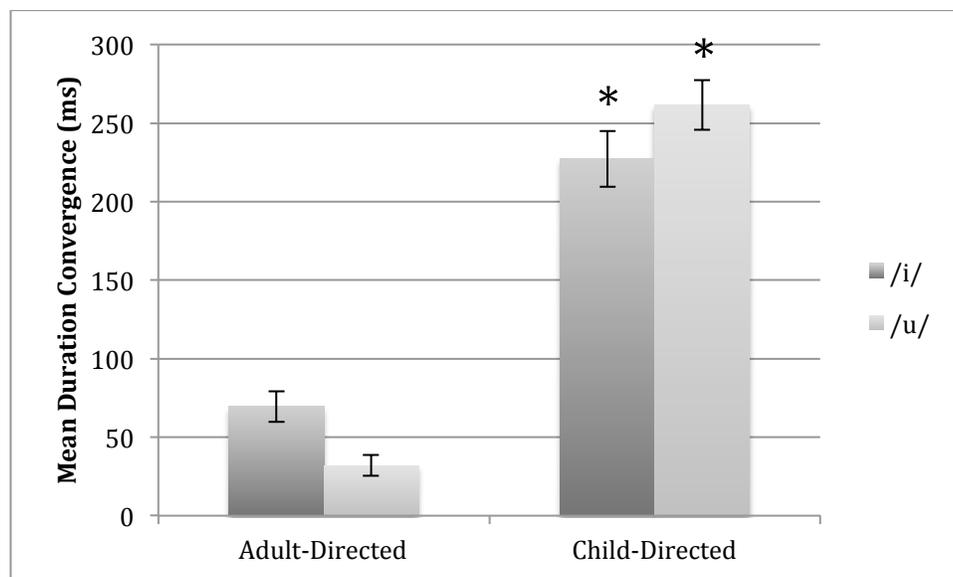


Figure 5.8. Comparison of pre-test and post-exposure phonetic distance in each vowel by register for adult subjects in the duration measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

f0:

In our analysis of convergence for f0, we saw that in all conditions subjects showed a significant difference between the two repetitions for f0 across all data subsets, converging in all conditions except for female participants and for the adult-directed register. Looking at further subsets, we also saw that female subjects diverged in f0 for both vowels in the adult-directed register, but not the child-directed register.

		Overall	Register		Modality		Gender	
			Adult-Directed	Child-Directed	Auditory	Audiovisual	Males	Females
f0	/i/	✓	D	✓	✓	✓	✓	D
	/u/	✓	D	✓	✓	✓	✓	D

Table 5.9. Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for f0. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for fundamental frequency is presented in Table 5.10. Table 5.10 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction.

<i>Adult Pretest f0</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Pseudo-R²</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	62.03	21.13	2.94	0.004*	17.44	5	0.004	0.0000
Modality	-19.41	20.85	-0.93	0.354	2.77	5	0.736	0.0001
Gender	40.13	26.92	1.49	0.139	18.35	5	0.003	0.0000
Vowel	-1.83	4.46	-0.41	0.683	2.27	5	0.810	0.0015
AQ	0.29	1.03	0.28	0.777	121.72	5	<0.001	0.0067
RegisterXModality	-1.32	12.71	-0.10	0.917	0.00	1	1.000	0.0000
RegisterXGender	-21.57	13.33	-1.62	0.108	3.03	1	0.082	-0.0001
ModalityXGender	1.18	13.32	0.09	0.930	0.00	1	1.000	0.0000
RegisterXVowel	1.66	2.50	0.66	0.510	0.54	1	0.464	0.0004
ModalityXVowel	0.82	2.53	0.32	0.748	0.18	1	0.676	0.0001
GenderXVowel	1.79	2.62	0.68	0.496	0.57	1	0.452	0.0004
RegisterXAQ	-2.40	1.19	-2.01	0.047*	4.62	1	0.032	-0.0001
ModalityXAQ	1.60	1.18	1.35	0.179	2.12	1	0.146	0.0000
GenderXAQ	-0.38	1.53	-0.25	0.805	0.02	1	0.901	0.0000
VowelXAQ	-0.08	0.24	-0.33	0.743	0.18	1	0.672	0.0001

Table 5.10. Summary of results from the mixed effects model for f0 for adult participants in pretest tokens compared to post-exposure production tokens.

Overall, there was a significant main effect of Register, and a significant interaction of Register X AQ. Surprisingly, considering the results for overall f0 convergence, we did not see a significant interaction of gender and register. Although we know from the analysis of convergence that only male adults converged to f0 in both registers and female adults converged in the child-directed register, but diverged in the adult-directed register, in this model we simply see that adults demonstrated significantly more implicit imitation of f0 in the child-directed speech register (Figure 5.9). Post-hoc tests confirm that this register difference was significant ($p = 0.0002$). Like for duration, it is not surprising to find f0 convergence for the child-directed

register, because child-directed speech shows a very different f0 from adult-directed speech, and there is a greater span of values upon which subjects could converge.

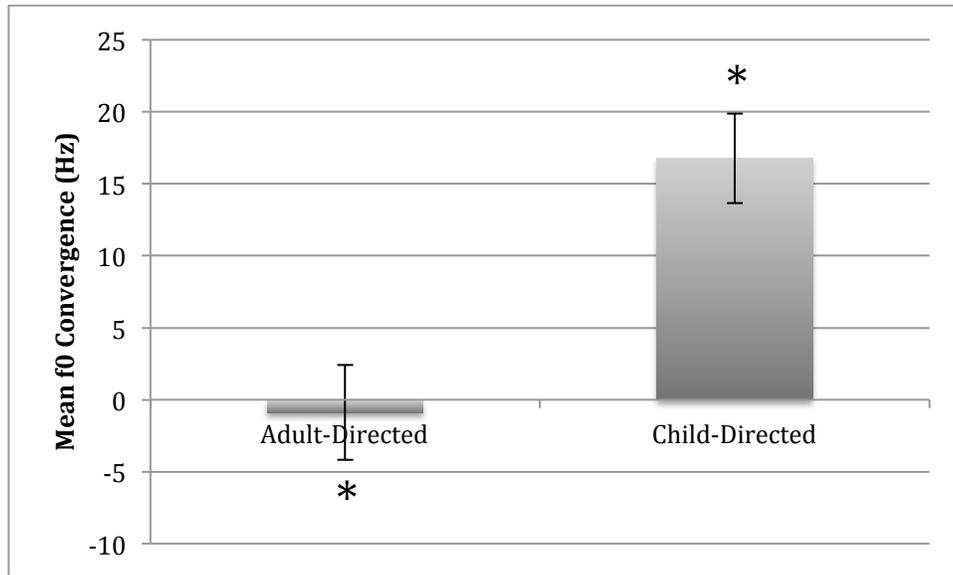


Figure 5.9. Comparison of pre-test and post-exposure convergence in each register for adult subjects in the F0 measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Finally, the last result for f0 showed that overall, adults displayed different behavior in each register as AQ score got higher (Figure 5.10): in the child-directed register, as AQ score got higher, imitation was lower ($p = 0.027$). In the adult-directed register, as AQ score got higher, imitation got *better* ($p = 0.049$).

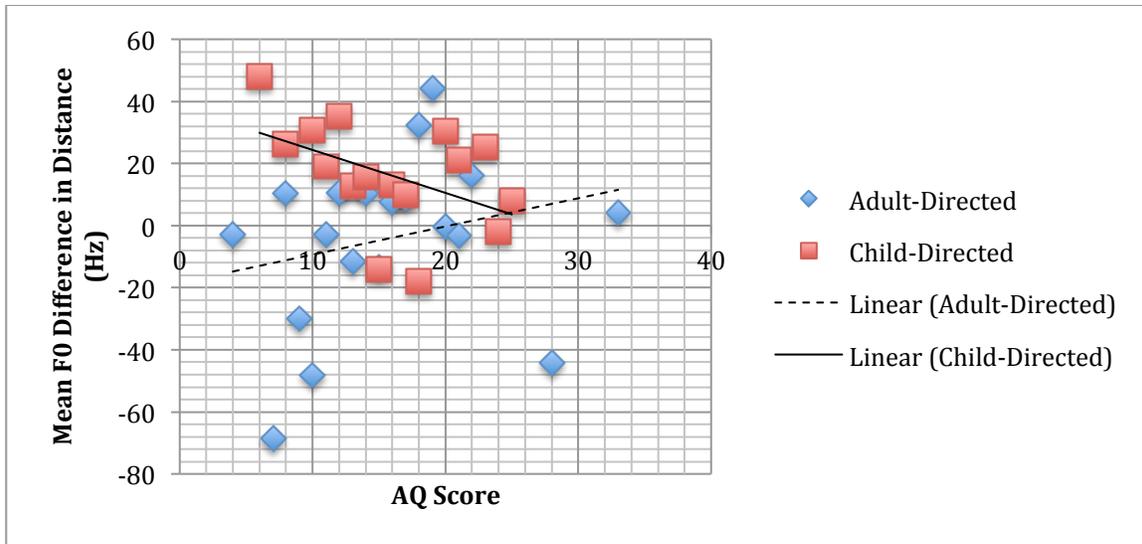


Figure 5.10. Scatterplot of each subjects' AQ and difference in pre-test and post-exposure f0 measurement by speech register. The lines are trend lines, and both slopes are significant.

Summary: Global Measures

In summary, the global measures showed a clear effect of speaking register, but inconsistent effects of modality. For both global measures, we saw an effect of register, favoring more implicit imitation when participants were exposed to child-directed speech. This was evident both in significant results of convergence, as well as in relative amounts of convergence within the measures. The increased imitation in the child-directed speaking register is likely due to the fact that duration and f0 are maximally different from participants' self-productions for child-directed speech and subjects had a greater span to converge on for this register.

With regards to modality, there was one significant effect related to modality, which was in the expected direction, favoring imitation in the audiovisual modality over the auditory

modality. This result was for overall convergence on duration for /i/; we saw a significant finding of convergence on /i/ in duration for the audiovisual but not the auditory modality.

Finally, there was one significant finding of a gender difference in the convergence analysis; female, but not male, subjects converged overall on the vowel /i/.

Aggregate Phonetic Measures

Euclidean Distance (F1 + F2 + F3):

In our analysis of overall convergence, recall that we saw that subjects in all conditions converged on the Euclidean Distance (F1 + F2 + F3) measure in the experiment. Further subset analyses revealed that this was true for all subset conditions.

The final output for the mixed effects model including all factors and interactions for the Euclidean Distance (F1 + F2 + F3) is presented in Table 5.11. Table 5.11 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. For the main effects, the likelihood ratio test represents taking out the main effect and all interactions. For the interactions, the likelihood ratio tests represent just taking out the single interaction.

<i>Euclidean F1+F2+F3</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Pseudo-R^2</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-1.05	0.46	-2.28	0.025*	31.71	5	<0.001	0.1136
Modality	-0.13	0.46	-0.29	0.771	5.69	5	0.338	0.0332
Gender	-0.13	0.58	-0.23	0.819	6.19	5	0.288	0.0263
Vowel	1.61	0.44	3.63	0.000*	117.85	5	<0.001	0.4403
AQ	-0.01	0.02	-0.45	0.656	63.83	5	<0.001	0.0917
RegisterXModality	0.16	0.27	0.59	0.553	0.37	1	0.545	0.0063
RegisterXGender	0.26	0.28	0.93	0.356	0.97	1	0.325	0.0043
ModalityXGender	0.34	0.28	1.21	0.229	1.68	1	0.195	0.0019
RegisterXVowel	0.95	0.25	3.81	0.000*	14.73	1	<0.001	0.0964
ModalityXVowel	0.35	0.25	1.41	0.161	2.18	1	0.140	0.0148
GenderXVowel	-0.33	0.26	-1.28	0.203	1.81	1	0.178	0.0110
RegisterX AQ	0.06	0.03	2.26	0.026*	5.81	1	0.016	-0.0116
ModalityX AQ	-0.01	0.02	-0.25	0.805	0.02	1	0.901	0.0075
GenderX AQ	-0.01	0.03	-0.39	0.697	0.12	1	0.724	0.0071
VowelX AQ	-0.04	0.02	-1.89	0.062	3.82	1	0.051	0.0318

Table 5.11. Summary of results from the mixed effects model for Euclidean Distance (F1 + F2 + F3) for adult participants in pretest tokens compared to post-exposure production tokens.

Overall, there were significant main effects of Register, Vowel, and significant interactions of Register X Vowel and Register X AQ. Results for the Register X Vowel interaction are shown in Figure 5.11. Since subjects converged on both registers and both vowels, then the significant difference was in the amount of convergence in these conditions; subjects converged more in the child-directed register than in the adult-directed register, and they converged more on the vowel /u/ than the vowel /i/. Post-hoc tests revealed that all pairwise Register X Vowel differences were significant ($p < 0.001$), except the convergence on the child-directed and adult-directed register for the vowel /i/ ($p = 0.999$).

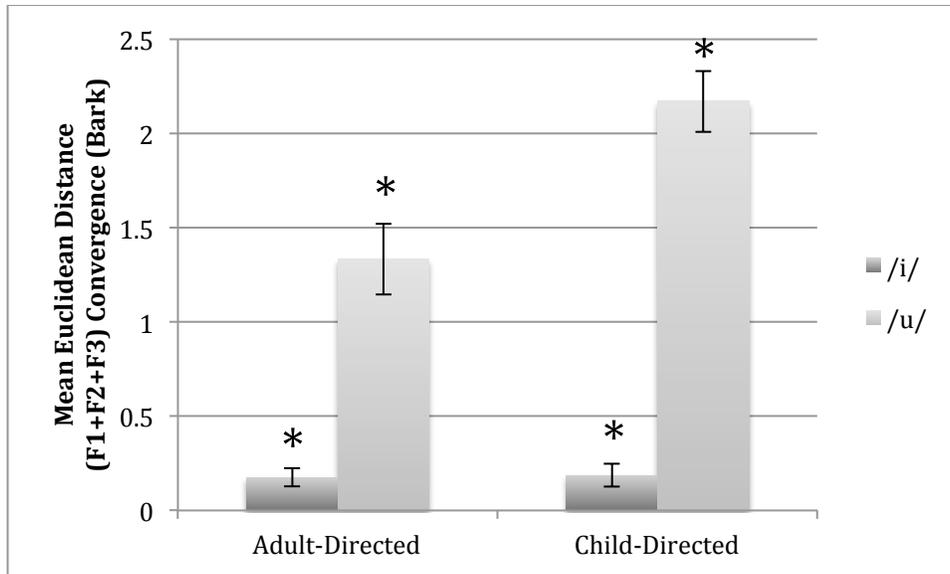


Figure 5.11. Comparison of pre-test and post-exposure phonetic distance in each vowel by register for adult subjects in the Euclidean Distance (F1+F2+F3) measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

The interaction of Register X AQ also showed a significant difference in imitation; post-hoc tests confirm that as AQ score got higher, imitation decreased, but only in the adult-directed register (slope = -0.042; $p = 0.0374$). The child-directed register showed no significant changes relating to AQ (slope = 0.018; $p = 0.58$). See Figure 5.12 for this plot.

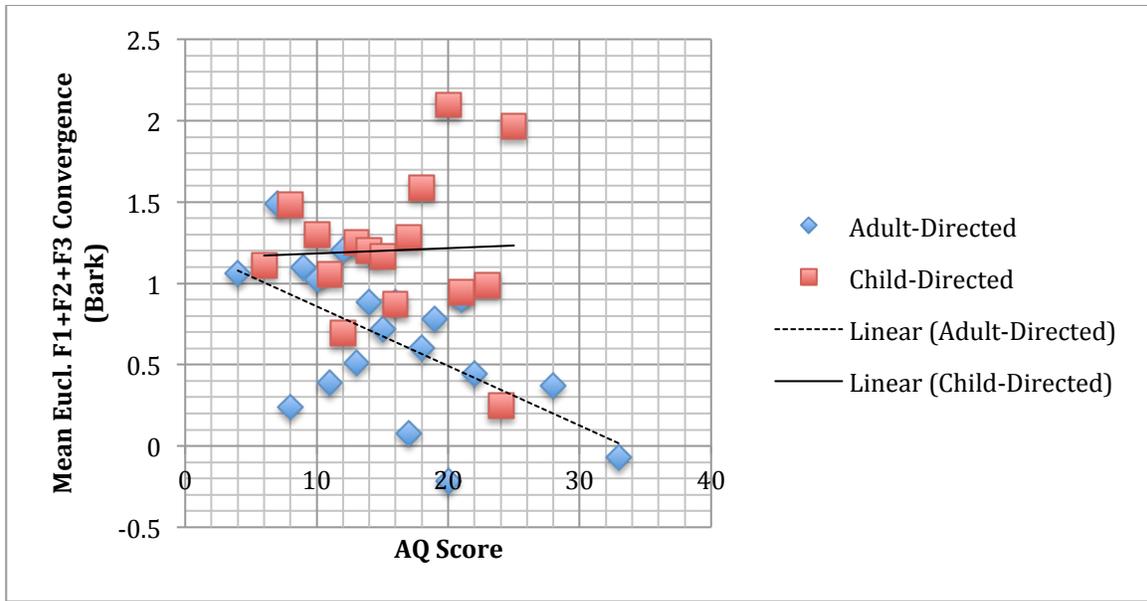


Figure 5.12. Scatterplot of each subjects AQ and the difference in pre-test and post-exposure Euclidean Distance (F1 + F2 + F3) measurement for each register. The lines are trend lines, and only the slope for the adult-directed register is significant.

Euclidean Distance (F1 + F2):

In our analysis of overall convergence, recall that we saw that subjects in all conditions converged on the Euclidean Distance (F1 + F2) measure in the experiment. Further subset analyses revealed that this was true for all subset conditions.

The final output for the mixed effects model including all factors and interactions for the Euclidean Distance (F1 + F2) measure is presented in Table 5.12. Table 5.12 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction.

<i>Euclidean Distance F1+F2</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Lambda)$-test</i>			<i>Pseudo-R^2</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-1.15	0.45	-2.57	0.012*	42.64	5	<0.001	0.1378
Modality	-0.06	0.44	-0.14	0.888	8.51	5	0.130	0.0441
Gender	-0.05	0.56	-0.10	0.924	6.21	5	0.287	0.0288
Vowel	1.52	0.43	3.50	0.001*	119.46	5	<0.001	0.4412
AQ	-0.01	0.02	-0.35	0.730	73.34	5	<0.001	0.1225
RegisterXModality	0.20	0.26	0.76	0.451	0.63	1	0.428	0.0054
RegisterXGender	0.19	0.27	0.70	0.488	0.52	1	0.469	0.0058
ModalityXGender	0.30	0.27	1.09	0.277	1.37	1	0.242	0.0030
RegisterXVowel	1.08	0.24	4.43	0.000*	19.92	1	<0.001	0.1020
ModalityXVowel	0.46	0.25	1.85	0.068	3.65	1	0.056	0.0299
GenderXVowel	-0.34	0.26	-1.35	0.181	2.00	1	0.157	0.0128
RegisterXAQ	0.06	0.02	2.66	0.009*	7.98	1	0.005	-0.0183
ModalityXAQ	-0.01	0.02	-0.48	0.630	0.22	1	0.637	0.0068
GenderXAQ	-0.01	0.03	-0.47	0.637	0.21	1	0.645	0.0068
VowelXAQ	-0.05	0.02	-2.23	0.028*	5.23	1	0.022	0.0464

Table 5.12. Summary of results from the mixed effects model for Euclidean Distance (F1 + F2) for adult participants in pretest tokens compared to post-exposure production tokens.

Overall, there are significant effects of Register and Vowel, and significant interactions of Register X Vowel, Register X AQ, and Vowel X AQ. Like for the first Euclidean Distance measure, the Euclidean Distance (F1 + F2) measure showed significantly more convergence for the child-directed speech, and for /u/ compared to /i/. Post-hoc tests again revealed that all pairwise Register X Vowel differences were significant ($p < 0.001$), except the difference between the child-directed and adult-directed register for /i/ ($p = 0.999$). This is the exact same pattern as that for Euclidean Distance (F1 + F2 + F3) seen above, and so we will not present a graph.

Like for the overall Euclidean Distance (F1 + F2 + F3) measure, for Euclidean Distance (F1 + F2) as well, the interaction of Register X AQ was significant; post-hoc tests once again confirmed that as AQ score got higher, imitation decreased, but only in the adult-directed register (slope = -0.046; $p = 0.018$). Once again, we do not plot this result because it is the exact same as for the previous measure, including F3. Unlike the overall Euclidean Distance (F1 + F2 + F3) measure, the Euclidean Distance (F1 + F2) measure also showed a significant interaction of Vowel X AQ (as AQ got higher, imitation of /u/ was worse) but post-hoc testing revealed that this was not significant.

Summary: Aggregate Phonetic Measures

Recall that we saw overall imitation on all data subsets for both vowels, but with the complex mixed effects models, we saw differences in the amount of imitation. Overall, the analysis of the aggregate phonetic measures showed an effect of register and some effects of the vowel being imitated. We saw greater imitation for /u/ than for /i/, and for /u/ only, we saw a register effect, favoring more imitation in the child-directed register. Additionally, we saw an influence of AQ scores on imitation. Subjects with higher AQ scores were less likely to implicitly imitate in the adult-directed register.

Individual Phonetic Measures

F1:

In our analysis of whether subjects showed convergence, recall that we did not see overall convergence on F1 for either vowel type. However, further subset analysis revealed that females converged on both vowels, and that there was overall convergence in the child-directed register for /u/. In order to see if further subsets would be more informative, we analyzed the data from males and females separately on this variable.

Both males and females showed convergence on F1 to both vowels, but only in particular sub-conditions. Females showed convergence on F1 for both vowels in the auditory modality, to /i/ in the child-directed register, and to /u/ in the adult-directed register. Males showed convergence to both vowels only in the adult-directed register, but did not show any modality differences. A summary of these results is shown below in Table 5.13.

		Overall	Overall			Child-Directed			Adult-Directed			Auditory			Audiovisual		
			M	F	Overall	M	F	Overall	M	F	Overall	M	F	Overall	M	F	
F1	/i/	✗	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗	✗	✓	✗	✗	✗
	/u/	✓	✓	✓	✓	✗	✗	✓	✓	✓	✗	✗	✓	✗	✗	✗	✗

Table 5.13. Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for F1. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for the first formant is presented in Table 5.14. Table 5.14 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction.

<i>Adult Pretest F1</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Pseudo-R^2</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	0.12	0.25	0.49	0.624	3.16	5	0.675	0.0129
Modality	-0.55	0.25	-2.25	0.026*	10.26	5	0.068	-0.0269
Gender	-0.14	0.32	-0.46	0.648	4.92	5	0.426	0.0179
Vowel	-0.17	0.16	-1.09	0.280	6.45	5	0.265	0.0316
AQ	-0.02	0.01	-1.32	0.190	6.41	5	0.268	-0.0155
RegisterXModality	-0.13	0.15	-0.88	0.380	0.87	1	0.350	-0.0010
RegisterXGender	0.08	0.15	0.54	0.594	0.29	1	0.593	0.0011
ModalityXGender	0.09	0.15	0.56	0.580	0.31	1	0.576	0.0010
RegisterXVowel	0.11	0.09	1.27	0.206	1.79	1	0.181	0.0097
ModalityXVowel	0.07	0.09	0.77	0.442	0.71	1	0.401	0.0023
GenderXVowel	-0.15	0.09	-1.65	0.103	2.93	1	0.087	0.0175
RegisterXAQ	-0.01	0.01	-0.55	0.585	0.30	1	0.582	0.0010
ModalityXAQ	0.03	0.01	2.13	0.036*	5.18	1	0.023	-0.0163
GenderXAQ	0.00	0.02	0.16	0.873	0.00	1	1.000	0.0022
VowelXAQ	0.01	0.01	1.08	0.281	1.32	1	0.250	0.0065

Table 5.14. Summary of results from the mixed effects model for F1 for adult participants in pretest tokens compared to post-exposure production tokens.

Overall, there was a significant main effect of Modality, and a significant interaction of Modality X AQ. The main effect of Modality was only marginally significant in the likelihood ratio tests. Adults showed significantly more imitation of F1 in the auditory modality compared to the audiovisual modality (Figure 5.13). This corresponds well with the finding of significant convergence in the auditory, but not audiovisual, modality, and it is likely driven by the female participants who showed significant convergence in this condition. However, post-hoc tests reveal that the pairwise comparison of the two modalities is barely significant ($p = 0.05$).

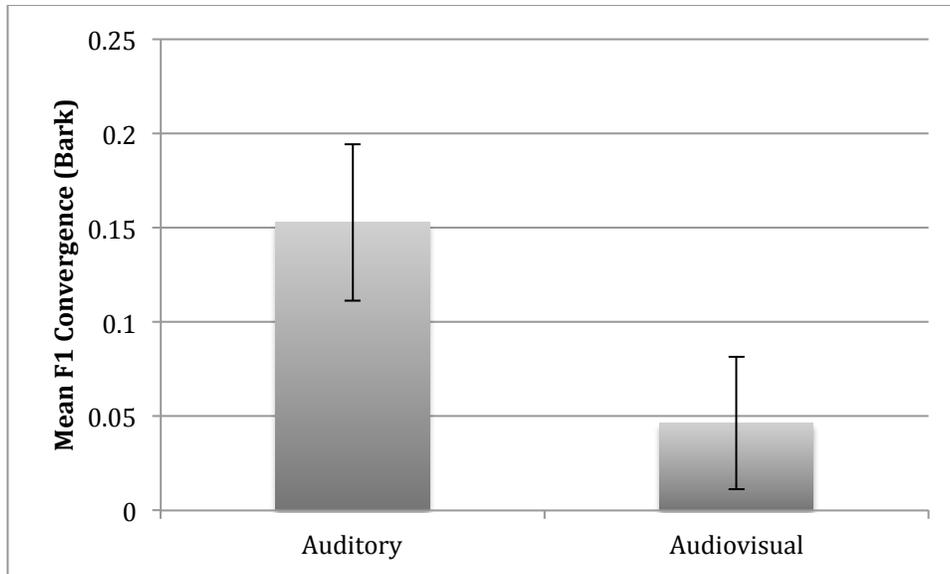


Figure 5.13. Comparison of pre-test and post-exposure convergence in each modality for adult subjects in the F1 measurement.

With regards to the interaction between Modality X AQ score (shown in Figure 5.14), post-hoc testing revealed that in the auditory modality, there were no significant changes based on AQ score, but in the audiovisual modality, as AQ score got higher, participants converged more on F1 with the talker’s articulations (auditory modality, slope = -0.015; $p = 0.11$; audiovisual modality, slope = 0.014; $p = 0.038$).

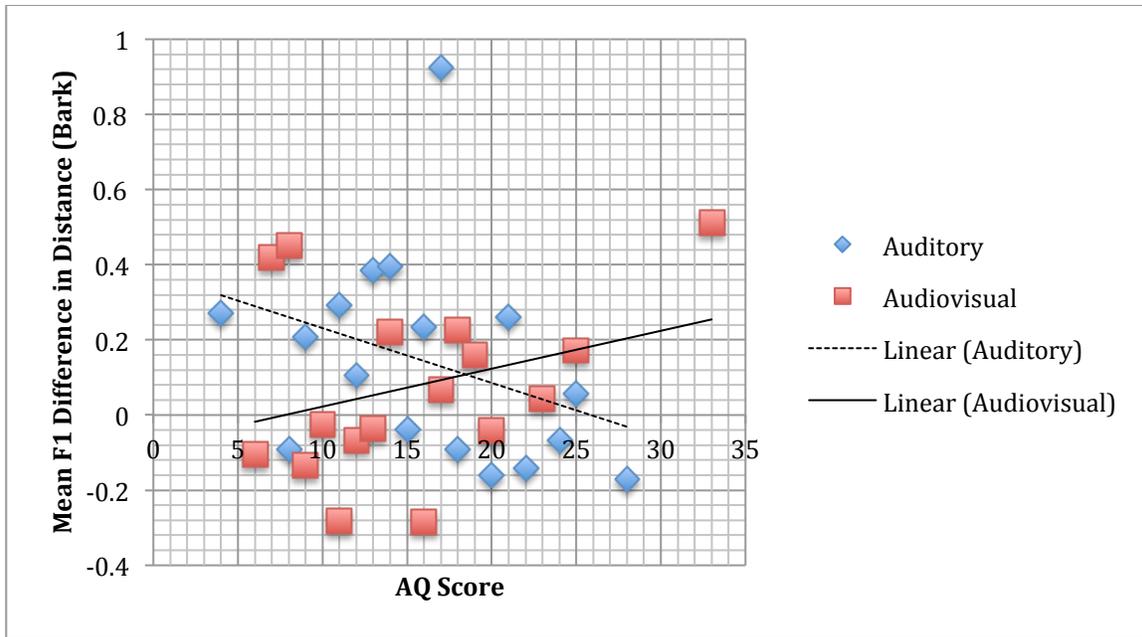


Figure 5.14. Scatterplot of each subjects AQ and the difference in pre-test and post-exposure F1 measurement for each modality. The lines are trend lines, and only the slope for the audiovisual modality is significant.

F2:

Recall that we saw convergence for both vowels overall on the F2 measure. Looking at each modality, register, and gender individually, we saw convergence for all subcategories of F2, except for /i/ for the male participants. Further analysis revealed that this lack of /i/ imitation in males was consistent across registers, but was only evident in the auditory modality; in the audiovisual modality, males did show significant F2 imitation for /i/ ($\beta = -0.19, t() = -2.35, p = 0.027$).

		Overall	Overall			Child-Directed			Adult-Directed			Auditory			Visual		
			M	F	Overall	M	F	Overall	M	F	Overall	M	F	Overall	M	F	
F2	/i/	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	
	/u/	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

Table 5.15. Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for F2. The full statistical output is shown in the appendix.

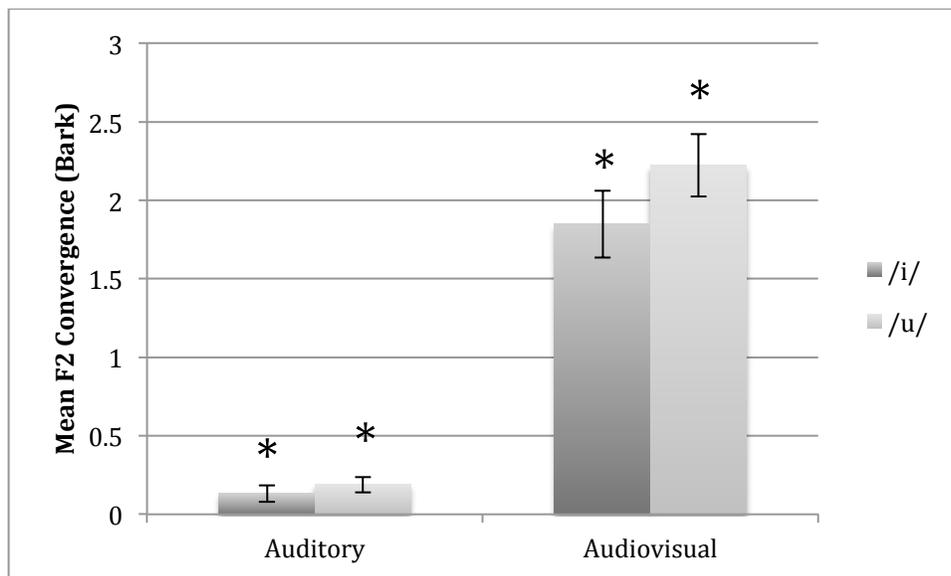


Figure 5.15. Comparison of pre-test and post-exposure convergence in each vowel by modality for adult subjects in the F2 measurement. Significance stars represent significant convergence or divergence (difference from zero) with a *p*-value of 0.05 or smaller.

The final output for the mixed effects model including all factors and interactions for the second formant is presented in Table 5.16. Table 5.16 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction.

<i>Adult Pretest F2</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Pseudo-R^2</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-1.31	0.48	-2.71	0.008*	30.27	5	<0.001	0.1025
Modality	0.00	0.48	0.01	0.995	8.90	5	0.113	0.0305
Gender	0.01	0.61	0.01	0.991	6.84	5	0.233	0.0233
Vowel	2.23	0.48	4.60	0.000*	136.17	5	<0.001	0.4629
AQ	-0.01	0.03	-0.22	0.826	69.11	5	<0.001	0.1146
RegisterXModality	0.30	0.28	1.06	0.291	1.28	1	0.258	0.0043
RegisterXGender	0.21	0.29	0.70	0.484	0.56	1	0.453	0.0019
ModalityXGender	0.37	0.29	1.28	0.204	1.85	1	0.174	0.0062
RegisterXVowel	0.89	0.27	3.26	0.002*	11.61	1	0.001	0.0400
ModalityXVowel	0.37	0.28	1.34	0.183	2.04	1	0.153	0.0069
GenderXVowel	-0.35	0.29	-1.21	0.229	1.66	1	0.197	0.0056
RegisterX AQ	0.07	0.03	2.77	0.007*	8.54	1	0.003	0.0181
ModalityX AQ	-0.02	0.03	-0.58	0.561	0.39	1	0.533	0.0013
GenderX AQ	-0.02	0.03	-0.71	0.481	0.57	1	0.450	0.0019
VowelX AQ	-0.06	0.03	-2.42	0.017*	6.51	1	0.011	0.1025

Table 5.16. Summary of results from the mixed effects model for F2 for adult participants in pretest tokens compared to post-exposure production tokens.

Overall, there were significant main effects of Register and Vowel, and significant interactions of Register X Vowel, Register X AQ, and Vowel X AQ. Overall, subjects converged more for the child-directed register than the adult-directed register, and for /u/ than /i/. Post-hoc testing revealed that the main effects and all pairwise Register X Vowel interactions were significant ($p < 0.01$) except the difference in imitation of F2 for /i/ in the two registers ($p = 0.998$). The results from this interaction are presented in Figure 5.16.

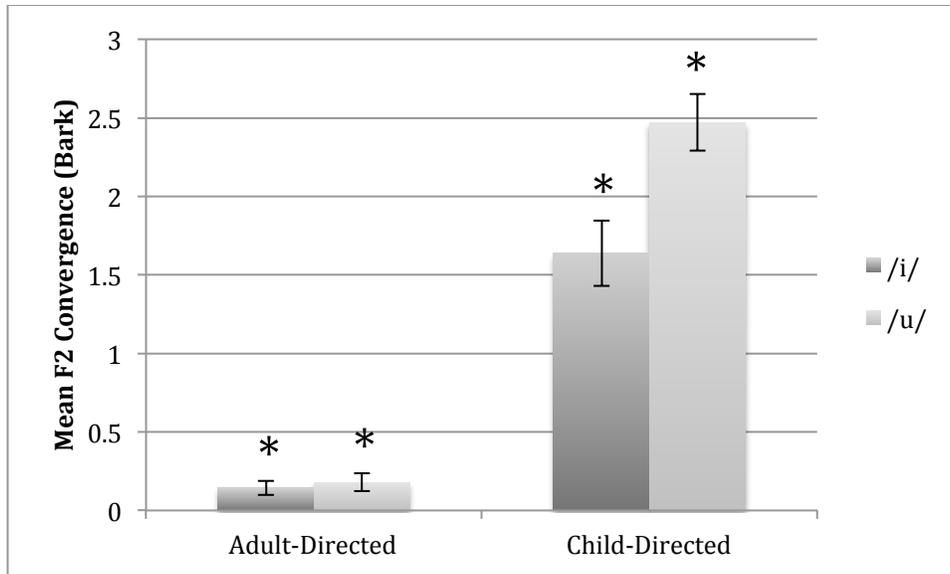


Figure 5.16. Comparison of pre-test and post-exposure phonetic distance in each vowel by register for adult subjects in the F2 measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Additionally, register also interacted with AQ score (Figure 5.17). As AQ score got higher, participants were worse at imitation, but only in the adult-directed register (slope = -0.052; $p = 0.028$); the child-directed register did not show significant influence of AQ score on imitation (slope = 0.028; $p = 0.55$). Finally, participants showed different imitation patterns for each vowel based on AQ score (Figure 5.18): while /i/ was imitated to the same extent by subject with different AQ scores (slope = 0.011; $p = 0.14$), subjects with higher AQ scores were less likely to imitate F2 of /u/ (slope = -0.057; $p = 0.044$).

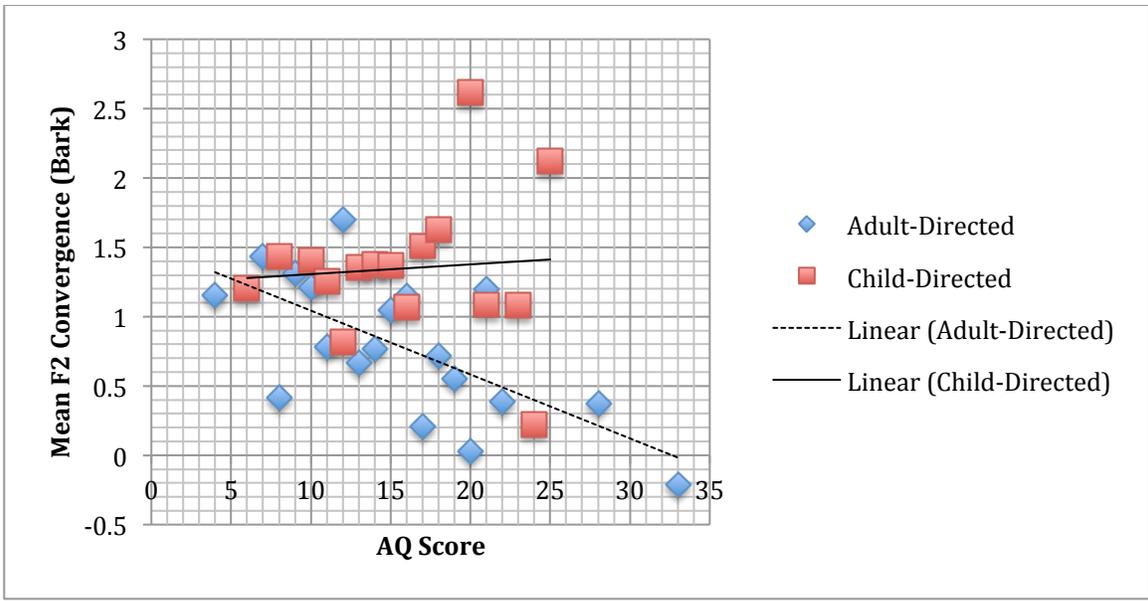


Figure 5.17. Scatterplot of each subjects AQ and the difference in pre-test and post-exposure F2 measurement for each register. The lines are trend lines, and only the slope for the adult-directed register is significant.

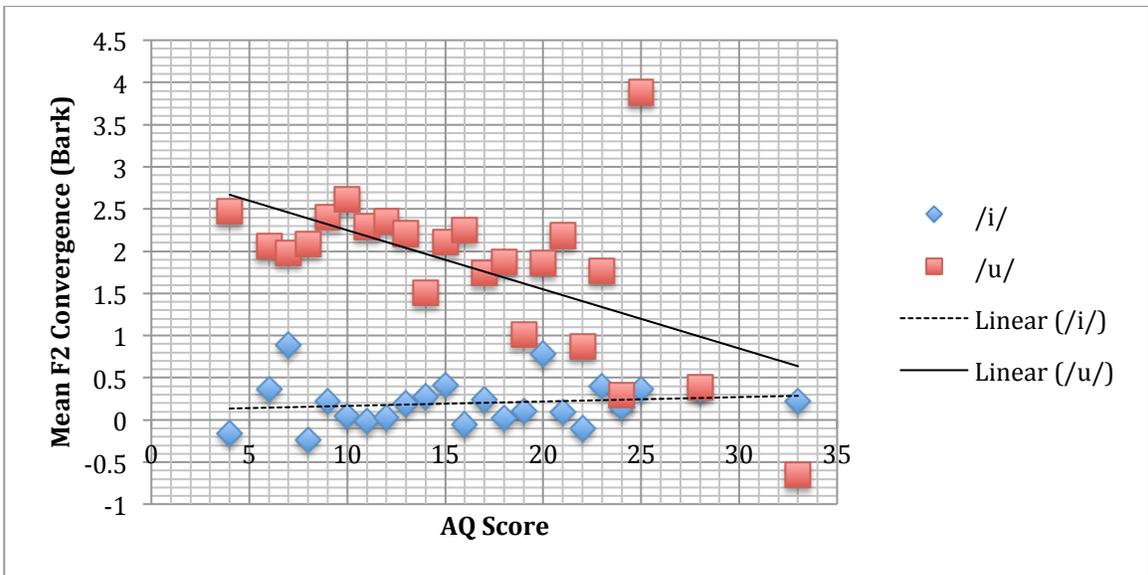


Figure 5.18. Scatterplot of each subjects AQ and the difference in pre-test and post-exposure F2 measurement for each vowel. The lines are trend lines, and only the slope for /u/ is significant.

F3:

Recall that subjects did not converge with the model talker on F3, so we will not look any further into this variable.

Summary: Individual Phonetic Measures

Looking at the individual phonetic measures, we saw effects of both register and modality, but they were not as clear and systematic as for the global measures. For register, we saw some register effects in both F1 and F2. For F1, we saw a difference in significant convergence; we only found a significant finding of convergence for child-directed /u/ (neither child-directed /i/ nor adult-directed /i/ or /u/ was significant). For F2, the picture was much clearer with regards to speaking register; while we found convergence in both registers, there was significantly more convergence for the child-directed register than the adult-directed register. Also, we saw the same pattern for the interaction of register and AQ score with F2 that we saw for the aggregate phonetic measures (higher AQ score resulted in worse imitation in the adult-directed register). Based on these results, it seems that the speaking register effect on F2 is largely driving the effects for the aggregate phonetic measures.

In addition to register effects, we did also see one effect of modality. For F1 we saw that there was significantly more convergence in the auditory modality than the audiovisual modality. Recall that F1 is supposed to be especially salient in the auditory modality. One explanation of the lesser imitation of F1 in the audiovisual modality is that the other formants are more salient

and that information is getting in the way of clear perception and imitation of F1. Further testing would be needed to analyze this hypothesis. One other modality effect was found, also for F1. This result was that there was an interaction of modality and AQ score, and there was less imitation of F1 in the auditory modality as AQ score got higher. None of the hypothesis of the current experiment can appropriately account for this finding of only lesser convergence at higher AQ scores in the auditory modality, although we predicted less imitation overall.

We found a main effect of vowel for F2, and this finding is in accordance with our hypothesis that F2 of the vowel /u/ would be better imitated due to its more variable pronunciation. Indeed we saw significantly more F2 imitation of /u/ than /i/, and we also saw a significant interaction of vowel with AQ score; this /u/ imitation was lesser as AQ score got higher. For whatever reason (perhaps due to this same social nature of F2 in /u/) participants with higher AQ scores were less likely to converge on this dimension.

Finally, we found some evidence that women imitate more than men. While there were no differences in the amount of imitation, we saw differences in whether there was significant convergence for F1 and F2 by gender. For F1, we only found convergence to be significant overall for female participants. For F2, male participants did not significantly converge, but this was only for /i/.

Interim summary: English-like vowels, pre-test vs. post-exposure convergence

Direction of Convergence:

The sections above have given information about whether there was significant convergence or divergence exhibited by subjects in the pre-test/post-exposure comparison. However, there is information missing when just asking “is there convergence?”, namely, how subjects are actually changing their productions to converge with the model talker. In this section, we show graphs that can indicate the *direction* of convergence.

The graphs below summarize the results for production in each of our experimental measures according to register and modality, separated by vowel. In each of these graphs, rather than indicating the *amount* of convergence (as in the graphs previously shown), we plot mean values in the pretest and post-exposure productions.

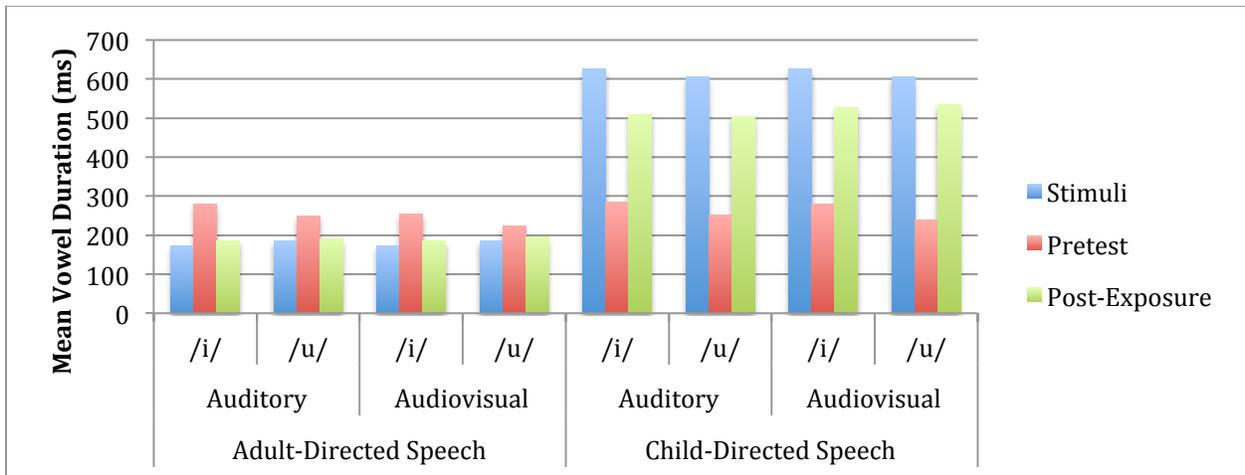


Figure 5.19. Mean duration of vowels produced by subjects in in the pretest and post-exposure phases, compared with the mean duration of the stimuli, separated by vowel, modality, and register.

In the duration graph above, we can see that in the adult-directed speech condition, subjects shortened their vowel duration to converge with the model talker. This is in contrast to what we see for the child-directed speech condition, in which they had to lengthen their vowel duration to converge with the model talker.

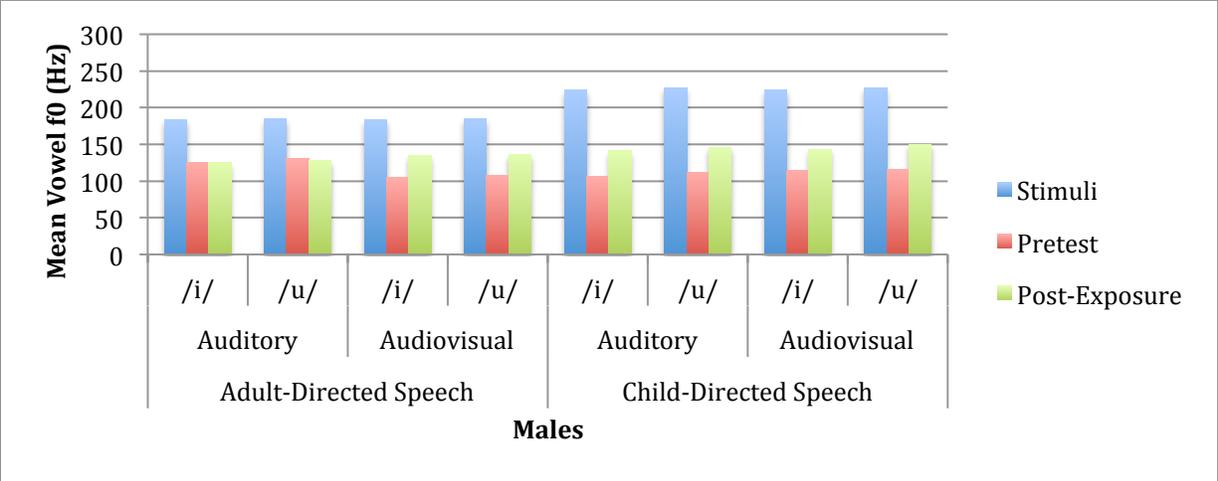


Figure 5.20. Mean f0 of vowels produced by male subjects in in the pretest and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel, modality, and register.

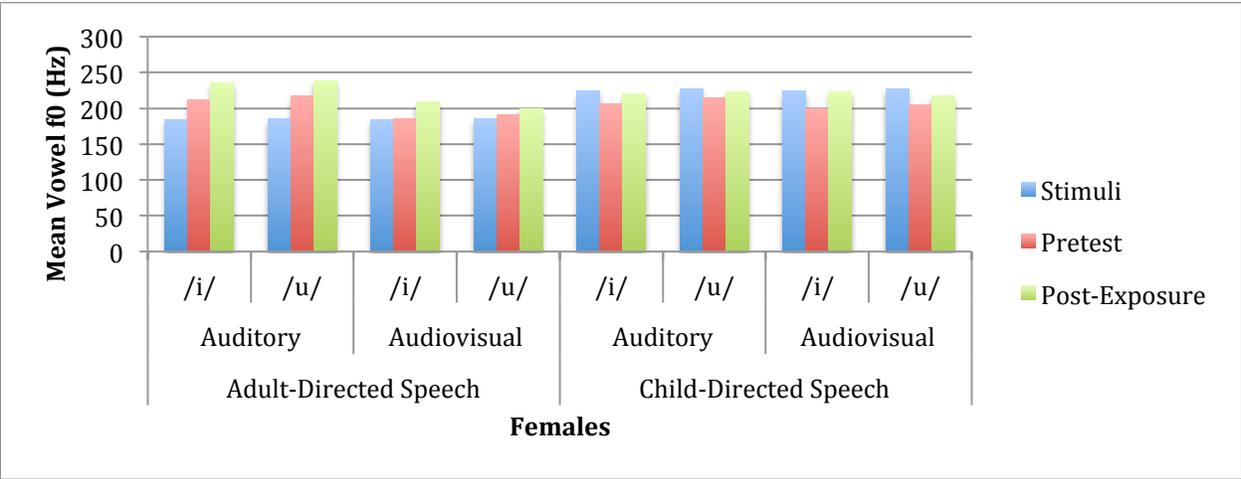
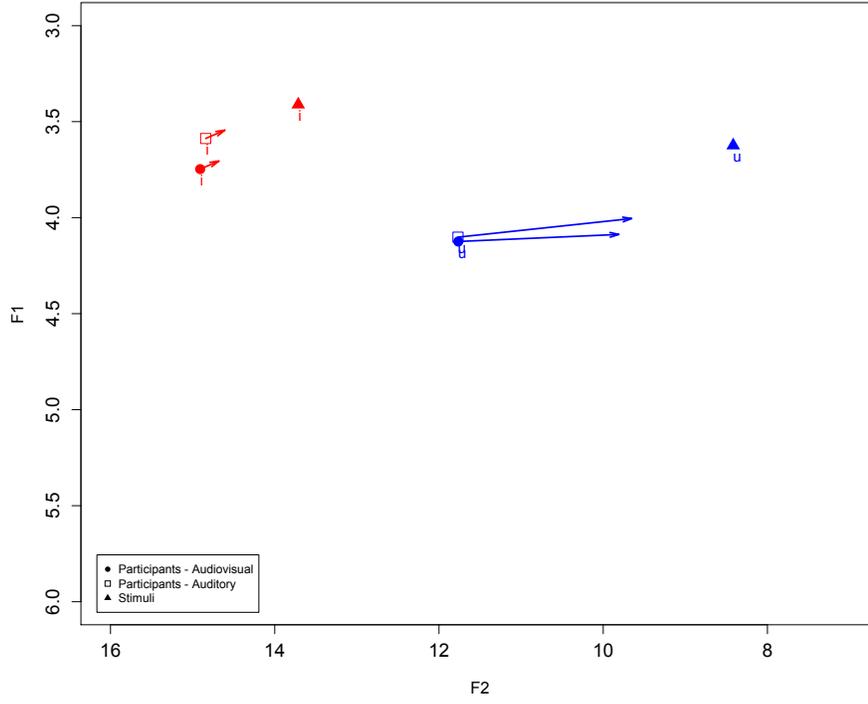


Figure 5.21. Mean f0 of vowels produced by female subjects in in the pretest and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel, modality, and register.

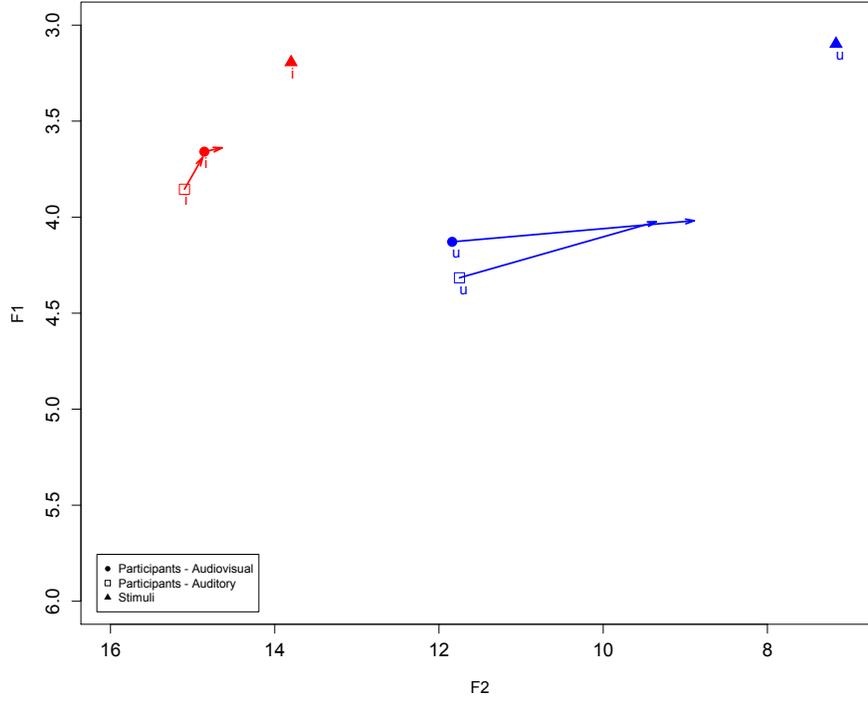
The graphs above show the direction of convergence to f0 for males and females separately, since f0 of male and female talkers often differs. We see that for the male talkers, the subjects raised their f0 to converge to the model talker in all conditions. For the female talkers, however, we notice that for the child-directed speech condition, they make their speech converge to the f0

of the model talker by raising it, but for the adult-directed speech condition, they also raise the f_0 of their speech, even though this makes their speech *less* like the model talker. An explanation for this would be that these talkers hear a raised f_0 in the speech of the male talker (the speaker's f_0 was very high for a male talker) in the adult-directed condition, and they in turn raise their f_0 . They still seem to be converging to the model talker, but they are normalizing the raised f_0 .

F1-F2 Imitation: Adult-Directed Speech



F1-F2 Imitation: Child-Directed Speech



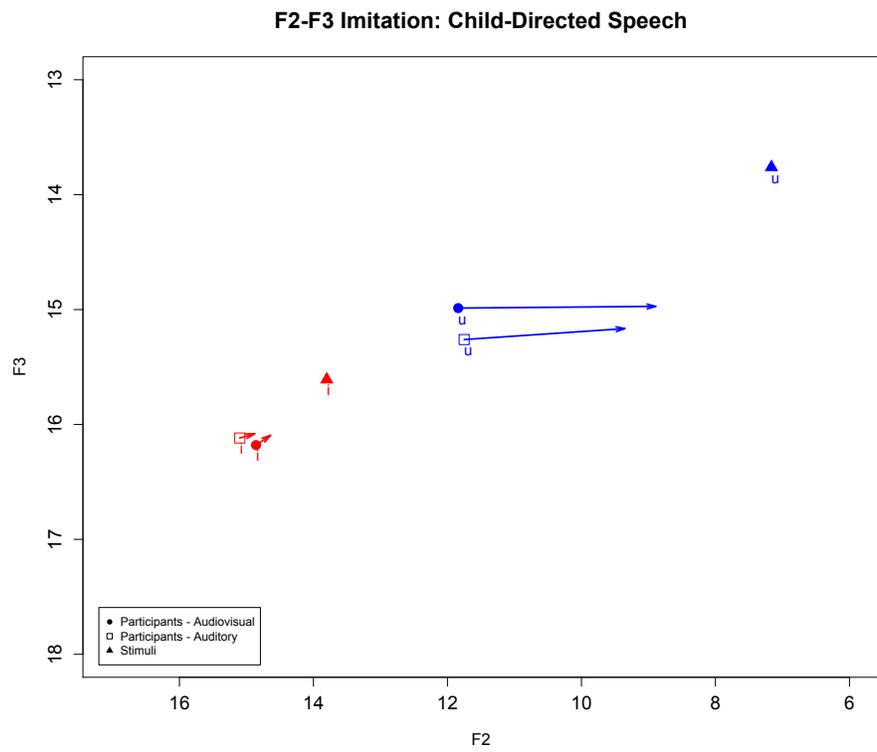
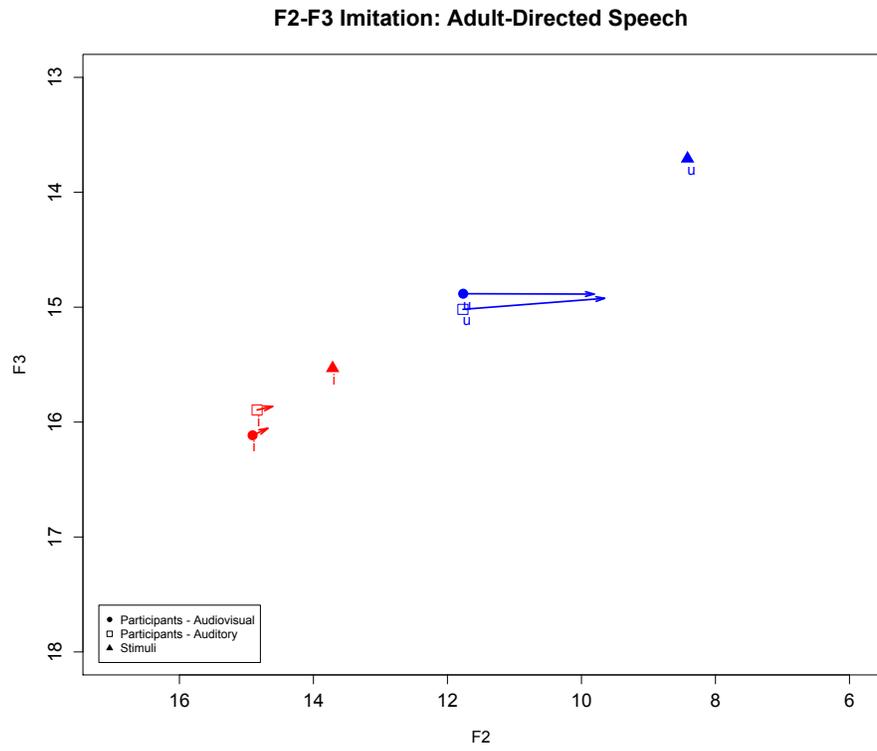


Figure 5.22. Formant plots of vowels produced by subjects in the pretest and post-exposure phases, compared with the mean formant values of the stimuli, separated by vowel, modality, and register.

Looking at the direction of convergence on the phonetic measures for the two vowels in the pre-test/post-exposure comparison, we see clear differences between the vowels. It appears that the majority of convergence for /u/ is on F2 (front/backness), which is confirmed in our data. The stimuli formant values for /u/ are much further back, and considerably raised, but subjects exhibit convergence almost exclusively on F2, ignoring the F1 and F3 differences. For /i/, we see that the changes in production are much more slight, but the vectors indicate convergence occurring on all of the phonetic measures, not limited to one measure as in /u/.

Register Effects:

For our overall measures, the subjects in this experiment converged between pre-test and post-exposure productions. Convergence for at least some data subsets was found on all of our acoustic measures except F3. For both the global measures, and the two aggregate phonetic measures, we saw clear and systematic effects of speaking register, favoring more implicit imitation in the child-directed register than in the adult-directed register (and even divergence in the adult-directed register for f0). Looking at the phonetic measures individually, this register difference was evident for F2 and indirectly evident for F1 (demonstrated by lack of significant convergence findings in the adult-directed register). Thus, register effects were shown in global measures, and in aggregate phonetic measures where they were mostly driven by F2. Overall convergence values for each measure by register are shown below in Figure 5.23.

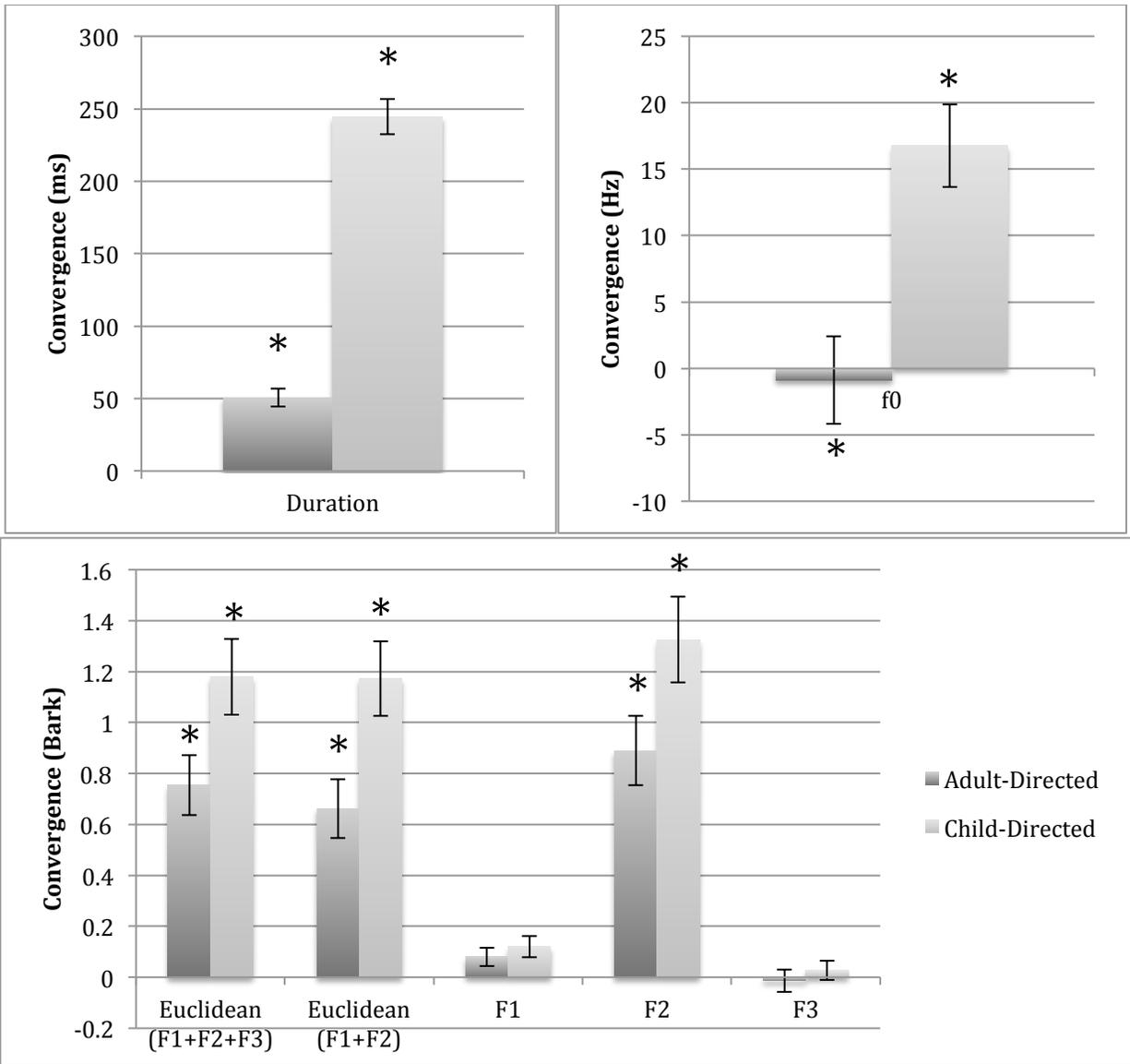


Figure 5.23. Results for convergence according to register for each measure in the pre-test/post-exposure comparison for adult participants. The darker bars represent results for the adult-directed register and the lighter bars represent results for the child-directed register. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Modality Effects:

Effects of modality (ignoring for the present time, the AQ effects) were less consistent. They were observed on global measures and individual phonetic measures, but not on aggregate phonetic measures. Starting with the global measures, duration convergence was evident in the audiovisual modality but not the auditory modality, but this was only the case for /i/. As for the individual phonetic measures, there was greater convergence for F1 in the auditory rather than audiovisual modality. This is consistent with the idea that F1 is best perceived in the auditory modality, however, this is specifically against an overall advantage of the visual cues, since this shows an advantage of the auditory cues. For F2, we saw the opposite, although expected) effect; an advantage for the audiovisual modality. We found significant convergence only in the audiovisual, and not auditory modality for /i/. Interestingly, this is very similar to the global measure of duration. For both of these measures, we saw significant convergence for /i/ only. Overall, in the pre-test/post-exposure comparison, the audiovisual modality favors convergence on duration and F2, whereas the auditory modality favors convergence on F1.

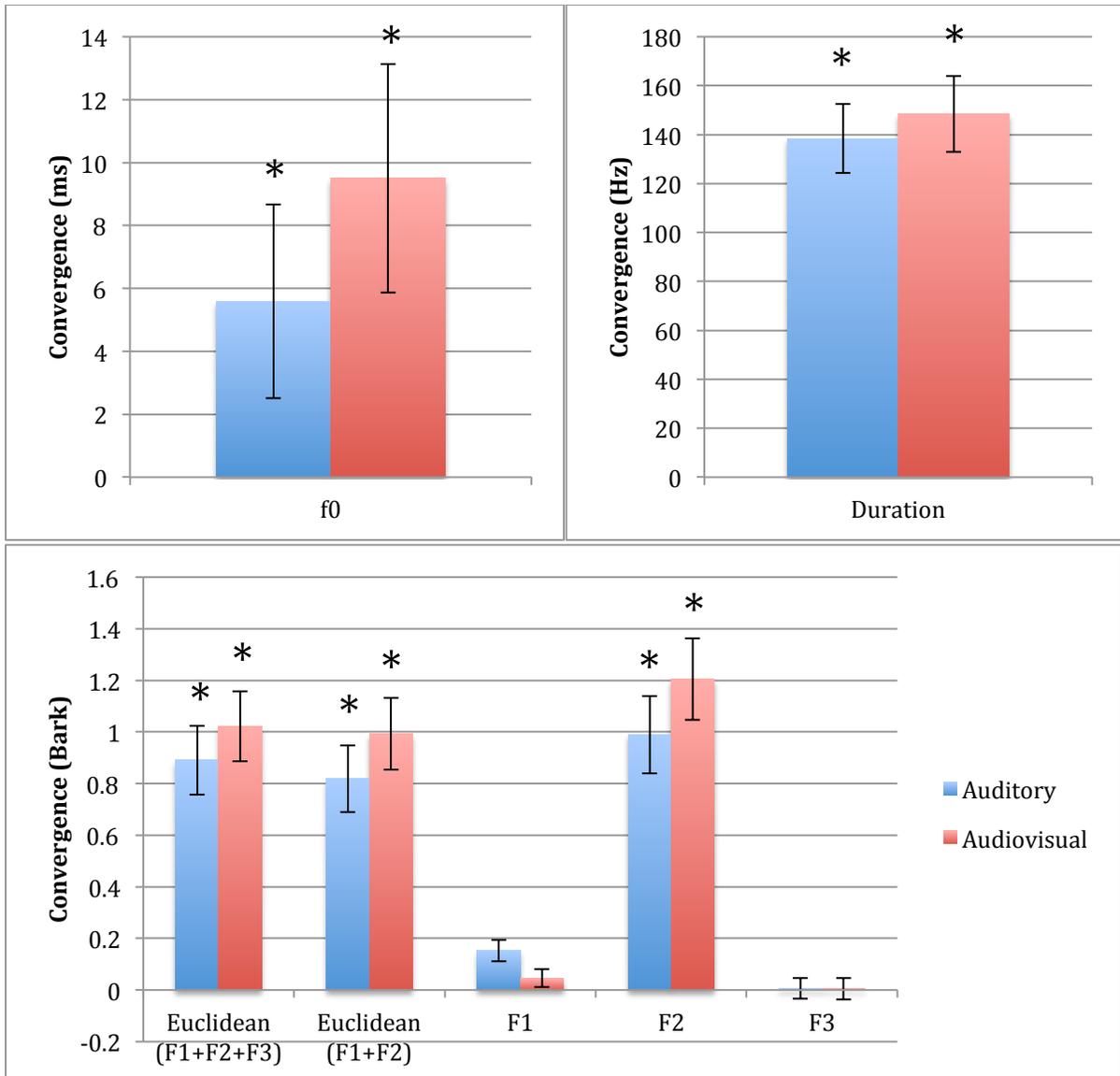


Figure 5.24. Results for convergence according to modality for each measure in the pre-test/post-exposure comparison for adult participants. The blue bars represent results for the auditory modality and the red bars represent results for the audiovisual modality. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Vowel Effects:

Additionally, because /u/ shows more variation within American English (recall, /u/-fronting from the section on imitation), we expected to see more convergence for /u/ on F2. This was confirmed, both in the complex mixed effect model, in which we saw more /u/ imitation than /i/ imitation, and in the overall convergence measure, in which we saw imitation for /u/ across all conditions, but not for /i/. This difference was exaggerated in child-directed speech.

AQ Effects:

The final point of interest for this section was in the results for AQ. We saw a few interesting results for AQ interacting with register, vowel, and modality. The AQ and register interactions were evident in the overall Euclidean Distance measures, and for the F2 measure. In all these cases, as predicted, there was overall less imitation with a higher AQ score, but only in the adult-directed register. Perhaps this is due to the social nature of child-directed speech; if people with higher AQ scores show autistic-like social functioning, then they may be less likely to imitate child-directed speech, which is a socially driven register. The interaction of AQ score and vowel also occurred in the F2 measure, with less F2 imitation at higher AQ scores, but only for /u/, regardless of modality. Finally, there was one additional interaction of the AQ measure with F1; here, the audiovisual modality allowed for better imitation of F1 for subjects with a higher AQ score. Because F1 is no more salient in the audiovisual modality compared to the auditory modality, imitation is expected to be worse at higher AQ scores, and subjects with higher AQ scores are less likely to use the visual modality, these findings are mysterious. Future studies are needed to evaluate the validity and consistency of this result. We should take note

that these AQ effects were only in phonetic measures (either individual or aggregate), and not in the global measures of duration and f0.

Did the nature of the exposure session affect convergence within the experiment?

In this section of the experiment, we will look at whether subjects converged to the model talker *within* the experiment. For this section, we will look at the subjects' productions in the pre-exposure and post-exposure phases, and compare the phonetic distance of the two productions with the productions of the model talker. For both analyses, we included two random factors in all models: Word and Subject. This allowed each subject to differ in his or her overall degree of convergence, and, since for this analysis we were comparing at the word-level, each word to have a unique pronunciation.

As in the pre-test/post-exposure comparison, we separated the analysis by variable, and for each variable, we report two difference analyses. In the previous section all statistical analyses were done with mixed effects models, using the lme4 package in R (R Development Core Team, 2008). These models have the advantage of including random effects in addition to fixed effects. In order to further interpret patterns, we subjected the data to a repeated measures analysis of variance, using Tukey's HSD test, also in R for all factors except AQ. For AQ, we computed analysis of covariance on data subsets to further clarify the results.

As for the pre-test/post-exposure comparison, each section will have two types of analysis. The first portion of the analysis looks at overall convergence between the two

experimental sessions. For this part of the analysis, the variable was the phonetic *difference* (for each phonetic measure) between the model talker and the participant, with a factor of Repetition (for this section, *pre-exposure* or post-exposure) to differentiate the two productions. We ran mixed effect models for each type of vowel separately (English-like or foreign vowels) with the fixed effect of Repetition and the random effects of Subject and Word.

For the second part of the analysis, we looked at how the experimental factors affected the convergence measure in the experiment. We model how each of 7 dependent variables (duration convergence, f0 convergence, F1 convergence, F2 convergence, F3 convergence, Euclidean distance with just F1 and F2, and Euclidean distance with F1, F2, and F3) vary according to the experimental design factors. The experimental design was a 2 (Gender: male or female) X 2 (Modality: audio or visual) X 2 (Register: child-directed or adult directed speech) X 2 (Vowel Type: English-like or foreign) X 2 (Carrier Type: CVC or CV) factorial design, with the additional variable of AQ score.¹⁸ The between-subject variables were Subject, Gender, Modality, Register, and AQ score, and the within-subject variables were Vowel Type, Carrier Type, and Word. We provide full models for each variable, including main effects and two-way interactions (resulting in fifteen possible fixed effects, and two random effects). We did not include the two-way interactions of AQ X Carrier Type and AQ X Vowel Type because there is no intuitive interpretation of the results of these interactions.

The number of variables for analysis makes statistical analysis very complicated. To simplify the model, we analyzed the effects of convergence separately for male and female

¹⁸ There were a few other factors not considered in the modeling: the place of articulation of the onset consonant, the specific carrier ('d_', 'g_k', etc.), the final consonant (/g/, /k/, none. Carrier type, CV or CVC, was used instead.), the voicing of the consonants (although this should actually affect duration measurements, it is not intrinsically interesting because duration is known to vary with voicing of surrounding consonants), and, finally, the word itself. The combination of carrier type, vowel, and place of articulation of the onset consonant (analyzed separately) were thought to capture the variation that could potentially be seen in the Word factor.

subjects (see also Babel, 2009) given that gender differences in convergence are well established in the literature (Pardo, 2006).

We first present the male data, and following that we will have an interim summary before moving on to the female data. At the end of this section, we will have another discussion, comparing the female performance to the male performance.

Is there convergence between the pre-exposure and post-exposure productions?

Before we delve into the analysis, we wanted to ensure that subjects did show significant findings of convergence. In Table 5.17, we show the output for mixed effects models for the Repetition factor with all the data included in the analysis. Each row in the table represents the output for the factor of Repetition (pre-test or post-exposure) in separate mixed effect models for each factor for each vowel. For the global measures of duration and f0, the results were split, with overall convergence on duration but not f0, however, further subset analyses showed significant results for f0 convergence (to be discussed in the subsections below). Also, there were overall convergence findings for at least one vowel type for each of the aggregate phonetic measures. Looking at the individual phonetic measures, we saw overall convergence for at least one vowel type for F1 and F2, but no significant findings for F3. As in the pre-test/post-exposure analysis, since no findings of convergence for F3 were shown overall, or for any smaller subsets, we will not discuss this variable further. An inspection of the means reveal that the significant overall results are all results of convergence, not divergence.

	Variable	Vowel Type	Estimate	Std. Error	t-value	p-value	Significant?
Global Measures	Duration	English-like	-12.66	2.327	-5.441	p < 0.0001	*
		Foreign	-14.566	1.951	-7.467	p < 0.0001	*
	F0	English-like	-0.8445	0.4996	-1.69	0.091	ns
		Foreign	-0.5346	0.4123	-1.297	0.1948	ns
Global Phonetic Measures	Euclidean F1+F2+F3	English-like	-0.06332	0.02959	-2.14	0.0324	*
		Foreign	-0.02807	0.01472	-1.91	0.0567	ns
	Euclidean F1+F2	English-like	-0.08517	0.03018	-2.822	0.0048	*
		Foreign	-0.04297	0.0153	-2.808	0.005	*
Individual Phonetic Measures	F1	English-like	0.07428	0.01456	5.103	p < 0.0001	*
		Foreign	0.16876	0.01381	12.22	p < 0.0001	*
	F2	English-like	0.10083	0.04026	2.505	0.0123	*
		Foreign	0.006144	0.022158	0.277	0.7816	ns
	F3	English-like	-0.01243	0.01722	-0.722	0.4704	ns
		Foreign	0.007141	0.011721	0.61	0.5424	ns

Table 5.17. Results from overall convergence analyses, looking at whether there were significant findings of convergence across all measures in the comparison of pre-exposure and post-exposure productions.

Male Participants:

The following sections will present the data from male participants in the pre-exposure/post-exposure comparison. A discussion of the more global measures of convergence (duration and f0) will be presented first, followed by the aggregate phonetic measures (the Euclidean distance measures), and then finally, the individual phonetic measures, the formant measures. As before, we will first present the overall analysis of whether there was convergence, before looking at the relative amounts of convergence. At the end of this section, before moving on to female participants, we will summarize the results for this subject group.

The numbers of participants for these groups was as follows: the auditory exposure, adult-directed register group consisted of 9 male participants, the audiovisual exposure, adult-

directed register group consisted of 6 male participants, the auditory exposure, child-directed register group consisted of 4 male participants, and the audiovisual exposure, child-directed register group consisted of 7 males.

Global Measures

Duration

Overall, duration convergence was observed for both vowel types. Looking at further subsets, noticeably absent were significant results for adult-directed speech, and for English-like vowels in the auditory modality.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
Duration	English-like	✓	✓	✗	✗	✓
	Foreign	✓	✓	✗	✓	✓

Table 5.18. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for duration in male participants. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for duration for the male participants is presented in Table 5.19. Table 5.19 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effect of Subject [$\chi^2(1) = 111.53, p < 0.01$] but not Word [$\chi^2(1) = 0.00, p = 1$] significantly reduced model fit.

Overall, there were no significant main effects or interactions.

<i>Duration Males</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Lambda)$-test</i>			<i>Pseudo-R^2</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	22.16	75.70	0.29	0.770	6.69	5	0.245	0.0002
Modality	36.19	100.53	0.36	0.719	4.18	5	0.524	0.0006
Carrier Type	16.22	9.16	1.77	0.077	3.81	2	0.149	0.0007
Vowel Type	9.19	9.22	1.00	0.319	7.71	4	0.103	0.0013
AQ	1.37	4.82	0.28	0.777	3169.60	3	<0.001*	-0.0008
Register X Modality	15.25	38.26	0.40	0.690	0.00	1	1.000	0.0000
Register X Carrier Type	4.69	8.07	0.58	0.561	0.34	1	0.559	0.0001
Modality X Carrier Type	-13.81	8.31	-1.66	0.097	2.78	1	0.095	0.0005
Register X Vowel Type	8.01	7.77	1.03	0.303	1.07	1	0.300	0.0002
Modality X Vowel Type	5.13	8.00	0.64	0.521	0.42	1	0.519	0.0001
Carrier Type X Vowel Type	-15.30	8.22	-1.86	0.063	3.48	1	0.062	0.0006
Register X AQ	-0.40	4.67	-0.09	0.931	0.00	1	1.000	0.0000
Modality X AQ	-1.68	5.67	-0.30	0.767	0.00	1	1.000	0.0000

Table 5.19. Summary of results from the mixed effects model for duration in pre-exposure compared to post-exposure productions for male participants.

*f*0:

Overall, we saw convergence to both vowel types for the male participants. However, looking at further subsets, there was only significant convergence in the adult-directed register and audiovisual modality. Additionally, there was significant divergence for foreign vowels in the auditory modality.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
f0	English-like	✓	✗	✓	✗	✓
	Foreign	✓	✗	✓	D	✓

Table 5.20. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for f0 in male participants, and overall. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for f0 for the male participants is presented in Table 5.21. Table 5.21 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effect of Subject [Subject: $\chi^2(1) = 406.45, p < 0.01$] but not Word [$\chi^2(1) = 0.00, p = 0.97$] significantly reduced model fit.

f0 Males	Parameter estimates		Wald's test		$\Delta(-2\Delta)$ -test			Pseudo- R^2
	Estimates	S.E.	Z	p_z	c^2	df	p	
Register	57.82	32.08	1.80	0.072	6.77	5	0.238	0.0003
Modality	1.89	42.64	0.04	0.965	11.52	5	0.042	0.0020
Carrier Type	-1.84	2.22	-0.83	0.408	8.31	2	0.016	0.0025
Vowel Type	-3.00	2.23	-1.35	0.179	9.84	4	0.043	0.0030
AQ	1.28	2.05	0.63	0.532	2575.30	3	<0.001	0.0127
Register X Modality	-12.61	16.27	-0.78	0.438	0.44	1	0.506	0.0000
Register X Carrier Type	1.99	1.95	1.02	0.308	1.05	1	0.306	0.0003
Modality X Carrier Type	-4.71	2.01	-2.34	0.019*	5.50	1	0.019	0.0017
Register X Vowel Type	-0.63	1.88	-0.33	0.739	0.12	1	0.734	0.0000
Modality X Vowel Type	-2.10	1.94	-1.09	0.278	1.19	1	0.276	0.0004
Carrier Type X Vowel Type	5.36	1.99	2.69	0.007*	7.25	1	0.007	0.0022
Register X AQ	-3.89	1.98	-1.96	0.050*	4.65	1	0.031	0.0000
Modality X AQ	1.32	2.41	0.55	0.584	0.01	1	0.940	0.0000

Table 5.21. Summary of results from the mixed effects model for F0 in pre-exposure compared to post-exposure productions for male participants.

Overall, there were no significant main effects but there were significant interactions of Modality X Carrier Type, Carrier Type X Vowel Type, and Register X AQ. Post-hoc testing revealed that there were no significant pairwise effects of Carrier Type X Vowel Type (lowest was CV:foreign with CV:English-like at $p = 0.099$), but there were significant pairwise comparisons of Carrier Type X Modality. Across the two carrier types, there were no significant differences within a single modality (auditory: $p = 0.92$; visual: $p = 0.40$), but between modalities, all pairwise comparisons were significantly different (CV:Audiovisual – CV:Auditory, $p < 0.0001$; CVC:Audiovisual – CV:Auditory, $p < 0.0001$; CV:Audiovisual – CVC:Auditory, $p < 0.0001$; CVC:Audiovisual – CVC:Auditory, $p < 0.0001$), demonstrating a modality effect, and this data is shown in Figure 5.25 below.

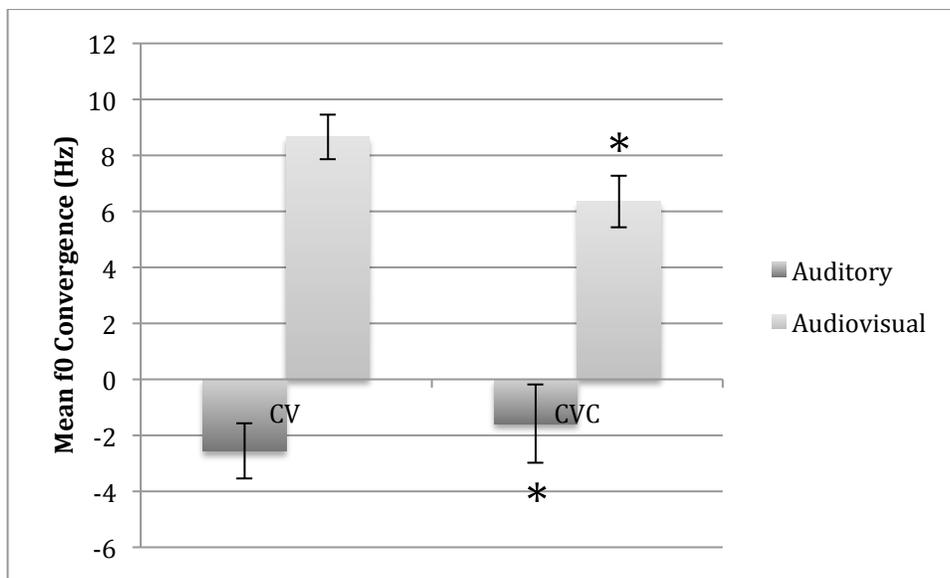


Figure 5.25. Results for convergence in f_0 by male participants by carrier type and modality of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Additionally, register also interacted with AQ score (Figure 5.26). As AQ score got higher, participants were better at imitation, but only in the adult-directed register ($p = 0.007$); the child-directed register did not show significant influence of AQ score on imitation ($p = 0.59$). This result does not relate to any of the hypotheses of the experiment and will not be discussed further.

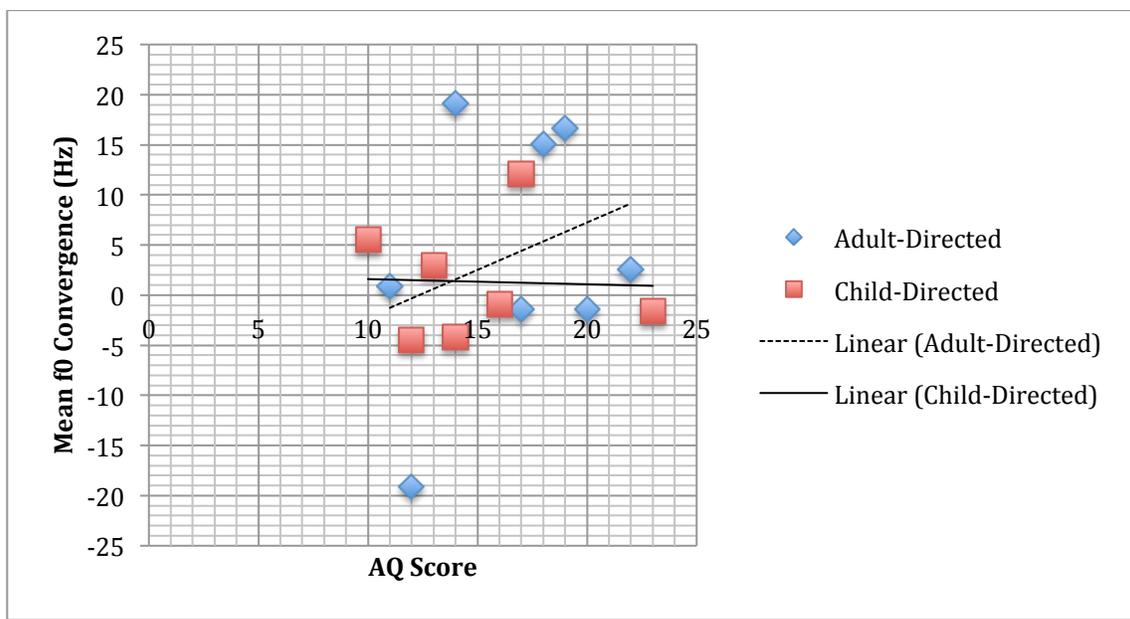


Figure 5.26. Scatterplot of each subjects AQ and the difference in pre-test and post-exposure f0 measurement for each register. The lines are trend lines, and only the slope for the adult-directed register is significant.

Summary: Global Measures – Male Participants

The results for the global measures in the pre-exposure/post-exposure comparison appear to pattern similarly to the rest of the data, favoring both an advantage of child-directed speech, with one exception, and the audiovisual modality. Taking the modality effects first, we saw a

difference in significant convergence results related to modality. For male participants, we saw that there was no significant finding of convergence to duration in the auditory modality; rather there was only convergence in the audiovisual modality, for English-like vowels only. For foreign vowels, convergence was shown in both modalities for duration. For f_0 , the male participants showed significant divergence to the auditory modality for foreign vowels, and convergence to both vowel types in the audiovisual modality, demonstrating a modality effect. Additionally, in the analysis of relative amounts of convergence for f_0 , convergence was shown to be significantly greater for the audiovisual modality than for the auditory modality (this appeared greater for CV than CVC syllables).

Looking at the register effects, the duration results are much more clear than the f_0 results. We saw an overall register effect, with convergence only for child-directed speech on the duration measure. For f_0 , contradictory to the prediction, there was a significant finding of convergence for adult-directed speech, and not for child-directed speech. It is not clear why specifically males were less inclined to implicitly imitate f_0 in child-directed speech; it could be due to that the f_0 of child-directed speech is more different from their natural pronunciation, or it could be associated with social factors. Regardless, it will not be discussed further.

Aggregate Phonetic Measures

Euclidean Distance ($F1 + F2 + F3$):

When we looked at convergence for the overall Euclidean Distance measure, male participants showed significant convergence only for English-like vowels in the audio-visual modality.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
Euclidean F1+F2+F3	English-like	×	×	×	×	✓
	Foreign	×	×	×	×	×

Table 5.22. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1 + F2 + F3) in male participants, and overall.

The full statistical output is shown in the appendix.

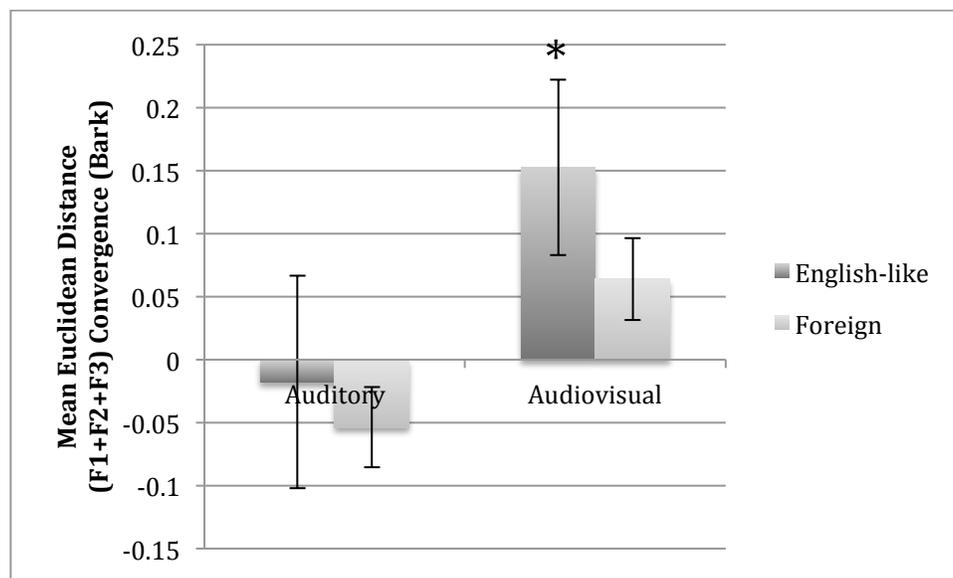


Figure 5.27. Results for overall convergence for male participants on the Euclidean Distance (F1+F2+F3) measure for English-like and foreign vowels by modality of exposure in the pre-exposure/post-exposure comparison.

Significance stars represent significant convergence or divergence (difference from zero) with a *p*-value of 0.05 or smaller.

The final output for the mixed effects model including all factors and interactions for Euclidean Distance (F1+F2+F3) for the male participants is presented in Table 5.23. Table 5.23 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effect of Subject [$\chi^2(1) = 7.00, p < 0.01$] but not Word [$\chi^2(1) = 0.00, p = 1.0$] significantly reduced model fit. Overall, there were no significant main effects or interactions.

<i>Euclidean Distance (F1/F2/F3) Males</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Pseudo- R²</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	0.94	0.58	1.61	0.107	3.35	5	0.646	0.0001
Modality	0.29	0.77	0.38	0.707	3.40	5	0.638	0.0000
Carrier Type	0.00	0.14	0.00	0.998	2.51	2	0.285	0.0001
Vowel Type	-0.13	0.14	-0.92	0.360	2.99	4	0.559	0.0001
AQ	0.01	0.04	0.40	0.687	755.92	3	<0.001	0.0002
Register X Modality	-0.23	0.29	-0.80	0.425	0.50	1	0.482	0.0000
Register X Carrier Type	-0.14	0.12	-1.15	0.252	1.33	1	0.250	0.0001
Modality X Carrier Type	0.01	0.13	0.12	0.908	0.02	1	0.891	0.0000
Register X Vowel Type	-0.03	0.12	-0.29	0.775	0.09	1	0.770	0.0000
Modality X Vowel Type	-0.03	0.12	-0.26	0.796	0.07	1	0.785	0.0000
Carrier Type X Vowel Type	0.14	0.13	1.08	0.278	1.19	1	0.275	0.0001
Register X AQ	-0.04	0.04	-1.19	0.235	1.58	1	0.210	0.0000
Modality X AQ	0.00	0.04	0.04	0.971	0.00	1	1.000	0.0000

Table 5.23. Summary of results from the mixed effects model for Euclidean Distance (F1+F2+F3) in pre-exposure compared to post-exposure productions for male participants.

Euclidean Distance (F1 + F2):

Again, male participants only converged on the Euclidean Distance (F1 + F2), when presented English-like vowels in the audiovisual modality.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
Euclidean F1+F2	English-like	×	×	×	×	✓
	Foreign	×	×	×	×	×

Table 5.24. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1 + F2) in male participants, and overall. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for Euclidean Distance (F1+F2) for the male participants is presented in Table 5.25. Table 5.25 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effects revealed no significant differences in the model fit [Subject: $\chi^2(1) = 2.07, p = 0.150$; Word: $\chi^2(1) = 0.00, p = 1$]. Overall, there were no significant main effects or interactions.

<i>Euclidean Distance (F1/F2) Males</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Pseudo- R²</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>		<i>p</i>	
Register	0.37	0.49	0.76	0.448	1.48	5	0.915	0.0001
Modality	0.40	0.64	0.63	0.532	6.30	5	0.278	0.0000
Carrier Type	-0.02	0.13	-0.16	0.871	0.55	2	0.760	0.0000
Vowel Type	0.06	0.13	0.46	0.648	4.59	4	0.332	0.0001
AQ	0.02	0.03	0.68	0.495	703.08	3	<0.001*	0.0000
Register X Modality	-0.10	0.24	-0.40	0.686	0.00	1	1.000	0.0000
Register X Carrier Type	-0.06	0.12	-0.50	0.617	0.26	1	0.612	0.0000
Modality X Carrier Type	0.00	0.12	0.03	0.974	0.01	1	0.941	0.0000
Register X Vowel Type	-0.08	0.11	-0.67	0.502	0.46	1	0.500	0.0000
Modality X Vowel Type	-0.19	0.12	-1.66	0.097	2.79	1	0.095	0.0001
Carrier Type X Vowel Type	0.06	0.12	0.53	0.594	0.29	1	0.588	0.0000
Register X AQ	-0.01	0.03	-0.25	0.804	0.00	1	1.000	0.0000
Modality X AQ	0.00	0.04	-0.11	0.910	0.00	1	1.000	0.0000

Table 5.25. Summary of results from the mixed effects model for Euclidean Distance (F1+F2) in pre-exposure compared to post-exposure productions for male participants.

Summary: Aggregate Phonetic Measures – Male Participants

There were no observable register effects, and a non-comprehensive modality effect in male participants. Taking all subsets of male data into account, there was only one significant finding of convergence: for English-like vowels in the audiovisual modality. This effect was seen on both aggregate phonetic measures.

Individual Phonetic Measures

F1:

Male participants showed no convergence in any subset of English-like vowels. For foreign vowels, there were no modality effects, but there was a register effect: male subjects significantly converged to child-directed speech but diverged to adult-directed speech.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
F1	English-like	×	×	×	×	×
	Foreign	✓	✓	D	✓	✓

Table 5.26. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F1 in male participants, and overall. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for F1 for the male participants is presented in Table 5.27. Table 5.27 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effects revealed no significant differences in model fit [Subject: $\chi^2(1) = 3.58, p = 0.058$; Word: $\chi^2(1) = 0.00, p = 0.977$].

<i>F1 Males</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Pseudo-R^2</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-0.24	0.20	-1.19	0.236	9.50	5	0.091	0.0004
Modality	-0.14	0.27	-0.52	0.602	10.19	5	0.070	0.0005
Carrier Type	-0.12	0.05	-2.29	0.022*	2.88	2	0.237	0.0001
Vowel Type	-0.09	0.05	-1.73	0.083	11.12	4	0.025	0.0005
AQ	0.00	0.01	-0.32	0.749	295.03	3	<0.001	-0.0002
Register X Modality	-0.01	0.10	-0.09	0.927	0.00	1	1.000	0.0000
Register X Carrier Type	0.08	0.05	1.65	0.099	2.74	1	0.098	0.0001
Modality X Carrier Type	0.14	0.05	2.85	0.004*	8.14	1	0.004	0.0004
Register X Vowel Type	0.12	0.05	2.59	0.010*	6.70	1	0.010	0.0003
Modality X Vowel Type	0.07	0.05	1.61	0.107	2.62	1	0.106	0.0001
Carrier Type X Vowel Type	0.02	0.05	0.36	0.716	0.14	1	0.710	0.0000
Register X AQ	0.01	0.01	0.80	0.423	0.50	1	0.477	0.0000
Modality X AQ	0.00	0.02	-0.06	0.954	0.00	1	1.000	0.0000

Table 5.27. Summary of results from the mixed effects model for F1 in pre-exposure compared to post-exposure productions for male participants.

Overall, there was a significant main effect of Carrier Type and significant interactions of Modality X Carrier Type and Register X Vowel Type. However, post-hoc testing revealed no significant effects of Carrier Type ($p = 0.16$), of Modality ($p = 0.45$), of the interaction of Carrier Type X Modality ($p = 0.07$), or of any of the pairwise comparisons (the visual CV: visual CVC interaction was the closest to significance at $p = 0.107$).

Post-hoc testing of the Register X Vowel Type interaction revealed significantly greater convergence in the child-directed register than the adult-directed register, but only for foreign vowels ($p = 0.007$). Also significant was the pairwise comparison between convergence to foreign and English-like vowels in the child-directed register ($p = 0.037$). This data is shown in Figure 5.28.

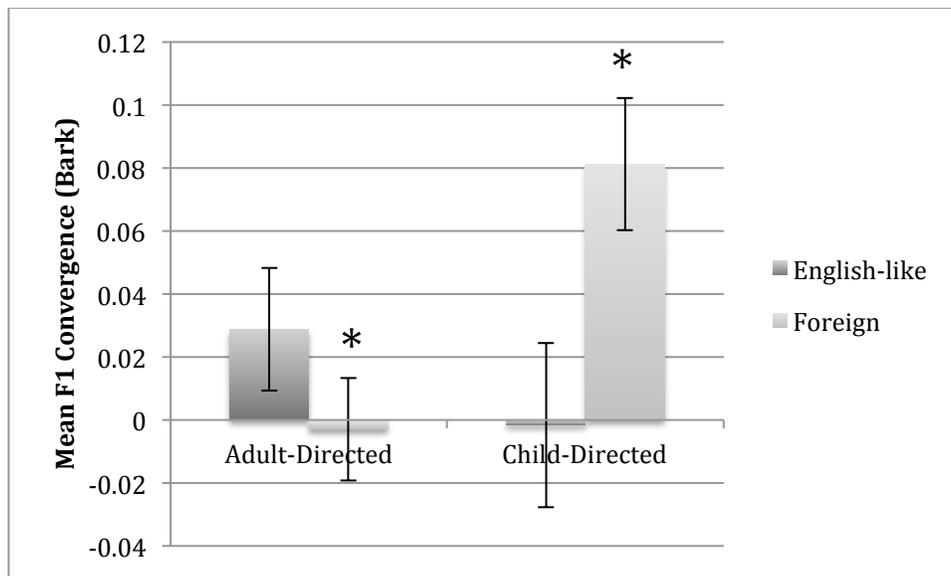


Figure 5.28. Results for overall convergence for male participants on the F1 measure by register for English-like and foreign vowels by modality of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

F2:

Results for overall analysis of convergence for F2 for male participants showed significant convergence only in the audiovisual modality and child-directed register, and only for English-like vowels. For foreign vowels, the only significant result was of divergence in the auditory modality.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
F2	English-like	✓	✓	✗	✗	✓
	Foreign	✗	✗	✗	D	✗

Table 5.28. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F2 in male participants, and overall. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for F2 for the male participants is presented in Table 5.29. Table 5.29 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effect of Subject [$\chi^2(1) = 6.71, p < 0.01$] but not Word [$\chi^2(1) = 0.00, p = 1$] significantly reduced model fit. Overall, there were no significant main effects or interactions.

<i>F2 Males</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Pseudo-R^2</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	0.48	0.57	0.86	0.393	2.70	5	0.746	0.0001
Modality	0.41	0.75	0.55	0.584	7.08	5	0.215	0.0001
Carrier Type	0.03	0.14	0.21	0.831	0.86	2	0.650	0.0000
Vowel Type	0.06	0.14	0.45	0.655	8.18	4	0.085	0.0004
AQ	0.02	0.04	0.63	0.527	730.67	3	<0.001	0.0002
Register X Modality	-0.03	0.28	-0.10	0.918	0.00	1	1.000	0.0000
Register X Carrier Type	-0.08	0.12	-0.68	0.496	0.47	1	0.491	0.0000
Modality X Carrier Type	-0.09	0.12	-0.73	0.468	0.53	1	0.467	0.0000
Register X Vowel Type	-0.16	0.12	-1.33	0.183	1.78	1	0.182	0.0001
Modality X Vowel Type	-0.21	0.12	-1.73	0.084	3.02	1	0.082	0.0001
Carrier Type X Vowel Type	0.08	0.12	0.62	0.537	0.39	1	0.531	0.0000
Register X AQ	-0.01	0.03	-0.41	0.684	0.00	1	1.000	0.0000
Modality X AQ	0.00	0.04	0.00	0.997	0.00	1	1.000	0.0000

Table 5.29. Summary of results from the mixed effects model for F2 in pre-exposure compared to post-exposure productions for male participants.

F3:

Recall that subjects did not converge with the model talker on F3, so we will not look any further into this variable.

Summary: Individual Phonetic Measures – Male Participants

For the individual phonetic measures for male participants, we saw the same pattern of results for modality and register (advantage to AV and child-directed speech), but once again, this was not complete across all measures and vowel types. Starting with register, we only saw a register effect for F2, and only for English-like vowels; there was significant convergence for the

child-directed register and not for the adult-directed register. For F1, we saw no register effects in whether there was convergence for either English-like or foreign vowels (there were convergence findings for both registers for foreign vowels, and no convergence findings for either register for English-like vowels). For F1, we also saw a significant effect that for foreign vowels, there was a greater amount of convergence in the child-directed register than for the adult-directed register.

When considering effects of modality, we saw some effects of modality, but only in F2. For F2, we saw significant convergence only in the audiovisual modality for English-like vowels, and divergence in the auditory modality for foreign vowels. There were no modality differences for F1.

Summary: Male Participants

Overall, male participants showed significant convergence findings on several different measures in the pre-exposure/post-exposure comparison, however, these effects were sometimes restricted to English-like vowels, and sometimes to foreign vowels. Further, there were many subsets in which we saw no evidence of these effects.

Register Effects:

For register, we saw an advantage of child-directed speech in the global measure of duration and the individual phonetic measures of F1 and F2, however, these were not comprehensive. For F1, we only saw a register effect for foreign vowels, and for F2, we only saw an effect for English-like vowels. Additionally, these were difference effects; for F1, the

effect was a significant difference in the amount of convergence, whereas for F2, this was a difference in whether there was a significant finding of convergence. Finally, there was one register effect in the unexpected direction, and that was for f0; males significantly converged to f0 in adult-directed but not child-directed speech. While this finding is in contrast to the hypothesis of this experiment as related to speaking register, it is not entirely unexpected, because there are a number of social factors that could plausibly influence male participants' imitation of f0 in the child-directed register, which we will not discuss, as they are not the focus of the current study.

Modality Effects:

There was evidence of modality effects for all of the different types of measures analyzed for this pre-exposure/post-exposure comparison in male participants. These all were in favor of imitation in the audiovisual, and not auditory, modality, but they were not robust. Starting with global measures, we saw convergence in the audiovisual and not auditory modality, but only for English-like vowels for duration. The effect for f0 was more robust, with greater convergence in the audiovisual modality compared to the auditory modality, regardless of vowel type. For both the aggregate phonetic measures, the only subset that showed significant convergence related to modality was for English-like vowels in the audiovisual modality (no other subsets showed this effect). This seems to be driven by the individual phonetic measure of F2, in which we see the same pattern (only significant convergence for audiovisual/foreign pairing), and also significant divergence for foreign vowels in the auditory modality. F1 does not seem to be driving these results for the global phonetic measures because it did not show any modality effect.

Vowel Type Effects:

Although there were several vowel type differences - male subjects imitated the English-like and foreign vowels differently - no consistent pattern emerged. For the global measures, we saw convergence on both vowel types. For the aggregate phonetic measures, we only saw significant convergence results for the English-like vowels, and for the individual phonetic measures, we saw significant convergence for F1 in foreign vowels, and for F2 in English-like vowels. Since these results present no clear picture of how English-like and foreign vowels were imitated, we will not discuss them further.

Female Participants:

The following sections will present the data from female participants in the pre-exposure/post-exposure comparison. A discussion of the more global measures of convergence (duration and f0) will be presented first, followed by the global phonetic measures (the Euclidean distance measures), and then finally, the individual phonetic measures, the formant measures. At the end of this section, before moving on a comparison between male and female participants, we will summarize the results for this subject group.

The numbers of participants for these groups was as follows: the auditory exposure, adult-directed register group consisted of 11 female participants, the audiovisual exposure, adult-directed register group consisted of 12 female participants, the auditory exposure, child-directed register group consisted of 13 female participants, and the audiovisual exposure, child-directed register group consisted of 11 females.

Global Measures

Duration:

Female participants converged overall on duration to both vowel types. Noticeably absent was convergence to English-like vowels in the auditory modality, and to foreign vowels in the adult-directed register.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
Duration	English-like	✓	✓	✓	✗	✓
	Foreign	✓	✓	✗	✓	✓

Table 5.30. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for duration in female participants, and overall. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for duration for the female participants is presented in Table 5.31. Table 5.31 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. The model also included two random intercepts – one for Subject, and one for word. Removing the random effects reduced model fit significantly [Subject: $\chi^2(1) = 361.99, p < 0.01$; Word: $\chi^2(1) = 7.18, p < 0.01$].

<i>Duration Females</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Pseudo-R^2</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-41.02	33.82	-1.21	0.225	10.48	5	0.063	-0.0001
Modality	-54.14	31.28	-1.73	0.084	15.04	5	0.010	0.0005
Carrier Type	-5.53	7.56	-0.73	0.464	7.40	2	0.025	0.0068
Vowel Type	0.89	7.70	0.12	0.908	9.26	4	0.055	0.0070
AQ	-2.33	1.57	-1.49	0.137	2448.10	3	0.000	-0.0018
Register X Modality	43.53	21.89	1.99	0.047*	4.28	1	0.038	0.0000
Register X Carrier Type	-0.56	6.01	-0.09	0.925	0.01	1	0.914	0.0000
Modality X Carrier Type	12.64	6.00	2.11	0.035*	4.45	1	0.035	0.0004
Register X Vowel Type	2.56	5.78	0.44	0.658	0.20	1	0.655	0.0000
Modality X Vowel Type	-7.01	5.76	-1.22	0.224	1.48	1	0.223	0.0001
Carrier Type X Vowel Type	2.81	7.64	0.37	0.713	0.11	1	0.742	0.0001
Register X AQ	2.68	1.95	1.37	0.170	2.01	1	0.157	0.0000
Modality X AQ	2.78	1.89	1.47	0.141	2.33	1	0.127	0.0000

Table 5.31. Summary of results from the mixed effects model for duration in pre-exposure compared to post-exposure productions for female participants.

Overall, there were significant main interactions between Register X Modality and Modality X Carrier Type. Post-hoc testing revealed significant pairwise interactions within Register X Modality: we saw significant differences between the imitation in the child and adult registers for audiovisual speech, differences between the visual and auditory modality for child-directed speech, and also a significant difference between Child:AudioVisual and Adult:Auditory (all $p < 0.0001$). These results are shown visually below in Figure 5.29. Overall, female subjects imitated duration better after exposure to child-directed audiovisual speech.

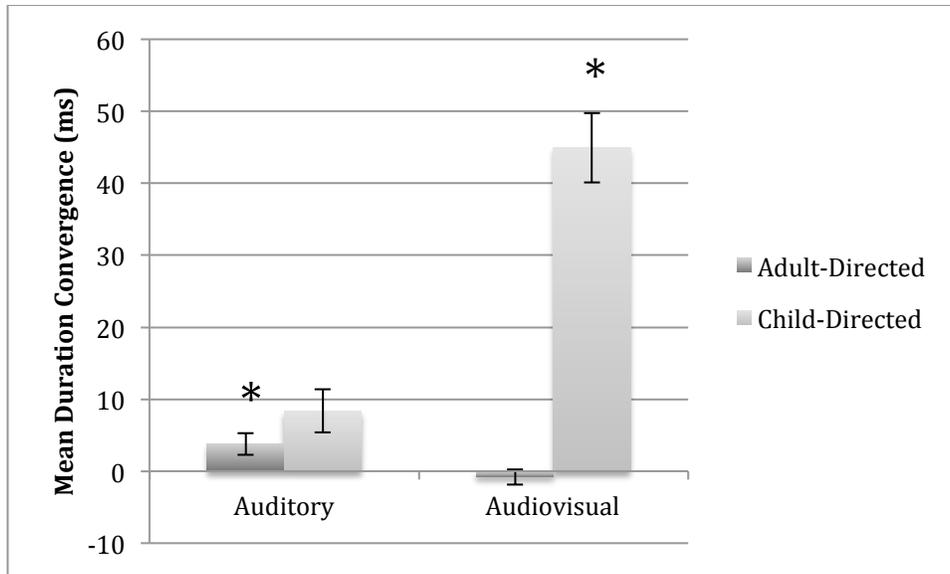


Figure 5.29. Comparison of pre-exposure and post-exposure phonetic distance in each register by modality for female subjects in the duration measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Finally, the interaction of Modality X Carrier Type also was confirmed to be significant in post-hoc testing ($p = 0.039$). Significant pairwise differences occurred between the Visual:CVC condition and the Auditory:CV condition ($p = 0.003$), as well as between the CVC words in the auditory and audiovisual modalities ($p < 0.0001$). Overall, subjects imitated duration better if they were imitating CVC words in the visual modality, compared to CVC words in the auditory modality. (Figure 5.30).

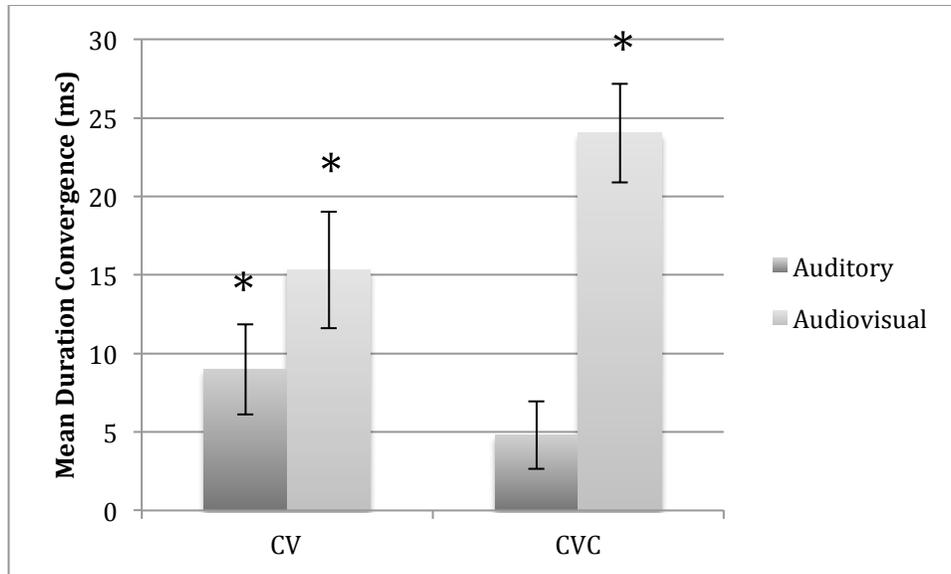


Figure 5.30. Comparison of pre-exposure and post-exposure phonetic distance in each modality by context type for female subjects in the duration measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

f0:

Female participants showed significant convergence for foreign vowels in the child-directed register. Additionally, we found significant divergence to both vowel types in the adult-directed register, enhancing the register effect. As for modality differences, female participants converged in the audiovisual modality and divergence in the auditory modality, but only for foreign vowels.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
f0	English-like	×	×	D	×	×
	Foreign	×	✓	D	D	✓

Table 5.32. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for f0 in female participants, and overall. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for f0 for the female participants is presented in Table 5.33. Table 5.33 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effect of Subject [Subject: $\chi^2(1) = 516.98, p < 0.01$] but not Word [$\chi^2(1) = 0.00, p = 0.98$] significantly reduced model fit.

For the fundamental frequency variable, no significant main effects or interactions were observed.

<i>f0 Females</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Pseudo-R^2</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	8.59	8.63	1.00	0.319	5.78	5	0.329	0.0002
Modality	5.19	7.98	0.65	0.515	4.08	5	0.538	0.0002
Carrier Type	2.13	1.46	1.46	0.145	2.64	2	0.268	0.0003
Vowel Type	0.11	1.48	0.07	0.942	6.31	4	0.177	0.0006
AQ	0.06	0.40	0.16	0.877	1884.00	3	0.000	0.0000
Register X Modality	-3.69	5.60	-0.66	0.510	0.35	1	0.556	0.0000
Register X Carrier Type	-0.91	1.32	-0.69	0.489	0.48	1	0.488	0.0000
Modality X Carrier Type	0.58	1.32	0.44	0.661	0.19	1	0.659	0.0000
Register X Vowel Type	1.75	1.27	1.38	0.168	1.91	1	0.167	0.0002
Modality X Vowel Type	1.67	1.27	1.32	0.189	1.73	1	0.188	0.0002
Carrier Type X Vowel Type	-1.96	1.34	-1.47	0.142	2.16	1	0.142	0.0002
Register X AQ	-0.16	0.50	-0.32	0.747	0.00	1	1.000	0.0000
Modality X AQ	-0.06	0.48	-0.13	0.895	0.00	1	1.000	0.0000

Table 5.33. Summary of results from the mixed effects model for F0 in pre-exposure compared to post-exposure productions for female participants.

Summary: Global Measures

The analysis of the global measures for female participants showed only slight modality and register effects in terms of significant convergence, but the mixed effects models provided additional evidence for an advantage of the audiovisual modality and the child-directed register, and this evidence additionally pointed towards an interaction between register and modality. Starting with duration, we saw that there was a register advantage in overall convergence only for foreign vowels (we saw significant convergence in the child-directed register but not the adult-directed register for this vowel type) and a modality advantage in overall convergence only for English-like vowels (we saw significant convergence in the audiovisual and not auditory modality for this vowel type). However, in the mixed effects models, female participants showed

register effects only in the audiovisual modality, and modality effects only within the child-directed register. Modality effects were additionally qualified by the interaction of Modality and Carrier Type, which showed that there was a significant advantage of modality in CVC words only.

For f0, the advantage of the child-directed register and audiovisual modality were only evident in overall convergence measures. For register, we found significant divergence in the adult-directed register for female participants for both vowel types, and also convergence for foreign vowels. These results all point to an advantage of the child-directed register, but they point towards this by showing a disadvantage of the adult-directed register. For female participants, we only saw a significant advantage to the audiovisual modality for foreign vowels (divergence in the auditory modality and convergence in the audiovisual modality). This advantageous effect only for English-like vowels supports the theory that the visual modality will aid more in imitation of foreign compared to English-like vowels.

Overall, the global measures showed significant overall effects of modality and register, all in the expected direction, and additionally, there was some support in these measures that the child-directed register and the audiovisual modality could perhaps aid more for learning foreign vowels. However, there was also one result contradicting this hypothesis, which was the advantage of modality on duration convergence only for *English-like* vowels. Perhaps this finding is actually not contrastive to the hypothesis, and the female subjects converged less on duration of this measure because they were ignoring duration to focus on other features made more evident in the audiovisual modality?

Aggregate Phonetic Measures

Euclidean Distance (F1 + F2 + F3):

For the Euclidean Distance (F1 + F2 + F3) measure, female participants converged in the child-directed register but not in the adult-directed register. However, modality effects in female participants were only seen for foreign vowels: convergence was only see in the audiovisual modality.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
Euclidean F1+F2+F3	English-like	✗	✓	✗	✗	✗
	Foreign	✓	✓	✗	✗	✓

Table 5.34. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1 + F2 + F3) in female participants, and overall. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for Euclidean Distance (F1+F2+F3) for the female participants is presented in Table 5.35. Table 5.35 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effect of Subject [$\chi^2(1) = 13.49, p < 0.01$] but not Word [$\chi^2(1) = 0.00, p = 0.99$] significantly reduced model fit.

<i>Euclidean Distance (F1/F2/F3) Females</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Pseudo- R²</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-0.21	0.14	-1.46	0.145	10.39	5	0.065	-0.0001
Modality	-0.05	0.13	-0.39	0.696	10.30	5	0.067	0.0000
Carrier Type	0.01	0.07	0.15	0.883	0.01	2	0.996	0.0000
Vowel Type	-0.09	0.07	-1.30	0.194	5.45	4	0.244	0.0001
AQ	0.00	0.01	-0.64	0.521	645.79	3	<0.001	-0.0004
Register X Modality	0.20	0.09	2.27	0.023*	5.58	1	0.018	-0.0001
Register X Carrier Type	0.00	0.06	0.06	0.956	0.01	1	0.942	0.0000
Modality X Carrier Type	0.02	0.06	0.31	0.754	0.10	1	0.751	0.0000
Register X Vowel Type	-0.01	0.06	-0.13	0.893	0.02	1	0.887	0.0000
Modality X Vowel Type	0.12	0.06	2.03	0.042*	4.15	1	0.042	0.0001
Carrier Type X Vowel Type	0.00	0.06	0.02	0.988	0.00	1	0.963	0.0000
Register X AQ	0.01	0.01	1.55	0.122	2.58	1	0.108	0.0000
Modality X AQ	-0.01	0.01	-1.10	0.273	1.24	1	0.266	0.0000

Table 5.35. Summary of results from the mixed effects model for Euclidean Distance (F1+F2+F3) in pre-exposure compared to post-exposure productions for female participants.

Overall, there were no significant main effects but there were significant interactions of Register X Modality and Modality X Vowel Type. Post-hoc tests revealed that there was an advantage of the child-directed register only in the audiovisual modality. The only significant pairwise comparison was the Child:AudioVisual with Adult:AudioVisual interaction ($p < 0.001$). This data is shown below in Figure 5.31.

Post-hoc testing revealed that all interactions of Modality X Vowel Type were not significant.

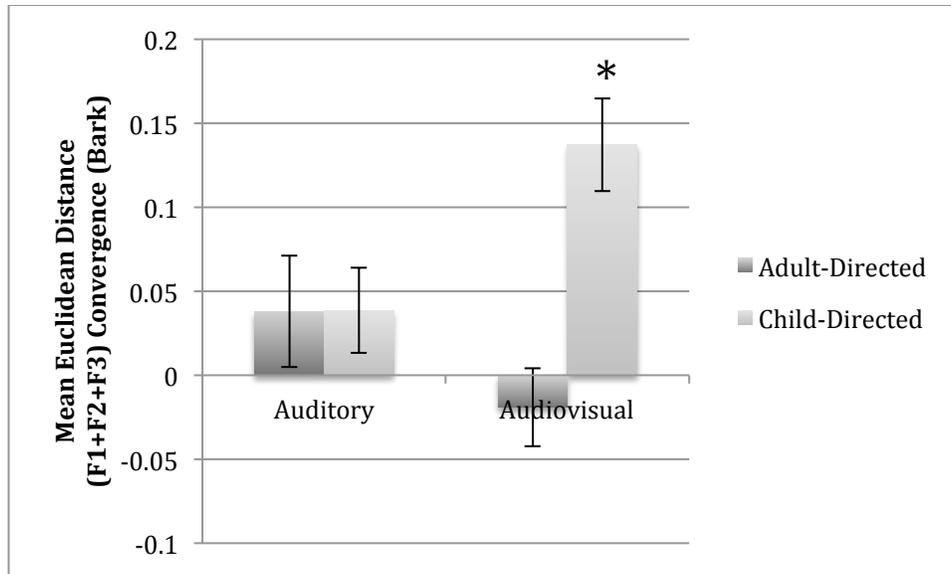


Figure 5.31. Comparison of pre-exposure and post-exposure phonetic distance in each register by modality for female subjects in the Euclidean distance convergence (F1/F2/F3) measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Euclidean Distance (F1 + F2):

We saw the same basic pattern in significant convergence findings for the Euclidean Distance (F1+F2) measure as for the more global measure of Euclidean Difference (F1+F2+F3). The register effect for this measure was exactly the same as for the measure including F3: convergence was only seen in the child-directed register. The modality subsets gave mixed results, convergence was observed only in the auditory domain for English-like vowels, and only in the audiovisual domain for foreign vowels.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
Euclidean F1+F2	English-like	✓	✓	✗	✓	✗
	Foreign	✓	✓	✗	✗	✓

Table 5.36. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1 + F2) in female participants, and overall. The full statistical output is shown in the appendix.

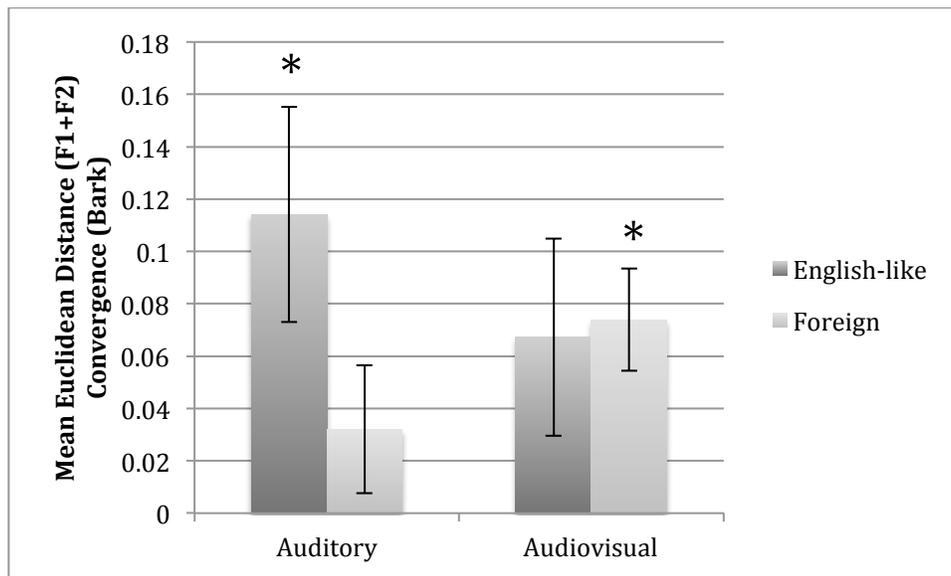


Figure 5.32. Comparison of pre-exposure and post-exposure phonetic distance in each modality by vowel type for female subjects in the Euclidean distance convergence (F1/F2) measurement. Significance stars represent significant convergence or divergence (difference from zero) with a *p*-value of 0.05 or smaller.

The final output for the mixed effects model including all factors and interactions for Euclidean Distance (F1+F2) for the female participants is presented in Table 5.36. Table 5.36 also presents results from likelihood ratio tests comparing the full model with models excluding

each factor or interaction. Removing the random effect of Subject [Subject: $\chi^2(1) = 15.3, p < 0.01$] but not Word [$\chi^2(1) = 0.00, p = 0.99$] significantly reduced model fit.

<i>Euclidean Distance (F1/F2) Females</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Pseudo-R²</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-0.21	0.16	-1.33	0.183	9.83	5	0.080	-0.0001
Modality	-0.12	0.15	-0.83	0.406	7.03	5	0.218	0.0001
Carrier Type	-0.01	0.07	-0.12	0.907	0.15	2	0.926	0.0000
Vowel Type	-0.14	0.07	-1.98	0.048*	7.90	4	0.095	0.0001
AQ	-0.01	0.01	-1.03	0.302	654.04	3	0.000	-0.0003
Register X Modality	0.15	0.10	1.59	0.112	2.73	1	0.098	0.0000
Register X Carrier Type	0.02	0.06	0.37	0.715	0.13	1	0.714	0.0000
Modality X Carrier Type	0.02	0.06	0.30	0.766	0.09	1	0.764	0.0000
Register X Vowel Type	0.05	0.06	0.82	0.413	0.68	1	0.411	0.0000
Modality X Vowel Type	0.13	0.06	2.10	0.036*	4.43	1	0.035	0.0001
Carrier Type X Vowel Type	-0.01	0.06	-0.13	0.900	0.02	1	0.894	0.0000
Register X AQ	0.01	0.01	1.43	0.152	2.20	1	0.138	0.0000
Modality X AQ	0.00	0.01	-0.39	0.698	0.01	1	0.908	0.0000

Table 5.37. Summary of results from the mixed effects model for Euclidean Distance (F1+F2) in pre-exposure compared to post-exposure productions for female participants.

Overall, there was a significant main effect of Vowel Type and a significant interaction of Modality X Vowel Type. As with the F1+F2+F3 Euclidean Difference measure, these effects relating to Vowel Type were not significant in post-hoc testing. However, they mirror effects shown in overall convergence analyses, and are plotted above in Figure 5.32.

Summary: Aggregate Phonetic Measures

For the aggregate phonetic measures, we once again saw an influence of register and modality. For both aggregate phonetic measures, we saw significant convergence for both vowel types in the child-directed register and not the adult-directed register. Additionally, in the Euclidean Distance ($F1 + F2 + F3$) we found that in the audiovisual modality there was significantly greater convergence in the child-directed register than the adult-directed register. This interaction of register and modality was the same as what was found for the global measure of duration.

As for modality, we saw an advantage of the audiovisual modality, but only for foreign vowels. We saw that for both the global phonetic measures, there was significant convergence in the audiovisual and not auditory modality for foreign vowels. Overall, we hypothesized that the foreign vowels could show greater effects of modality, because (a) they are unfamiliar, and participants may *need* to attend to these cues, and (b) the foreign vowels contain visually salient features. For English-like vowels, we actually saw an advantage of the auditory modality in the measure not including F3. We will look to the analysis of the individual phonetic measures in the section below to try and explain this finding.

Individual Phonetic Measures

F1:

There was significant convergence across all data subsets for F1, for female participant. No modality or register effects were shown in the analysis of whether there is convergence.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
F1	English-like	✓	✓	✓	✓	✓
	Foreign	✓	✓	✓	✓	✓

Table 5.38. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F1 in female participants, and overall. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for duration for the female participants is presented in Table 5.38. Table 5.38 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effect of Subject [Subject: $\chi^2(1) = 3.90, p < 0.01$] but not Word [$\chi^2(1) = 0.18, p = 0.67$] significantly reduced model fit.

F1 Females	Parameter estimates		Wald's test		$\Delta(-2\Delta)$ -test			Pseudo- R^2
	Estimates	S.E.	Z	p_z	c^2	df	p	
Register	-0.13	0.14	-0.93	0.350	8.32	5	0.140	0.0056
Modality	-0.16	0.13	-1.29	0.198	10.10	5	0.073	0.0003
Carrier Type	0.05	0.04	1.16	0.246	3.62	2	0.164	0.0000
Vowel Type	-0.01	0.04	-0.13	0.898	21.07	4	0.000	0.0002
AQ	0.00	0.01	-0.09	0.929	387.78	3	0.000	-0.0001
Register X Modality	0.07	0.09	0.87	0.385	0.72	1	0.395	0.0000
Register X Carrier Type	0.03	0.04	0.88	0.377	0.79	1	0.375	0.0000
Modality X Carrier Type	-0.02	0.04	-0.54	0.590	0.29	1	0.587	0.0000
Register X Vowel Type	0.09	0.04	2.30	0.021*	5.31	1	0.021*	0.0002
Modality X Vowel Type	0.11	0.04	2.89	0.004*	8.38	1	0.004*	0.0003
Carrier Type X Vowel Type	-0.07	0.04	-1.67	0.095	2.83	1	0.093	-0.0001
Register X AQ	0.00	0.01	0.54	0.593	0.17	1	0.676	0.0000
Modality X AQ	0.00	0.01	0.35	0.728	0.00	1	1.000	0.0000

Table 5.39. Summary of results from the mixed effects model for F1 in pre-exposure compared to post-exposure productions for female participants.

Overall, there were no significant main effects but there were significant interactions of: Register X Vowel Type and Modality X Vowel Type. Post-hoc testing confirmed a significant affect of Register X Vowel Type ($p = 0.009$) as well as of Modality X Vowel Type ($p < 0.001$). Pairwise comparison of the significant Register X Vowel Type interactions (Figure 5.33) revealed significant differences between English-like and foreign vowels in the child-directed register ($p = 0.008$), between foreign vowels in the two registers ($p < 0.001$), and for the Child:foreign-Adult:English-like comparison ($p = 0.006$). Overall, this shows that F1 of vowels was imitated significantly better in the child-directed register, but only for foreign vowels.

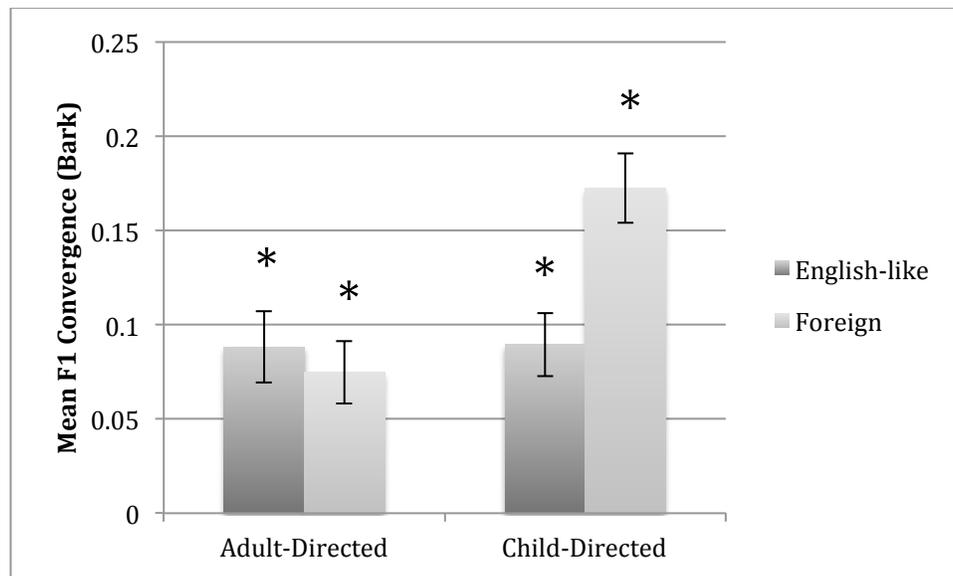


Figure 5.33. Comparison of pre-exposure and post-exposure phonetic distance for each type of vowel by register for female subjects in the F1 measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

For the pairwise analysis of the Modality X Vowel Type interaction (Figure 5.34), we saw significant pairwise comparisons of English-like vowels in the two modalities ($p = 0.002$), and of English-like and foreign vowels in the audiovisual modality ($p = 0.001$), and finally, of the English-like:Audiovisual – foreign:Auditory interaction ($p = 0.015$). Overall, subjects imitated F1 significantly better, for English-like vowels in the auditory modality, which could be due to that openness is salient auditorily. However, we also saw that in the audiovisual modality, they imitated foreign vowels greater than the English-like vowels (and at a level comparable to the other vowel types in the auditory modality).

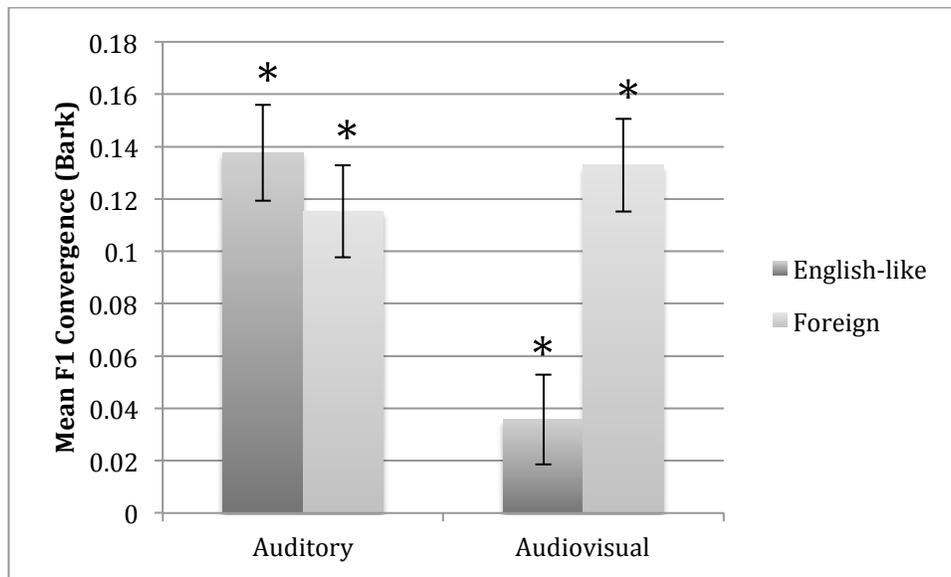


Figure 5.34. Comparison of pre-exposure and post-exposure phonetic distance for each type of vowel by modality for female subjects in the F1 measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

F2:

For female participants alone, we saw no significant overall finding of convergence on F2. Register effects were evident only for English-like vowels. As for modality, female participants significantly diverged in the auditory modality for foreign vowels, and showed no significant results of convergence or divergence in the audiovisual modality.

		Overall	Child-Directed	Adult-Directed	Auditory	Audiovisual
F2	English-like	×	✓	×	×	×
	Foreign	×	×	×	D	×

Table 5.40. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F2 in female participants, and overall. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for duration for the female participants is presented in Table 5.40. Table 5.40 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effect of Subject [Subject: $\chi^2(1) = 5.99, p = 0.01$] but not Word [$\chi^2(1) = 0.00, p = 0.97$] significantly reduced model fit.

<i>F2 Females</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Pseudo-R^2</i>
	<i>Estimates</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c^2</i>	<i>df</i>	<i>p</i>	
Register	-0.16	0.15	-1.05	0.292	6.36	5	0.273	-0.0001
Modality	-0.03	0.14	-0.23	0.818	2.98	5	0.703	0.0000
Carrier Type	-0.03	0.08	-0.46	0.644	0.28	2	0.871	0.0000
Vowel Type	-0.18	0.08	-2.38	0.017*	14.01	4	0.007*	0.0001
AQ	-0.01	0.01	-1.24	0.215	701.06	3	0.000	-0.0003
Register X Modality	0.10	0.09	1.05	0.292	1.13	1	0.287	0.0000
Register X Carrier Type	-0.01	0.07	-0.18	0.861	0.03	1	0.855	0.0000
Modality X Carrier Type	0.04	0.07	0.52	0.605	0.27	1	0.604	0.0000
Register X Vowel Type	0.01	0.07	0.21	0.830	0.05	1	0.825	0.0000
Modality X Vowel Type	0.07	0.07	1.07	0.287	1.14	1	0.286	0.0000
Carrier Type X Vowel Type	0.03	0.07	0.49	0.623	0.24	1	0.623	0.0000
Register X AQ	0.01	0.01	1.51	0.130	2.47	1	0.116	0.0000
Modality X AQ	0.00	0.01	-0.45	0.653	0.07	1	0.788	0.0000

Table 5.41. Summary of results from the mixed effects model for F2 in pre-exposure compared to post-exposure productions for female participants.

Overall, there was a significant main effect of Vowel Type but no significant interactions. Post-hoc testing revealed that the difference between imitation of English-like and foreign vowels was significant ($p = 0.005$), favoring more imitation for English-like vowels than foreign vowels (recall, in the analysis of convergence, we found that for foreign vowels in the auditory domain, there was significant divergence). This difference is shown visually in Figure 5.35.

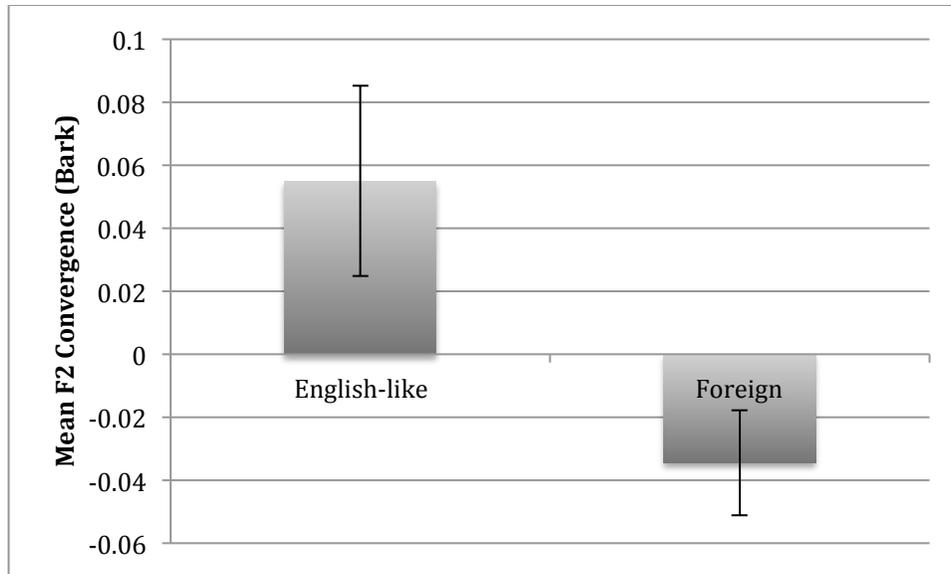


Figure 5.35. Comparison of pre-exposure and post-exposure phonetic distance for each type of vowel by for female subjects in the F2 measurement.

F3:

Recall that subjects did not converge with the model talker on F3, so we will not look any further into this variable.

Summary: Individual Phonetic Measures

For the analysis of the individual phonetic measures, we once again saw an advantage of the child-directed register and the audiovisual modality, but for these measures, the findings were strongly tied to vowel type (all measures were, however, in the expected direction). For F1, we saw an advantage of the child-directed register and audiovisual modality only on foreign

vowels (results of greater amounts of convergence), and for F2, we saw a modality advantage of the audiovisual modality on foreign vowels (there were only findings of convergence on the audiovisual and not auditory modality for this vowel type). We hypothesized that the child-directed register and the audiovisual modality could help improve subjects' imitation of foreign vowels, and these findings align with that hypothesis. However, at first glance there was one finding that was contrary to this hypothesis, which was for F2. We found that overall on F2, there was convergence in the child-directed register but not in the adult directed register, but only for English-like vowels. There was also significantly more convergence to F2 to English-like vowels than foreign vowels. We speculate that this result is likely due to increased convergence to F2 on the /u/ vowel. Recall in the pre-test/post-exposure comparison, we saw a large amount of convergence in F2 to /u/, attributed to the sociophonetic status of /u/ in dialects of American English. We once again see that there is a large amount of convergence for English-like vowels in this same measure, and it appears this is once again driving the results. The child-directed register likely serves to make this measure more salient, allowing for even better convergence on F2 of /u/.

Finally, we wanted to relate the results from these individual phonetic measures to the results we found favoring imitation in the auditory modality for English-like vowels, shown in the global phonetic measure not including F3. Looking at the data for F1, it appears that, although it is not significant, there is more convergence in the auditory modality for F1. This is likely what is driving that difference for the Euclidean Distance measure, but closer analysis would be able to determine definitively whether this is the case.

Summary: Female Participants

Overall, female participants showed a number of significant convergence findings in the pre-exposure/post-exposure comparison. These results also point to an advantage of child-directed speech, and of the audiovisual modality, on imitation. However, they are not comprehensive, and there are many subgroups in which we see no evidence of these effects.

Register Effects:

The global measures, the global phonetic measures, and the individual phonetic measures all provided evidence for increased implicit imitation in the child-directed register compared to the adult-directed register. In the overall convergence measures, there were clear effects of register. For nearly all of the measures, the female participants converged on child-directed speech (missing was f0 for English-like vowels, but female participants diverged on f0 for both vowel types in the adult-directed register, and f2 for foreign vowels). They only converged to adult-directed speech for duration of English-like vowels and F1 for both vowel types (although for F1, they converged on all measures). However, for both these measures there was evidence in the relative levels of convergence to support an advantageous effect of the child-directed register. For duration, there was a register effect supporting increased imitation for child-directed speech, but only for the audiovisual modality. For F1, there was still support for a register effect for foreign vowels, in which female subjects showed greater levels of convergence in the child-directed register. This effect was robustly seen in all types of measures.

Overall, we see clear, and compared to other analyses, *comprehensive* register effects, not limited to a particular vowel type or measure. It seems also that there is some evidence that the register differences are maximized in the audiovisual modality. Further analysis would have to

analyze specifically what characteristics in the visual input are aiding this combinatory advantage.

Modality Effects:

The modality effects were less robust than the register effects for the female participants. The overall convergence measures showed significant convergence findings for the audiovisual compared to the auditory modality for foreign vowels in both the global phonetic measures, for English-like vowels in the duration measure, for foreign vowels in the f_0 measure (in which there was also significant divergence for the auditory modality) and inversely for foreign vowels in the F2 measure (significant divergence in the auditory modality, no effect in the audiovisual modality). For all other measures except English-like vowels for the Euclidean Distance (F1+F2) measure, there was either convergence in both modalities or a lack of convergence in either modality. For English-like vowels for the Euclidean Distance (F1+F2) measure, there was only convergence in the auditory modality. In addition to the convergence differences, there was one significant result in regards to relative amount of convergence: there was significantly more convergence on duration in the audiovisual modality when considering just CVC syllables.

Overall, the modality effects for female participants are not comprehensive or robust, but there is some evidence for this effect, and one finding of evidence against it.

Vowel Type Effects:

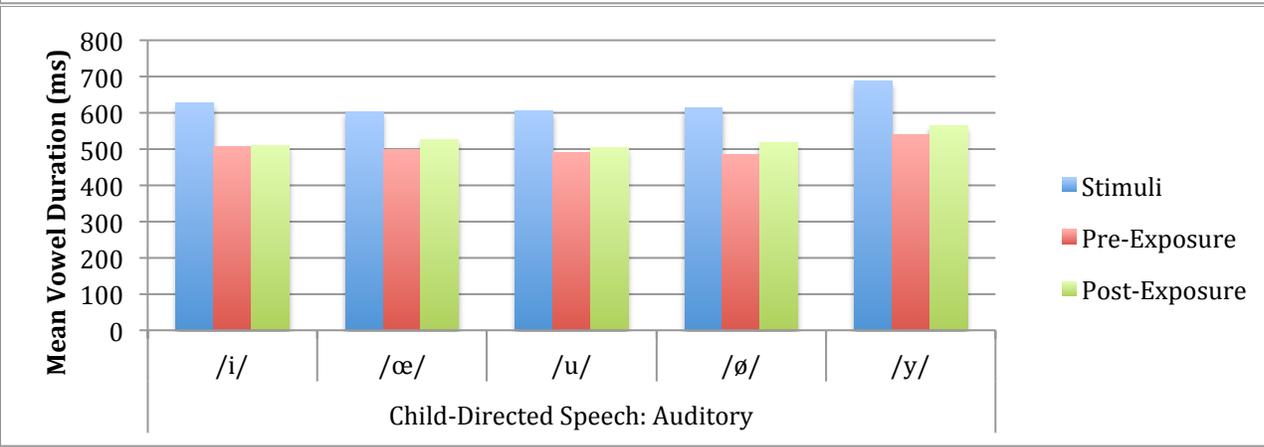
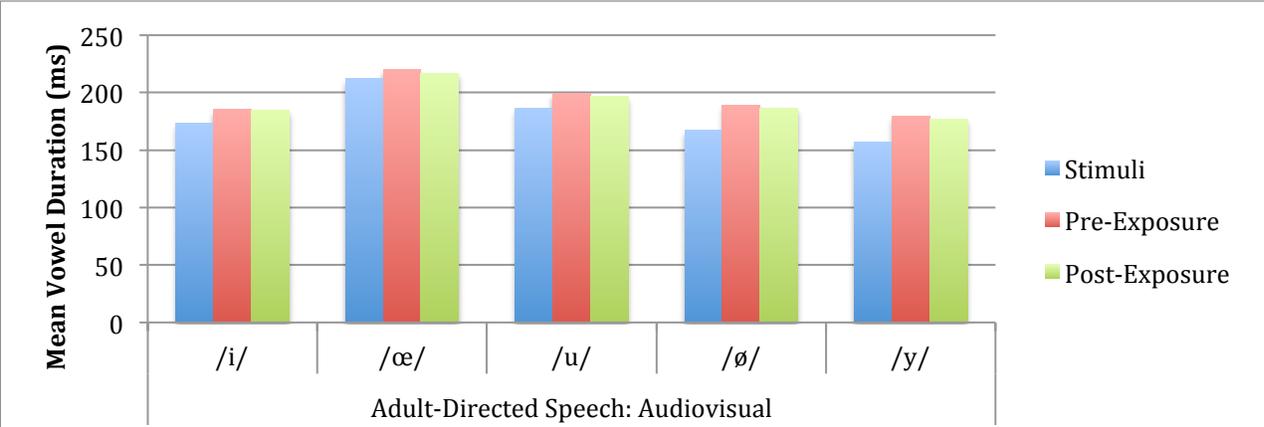
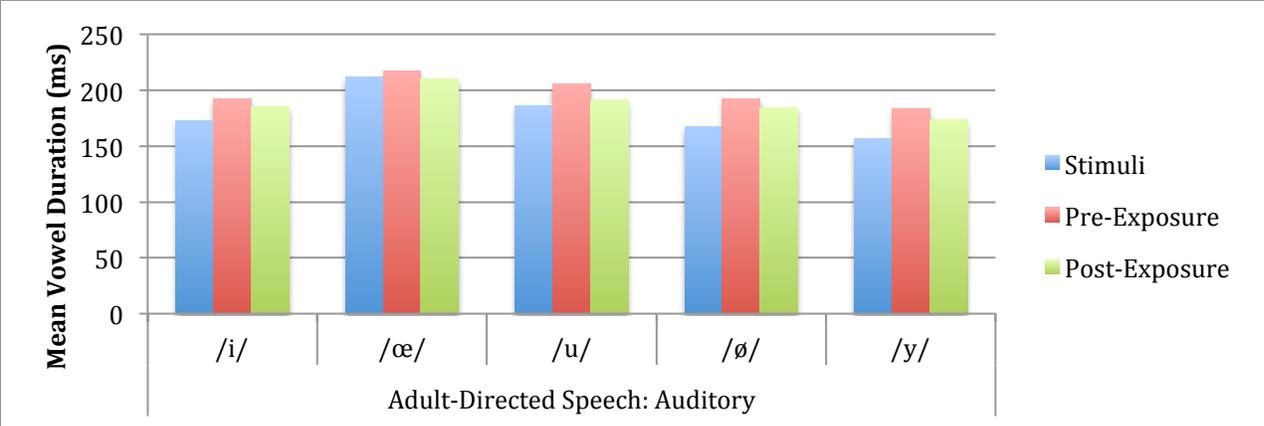
The vowel type effects for the female participants were not much clearer than the vowel type effects shown for the male participants. While they were still not comprehensive, they were more slightly more systematic than for the male participants, favoring an advantage of the visual

modality and the child-directed register for foreign vowels only. Significant convergence findings favoring an advantage of a register *or* modality effect only in foreign vowels was shown for the measures of f0 (register and modality), duration (register), Euclidean Distance (F1+F2+F3) (register), F1 (register and modality), and F2 (modality). However, there were only two instances that showed convergence effects for English-like and not foreign vowels: duration (modality) and F2 (register). Overall, while there is some evidence for increased effects of register and modality for foreign vowels, they are not widespread enough to draw definitive conclusions.

Discussion of male and female participants

Direction of Convergence:

The graphs below summarize the results for production in each of our experimental measures according to register and modality, separated by vowel. In each of these graphs, rather than indicating the *amount* of convergence (as in the graphs previously shown), we plot mean values in the pre-exposure and post-exposure productions.



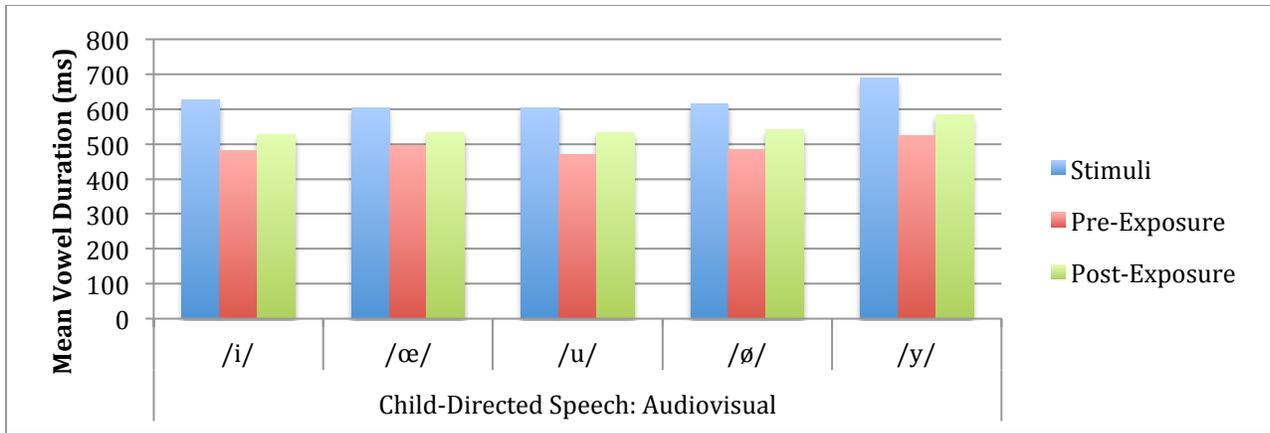
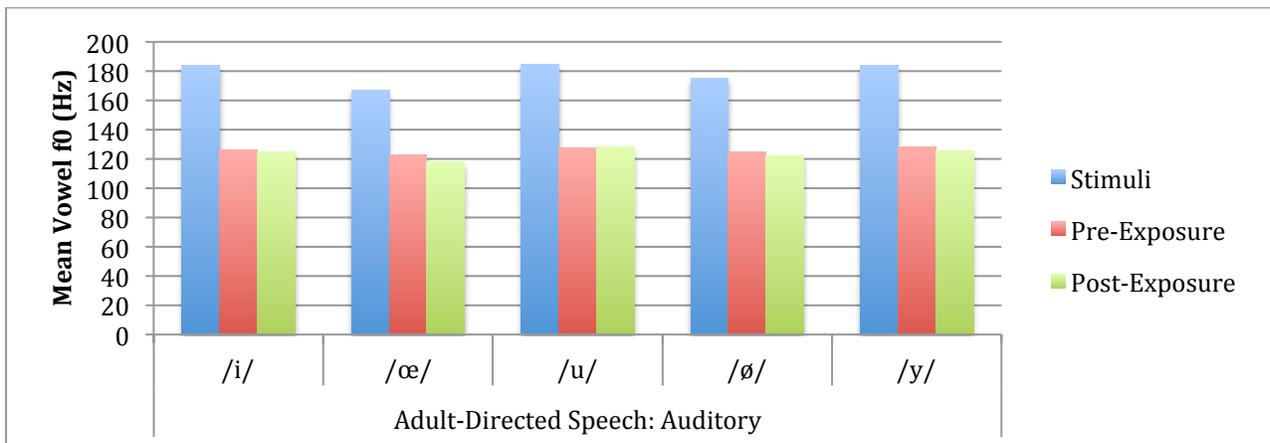


Figure 5.36. Mean duration of vowels produced by subjects in the pre-exposure and post-exposure phases, compared with the mean duration of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.

In the duration graphs above, we can see that in both of the adult-directed speech conditions, subjects shortened the durations of their productions to converge with the model talker. For both of the child-directed speech conditions, we saw the opposite effect: subjects had to lengthen their durations to converge with the longer duration of the model talkers' productions.



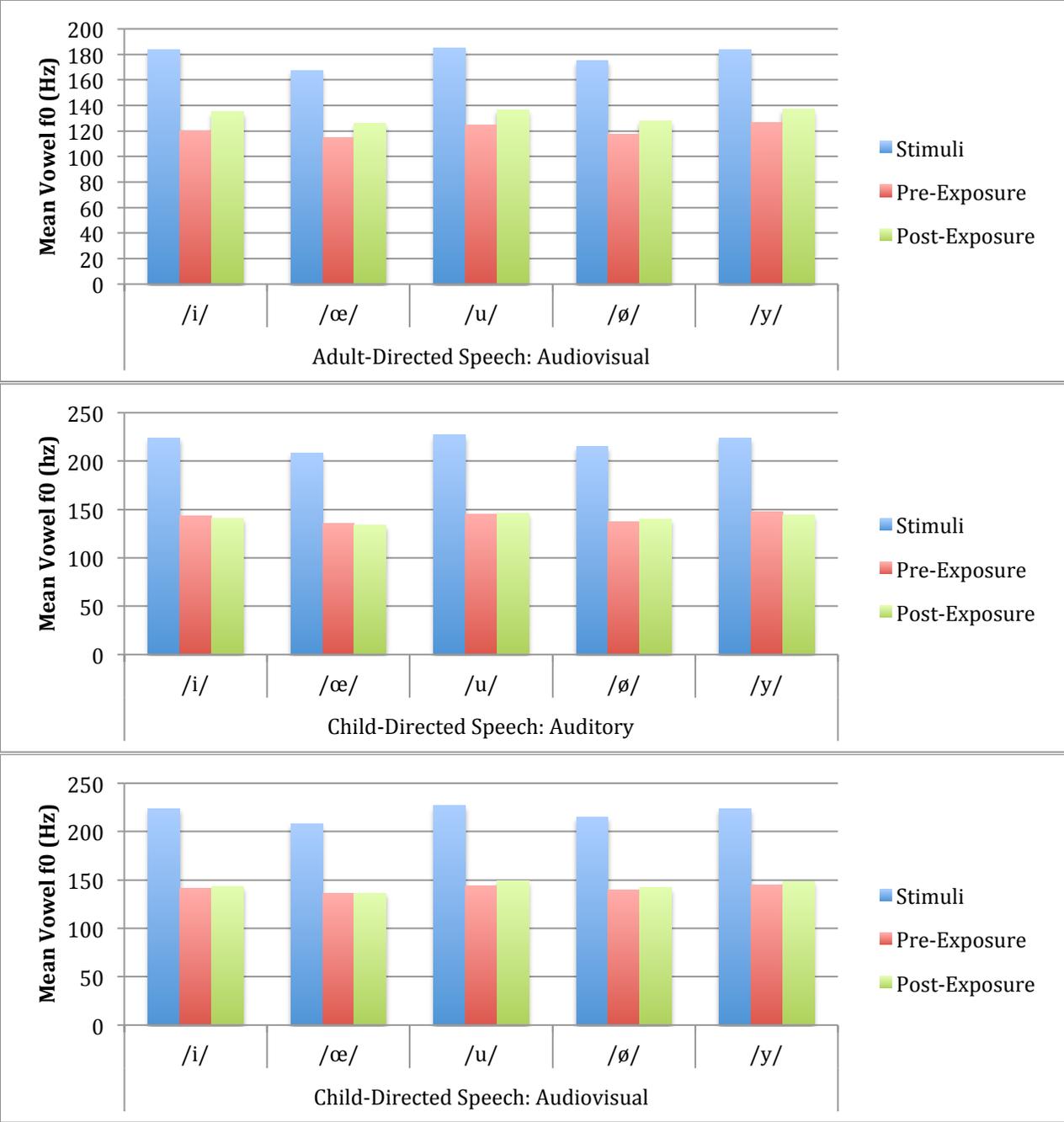
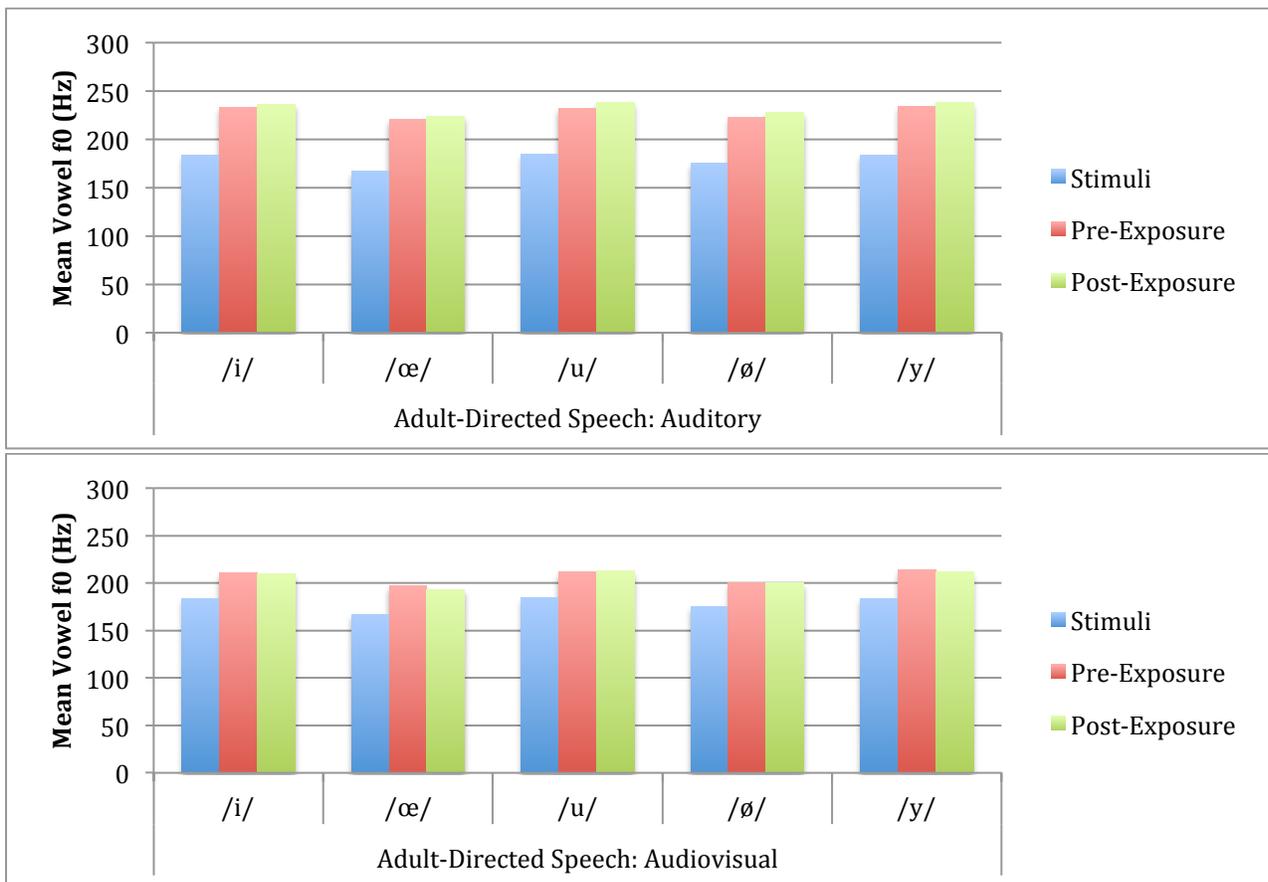


Figure 5.37. Mean f0 of vowels produced by male subjects in in the pre-exposure and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.

Male subjects showed no clear consistent pattern across the different conditions for how they adjusted their f0. The model talker's f0 values were considerably higher than the average for the males in any of the conditions. The only clear case of convergence was for the adult-directed speech:audiovisual condition. Subjects in this condition clearly raised their f0 to converge to the model talker. Slightly less clear, but mostly in the correct direction was the child-directed speech:audiovisual condition. In the other conditions, there were no clear patterns for how subjects adjusted their productions in response to the model talker.



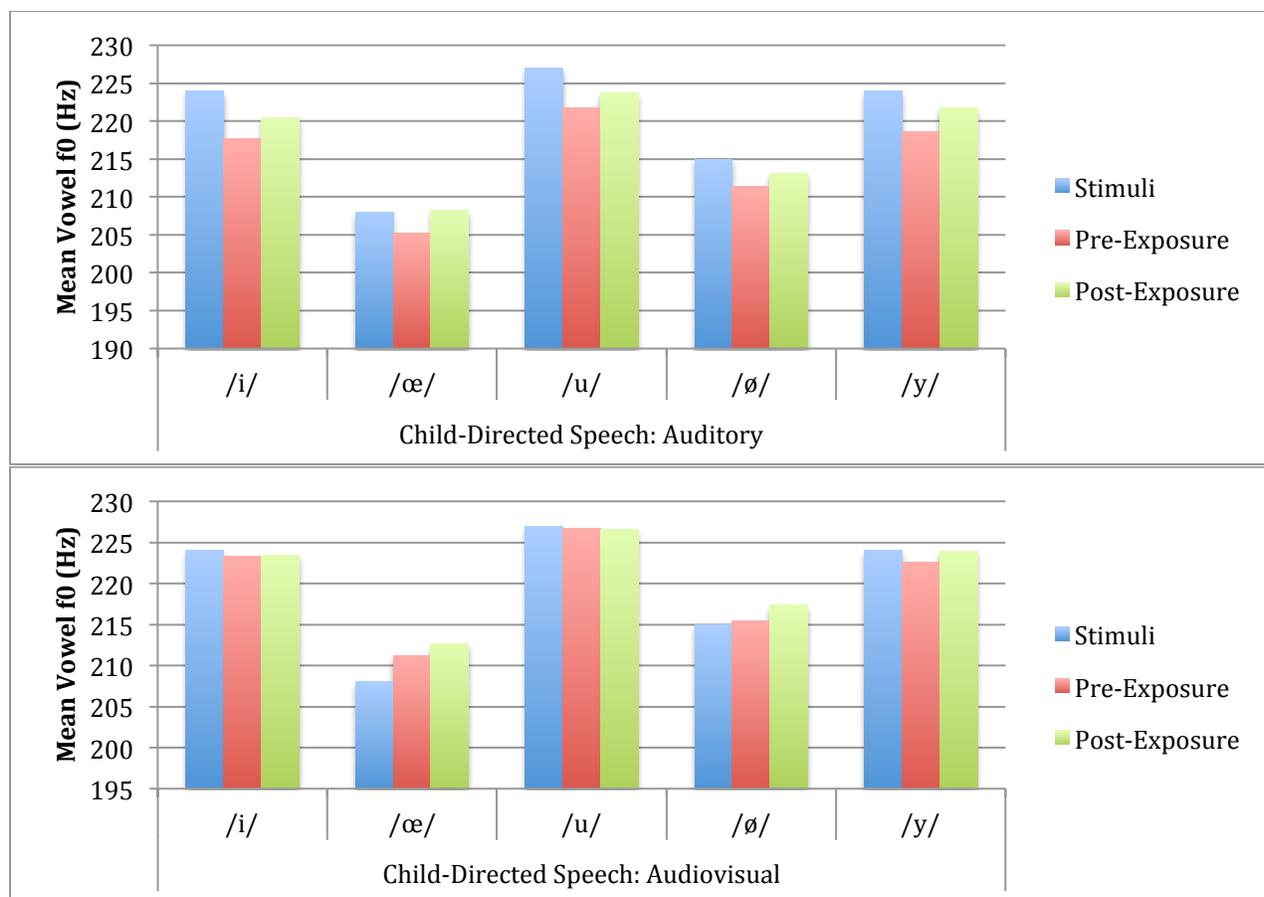
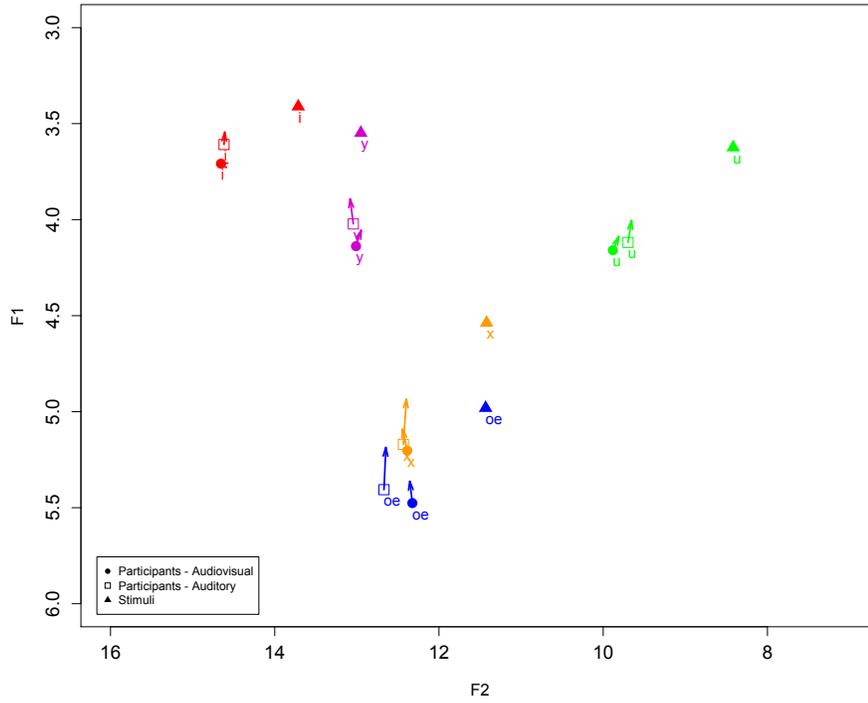


Figure 5.38. Mean f0 of vowels produced by female subjects in the pre-exposure and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.

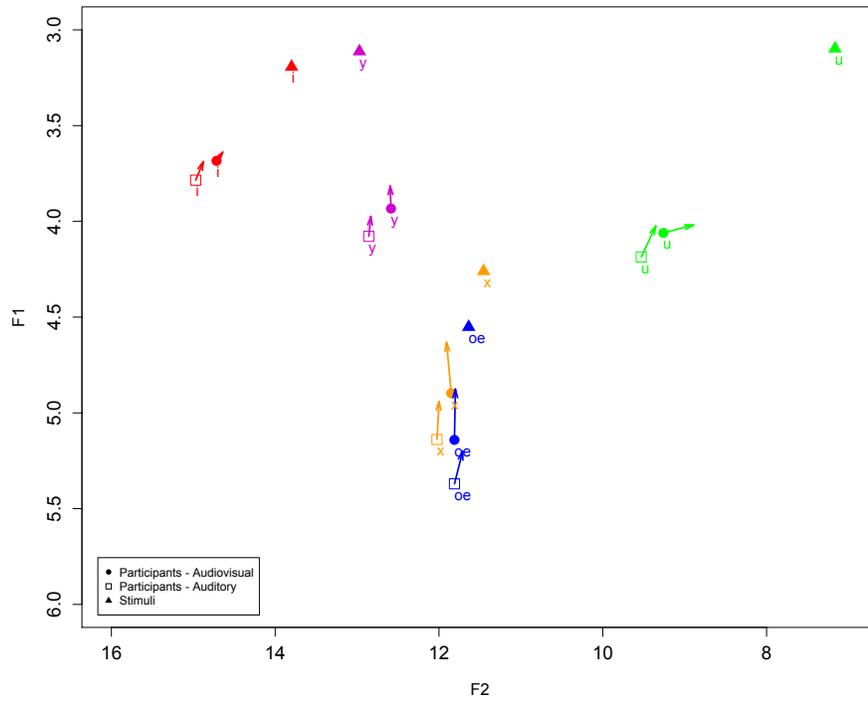
The graphs above show the direction of convergence to f0 for female subjects. Starting with the adult-directed speech condition, we can see different patterns for the two modalities. In the auditory modality, the female subjects raised their f0 in response to the model talker's productions. This is the same pattern as seen for females in the pretest/post-exposure comparison; it seems as though subjects in this condition could have been normalizing what they hear and raising their f0 in response to hearing a male with a high f0. For the audiovisual condition, however, for all vowels except /u/, female subjects appear to be lowering their f0,

converging with the model talker's raw f_0 . This is an interesting finding because we could imagine that with the addition of the visual cues, subjects would be more aware of the speaker's gender, and in this situation we might expect more normalization. For the child-directed speech:auditory condition, the female subjects almost uniformly raised their f_0 , converging to the talker's articulations. For the child-directed speech:audiovisual condition, we saw a difference according to vowel type; for the English-like vowels, we saw little change to f_0 , and this was a lowering of f_0 values (diverging with the talker). For the foreign vowels, however, the female subjects raised their f_0 across the experiment, and for all vowels but /y/, this resulted in divergence from the model talker (which could indicate possible normalization).

F1-F2 Imitation: Adult-Directed Speech



F1-F2 Imitation: Child-Directed Speech



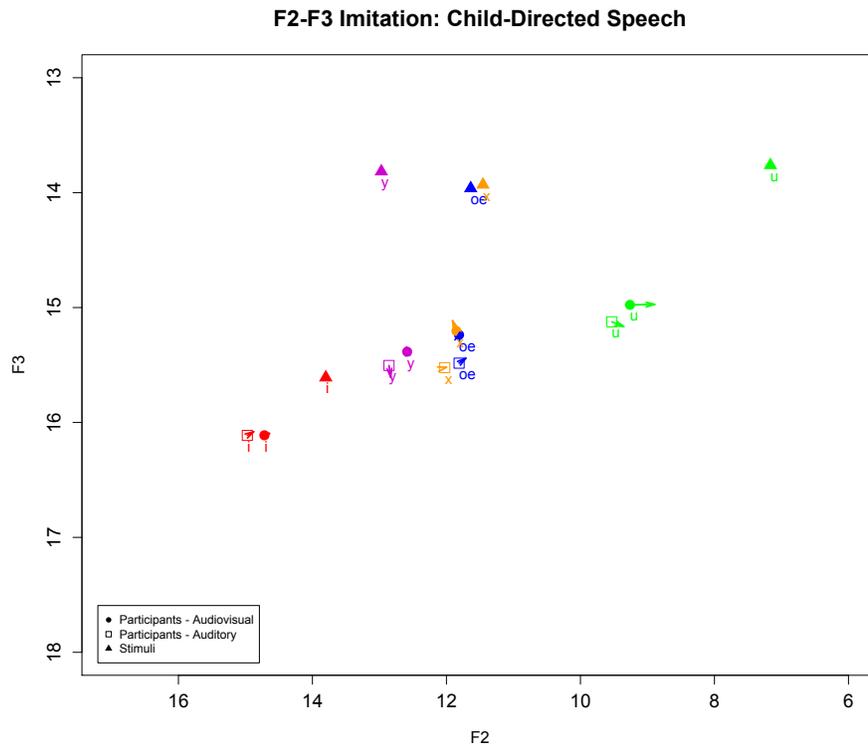
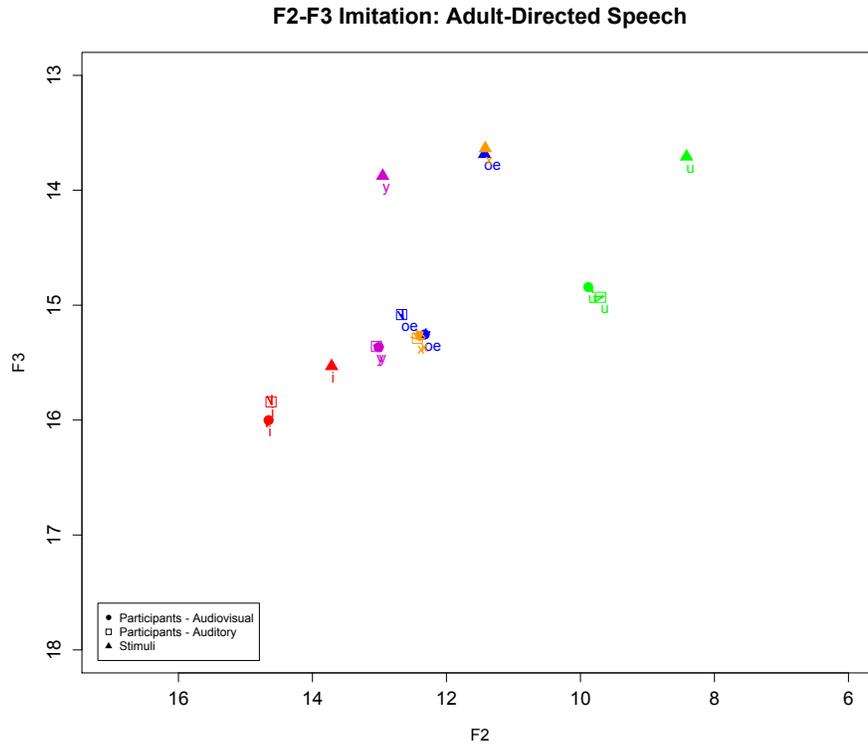


Figure 5.39. Formant plots of vowels produced by subjects in in the pre-exposure and post-exposure phases, compared with the mean formant values of the stimuli, separated by vowel, modality, and register.

Finally, looking at the direction of convergence on the phonetic measures for the vowels in the pre-exposure/post-exposure comparison, in the adult-directed speech condition, we saw that the majority of convergence occurred in F1, with the exception of /u/, in which there was convergence for F2 as well. Participants raised their F1 values to make them more similar to the model talker's F1. F2 modifications for vowels other than /u/ were often incorrect. For the child-directed speech condition, we also saw considerable convergence in subject's raising their F1 consistently in both modalities for all vowels. For this condition, however, we also saw more accurate and exaggerated convergence in F2 for /i/ and /æ/ (/æ/ only for the auditory condition). Once again, however, we saw considerable F2 convergence for /u/. For F3, the changes were so minimal, the direction of the convergence is mostly irrelevant.

Register Effects:

Overall in every measure for females, and every measure but f0 for males, there was more convergence in the child-directed register than the adult-directed register. However, not all of these effects were significant. The result overall for convergence to each register by gender are shown in the graph below.

*

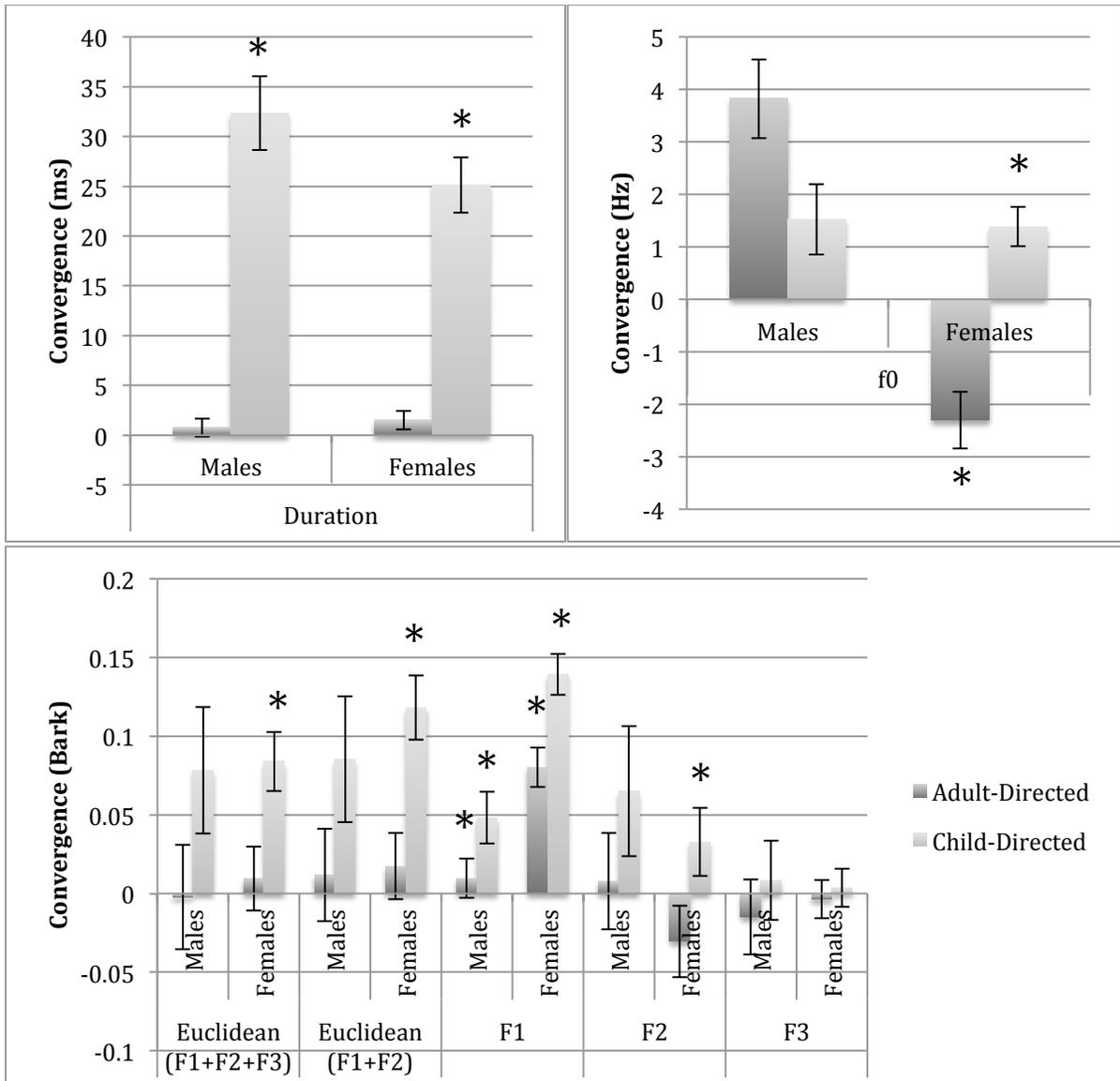


Figure 5.40. Results for convergence according to register for each measure in the pre-exposure/post-exposure comparison for adult participants, separated by gender. The dark bars represent results for the auditory modality and the light bars represent results for the audiovisual modality. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

For both the female and the male data, the register effects were shown in the global measures, the global phonetic measures, and the individual phonetic measures. For females, the

significant results were comprehensive for the global phonetic measures, showing a significant result of convergence in the child-directed register and no significant result for the adult-directed register (the more global phonetic measure captured the register effect better than the phonetic measures individually, leading us to believe that the register advantage is global with regards to the phonetic measures). Also for f_0 , we saw a complete register effect, but this was due to results of divergence in the adult-directed register, rather than just convergence in the child-directed register. For the other global measure of duration and the individual phonetic measures in females, we saw a register effect for at least one vowel type in every measure, but not for both. To reiterate the previous section, for female subjects, for all measures in which there was no effect on the “is there convergence?” measure, there was an effect seen in the relative amounts of convergence related to duration.

For the male data, however, there were no significant register effects for the global phonetic measures, and instead the register effects were more evident in other measures, and not as broadly observed. There was also one result in the unexpected direction, favoring more imitation in the adult-directed register, and that was for f_0 . Males were less likely to imitate f_0 in child-directed speech, even though they were more likely to implicitly imitate other measures in this register.

Overall, in the results comparing the two registers, it seems that while both males and females were likely to be better implicit imitators in the child-directed register, the effects were much more robustly observed for females than for males, and they were most clearly evident for females in the global phonetic measures and f_0 . Future work can confirm this with a statistical comparison of the influence of register on each gender.

Modality Effects:

There were significant modality effects seen in both the male and female participants. Interestingly, however, although not all differences were significant, for every measure for males, and every measure but F1 for females, mean convergence was greater in the audiovisual modality than the auditory modality. The result overall for convergence to each modality by gender are shown in the graph below.

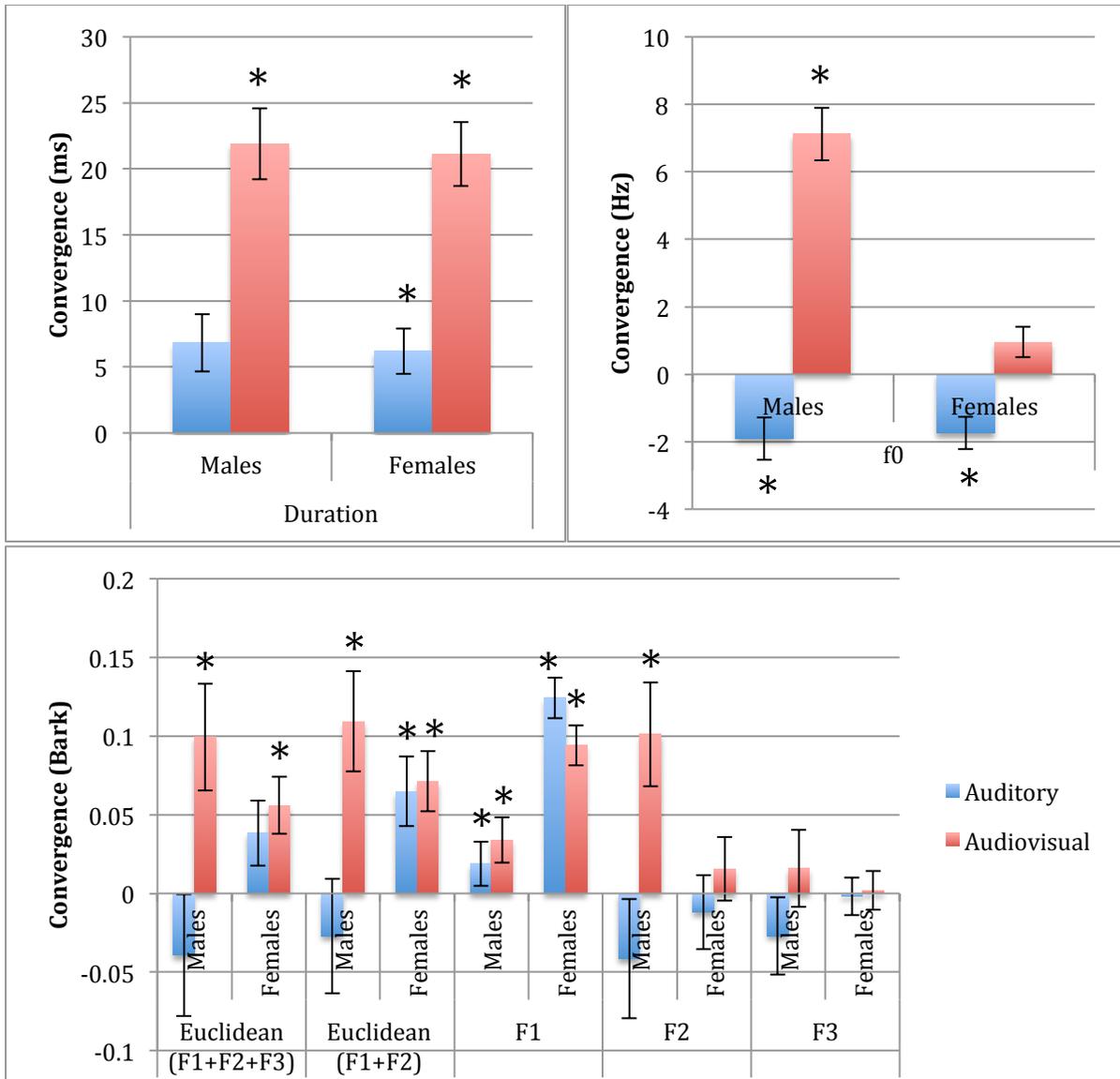


Figure 5.41. Results for convergence according to modality for each measure in the pre-exposure/post-exposure comparison for adult participants, separated by gender. The blue bars represent results for the auditory modality and the red bars represent results for the audiovisual modality. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Once again, like for register, we saw a nearly complete effect of modality when just looking at the mean convergence amounts. For all cases but for F1 in females, there was more

imitation in the audiovisual modality than the auditory modality. However, also like register, these results were not all significant, or significantly different.

The modality effects were shown to be relatively the same degree for males and females, but they were shown in different measures. Unlike for register, there were no clear measures in which we saw convergence to both vowel types for the audiovisual modality and not to the auditory modality. In other words, it appears that the modality differences were more strongly influenced by vowel type. For the global measures, we saw convergence to the audiovisual, but not auditory, modality for English-like vowels in males and foreign vowels in females (we also saw that females significantly converged to English-like vowels in the auditory modality, however). Another measure that was somewhat clear related to modality was f_0 in both males and females. Males significantly converged in both vowel types to the audiovisual, and not auditory, modality, and females showed this pattern for foreign vowels only. In fact, the female results overall for the audiovisual modality, there was a strong tie to foreign vowels; of the five cases in which females showed significant convergence to the audiovisual and not the auditory modality, four of them were for foreign vowels (the final was for duration), leading us to believe that at least for these female participants, there is reason to believe that the visual cues helped to learn the pronunciation of unfamiliar sounds. The female participants showed one overall effect in the opposite direction, with significant convergence in the auditory, and not audiovisual, modality: English-like vowels for the Euclidean Distance (F_1+F_2) measure. However, this seems to be driven by large amounts of imitation for /u/ in F_2 . The contrastive result does not seem to be as related to modality itself.

Did subjects generalize?

As a final step in the analysis, we looked at whether the analysis was affected by the addition of a factor defining generalized words. Recall, only a subset of the words were played during the exposure phase, but the analysis was run on all words. For both male and female participants, this factor did not affect the results for any analysis variable when it was added as a main effect into the mixed effects models. It appears that subjects did generalize their exposure to other words. See Table 5.42 for the statistical output for this variable *only*, as it appeared when added to the mixed effect models analyzing differences in convergence.

	<i>Parameter estimates</i>		<i>Wald's test</i>	
	<i>S.E.</i>	<i>Estimates</i>	<i>Z</i>	<i>p_z</i>
Duration	4.04	4.66	1.15	0.249
F0	0.69	-0.59	-0.85	0.397
F1	0.02	0.01	0.60	0.550
F2	0.04	0.00	0.00	0.999
F3	0.02	0.00	-0.10	0.919
F1F2	0.03	0.01	0.44	0.664
F1F2F3	0.03	0.01	0.36	0.722
Duration	4.27	3.38	0.79	0.429
F0	1.04	-1.19	-1.15	0.251
F1	0.02	0.03	1.41	0.159
F2	0.06	0.03	0.39	0.695
F3	0.04	0.02	0.40	0.690
F1F2	0.06	0.04	0.69	0.492
F1F2F3	0.07	0.05	0.74	0.460

Table 5.42. Output for the factor of whether a word was generalized.

The null result for this factor indicates that subjects did not change in their imitation as a function of whether or not they heard the word in the exposure phase. Exposure to the vowels in only some of the contexts made for consistent imitative behavior in all of the contexts. However, this generalization is only at the word level, as we did not analyze generalization to new vowels or acoustic features.

Place of articulation of the onset consonant

In a separate analysis, we included the factor of Place of Articulation of the Onset consonant because we wanted to ensure that our use of the alveolar consonants /t/ and /d/ did not obscure visual cues that could be used in imitation. Our main interest was in whether these consonants would impact imitation in the visual modality. We only saw one interaction of modality and place of articulation of the onset. This finding revealed that for F3, imitation for females was best in the auditory context for velars and the audiovisual context for alveolars. There were no detrimental effects for using alveolars rather than just glottals and velars.

While we did not see interactions of the place of articulation of the onset with the modality that would have suggested that alveolar consonants obscure the vowels' visual cues, we did see a few effects of place of articulation. Overall, duration and F1 were imitated better by males following a non-alveolar consonant, but for the case of F1, this was only for English-like vowels. In complete contradiction, overall convergence in Euclidean distance (F1+F2+F3) for English-like vowels was *better* for females in the post-alveolar context. There was also one interaction of place of articulation of the onset consonant and context type, and that showed that for the measure of Euclidean distance (F1+F2), female participants were worse at imitating CV

syllables when the onset was an alveolar, or CVC syllables when the onset was velar. The statistical output of all of these models is shown in the appendix. While these findings are interesting, since they are not significant to the overall research questions, they will not be discussed further.

Discussion and comparison between pre-test/post-exposure results and pre-exposure/post-exposure results

In this final section, we will look at the results from both of the analyses completed for this chapter. We will look at both the comparison of pre-test and post-exposure productions, and the comparison of pre-exposure and post-exposure productions, in order to make generalizations about the imitative behavior of adult subjects in this study. Specifically, we will evaluate whether the predictions and hypotheses for this study held for our experimental results. We will begin by reviewing these hypotheses, then we will look at whether they held in our results.

Review of predictions

In this experiment we had two main goals. First, we wanted to determine if the differences between audiovisual and auditory-only imitation could be shown as quantifiable in any particular acoustic measures. We were interested in whether imitation differences could be measured, in order to determine the exact impact of the visual modality on imitation.

Second, we wanted to test how imitation differed based on situational factors such as the register of speech presentation, the modality presented, and the familiarity of the vowel. While we had no specific predictions about how adults would imitate vowels based on their familiarity, we did have predictions about the modality and register effects:

- Audiovisual speech was expected to facilitate convergence. Previous research using perceptual analysis measures showed increased rates for convergence when subjects were exposed to the audiovisual rather than auditory domain.

Prediction = Supported

- Audiovisual speech was expected to facilitate convergence of visually salient cues.
 - In particular, a measure relating to vowel rounding – F3, which is visually salient, was expected to show increased convergence in the audiovisual domain, relative to cues for vowel height (F1) and backness (F2).

Prediction = Not supported

- Child-directed speech was expected to facilitate convergence, as it provides longer durations of vowel exposure, and it is a learning register.

Prediction = Supported

In addition to the above bullets, we also had a prediction about imitation of the two English-like vowels /i/ and /u/. Since the pronunciation in /u/ in English allows for more stylistic variation, results from previous convergence studies indicate that participants are likely to show more convergence for /u/ compared to /i/. This prediction was confirmed in our results.

One last question we were interested in is whether advantages in imitation by experimental conditions are motivated by the type of measure being analyzed (i.e. global measures vs. aggregate phonetic measures vs. individual phonetic measures). What specifically are the different cues adding to overall measures of convergence?

Did audiovisual speech facilitate convergence?

Exposure to the audiovisual modality resulted in greater convergence for both the pre-test/post-exposure comparison as well as for the pre-exposure/post-exposure comparison. There were noticeable differences, however, for these results. For the pre-test/post-exposure comparison (the more absolute measure of convergence) we found significant results related to modality for the global measures and aggregate phonetic measures. For the pre-exposure/post-exposure comparison, we saw significant effects of modality on all measures, but these results were not as clear and complete as for the previous result.

Looking at the results according to vowel type, there was also some evidence, for female participants in the pre-exposure/post-exposure comparison that modality interacted with vowel type (there could be no effects of vowel familiarity for the pre-test/post-exposure comparison). Females only showed an audiovisual modality advantage in the aggregate phonetic measures and f_0 for *foreign* vowels. However, there was also evidence that males show the opposite pattern in the global phonetic measures: imitation for English-like vowels in the audiovisual modality. As the results here are inconclusive, we can conclude that while there is some evidence of this interaction, it is not enough to make a statement on the visual modality with regards to learning a new sound, rather than modifying a familiar sound. Additionally, considering these results in context with studies showing increased convergence in the audiovisual modality (Dias & Rosenblum, 2011), we are posed with an interesting question. While we saw some effects of modality, they were not nearly comprehensive, so why are increased levels of convergence in the audiovisual modality so robust? Perhaps to model what these studies show using perceptual

measures, we need a more encompassing measure, such as a Euclidean Distance measure including duration and f_0 , taking all aspects of imitation into account.

Did modality effects selectively enhance uptake of particular cues salient in that modality?

We know that the audiovisual modality facilitates convergence, and results in increased levels of convergence, measureable at a perceptual level (Dias & Rosenblum, 2011). What is not known, however, is what aspects of the speech are imitated better. Based on previous studies looking at audiovisual speech integration, we thought perhaps there could be influences on formant frequencies (better imitation of visually-salient formant cues in the audiovisual modality, namely F3). These studies hypothesize not just that there are some cues perceived better in the visual modality, but also that there are cues perceived better in the auditory modality, namely F1. These studies led us to hypothesize that F1 would be best perceived in the auditory modality, and F3 would best be perceived in the audiovisual modality. While we found no overall results for F3, we did find a couple of significant findings favoring F1 imitation in the auditory modality. In the pre-test/post-exposure comparison, we saw findings of convergence to F1 in the auditory, but not audiovisual, modality. In the pre-exposure/post-exposure comparison, we also saw that female subjects imitated F1 significantly better in the auditory modality, though only for English-like vowels. Overall, these F1 results suggest that F1 is better imitated in the auditory modality, which is in line with our hypothesis.

Did child-directed speech facilitate convergence?

We made the prediction that the child-directed speech register would allow for increased imitation in our study. Nielsen's (2011b) study on speech imitation in children showed increased imitation for children compared to adults, and one suggestion she had for this result was that it was due to use of a child-directed speech register. The child-directed speech register is both slower, which can allow for increased exposure to the stimuli, and more exaggerated, and shares common characteristics with speech-to-foreigners and "clear speech", which are learning registers. Due to this, we hypothesized that child-directed speech registers would facilitate imitation. Perhaps the strongest result that we saw within the study was the result for speaking register.

Overall, child-directed speech increased convergence. Given the larger original distance between the child-directed pronunciations and the adults' natural pronunciations, it was expected that convergence levels could be greater, for at least the cues that are known to be maximally different: the global measures. However, for both the pre-test/post-exposure analysis and the pre-exposure/post-exposure analysis, we found effects of convergence in both the aggregate phonetic measures in addition to the global measures. For the pre-exposure/post-exposure comparison, we found additional effects of register, favoring child-directed speech, in the individual phonetic measures (which is not surprising given the significant results for the global phonetic measures). Overall, these results point to an advantage of the child-directed register. In other words, the longer durations and more exaggerated pronunciations helped when male and female participants were not familiar with the sound they were imitating.

Noticeable in the pre-exposure/post-exposure comparison was the difference in the register effect for male and female participants. While female participants showed a register effect on nearly every measure, the results for male participants were fewer. There was also one result in favor of imitation in the adult-directed, and not child-directed register. This was for f0 in male participants in the pre-exposure/post-exposure comparison. For whatever reason, males were not inclined to imitate the pitch of the speaker when the speaker was producing child-directed speech. Additionally, for female participants there was an interaction between register and modality where the effects of the audiovisual modality were only significant within the child-directed register.

There was evidence in both comparisons and both genders that child-directed speech facilitated convergence. The results of this study cannot answer whether this was due to simply the longer durations or because it is specifically child-directed speech; the stronger influence of register on females over males seems to imply that it is not just longer durations. There is no reason to suspect that longer durations would aid females more than males, but there are social reasons why females may be more likely to imitate child-directed speech. A similar study on clear speech, with longer durations, but not the “child-directed” social status, is needed in order to disambiguate these two explanations.

Finally, one additional hypothesis regarding child-directed speech was that the child-directed register would minimize modality effects on individual phonetic measures, since there are different features that are maximized visually in child-directed speech (Green et al., 2010). However, there was no evidence to support this hypothesis, as increased imitation in the child-directed register was found across nearly all measures.

Did subjects imitate /i/ and /u/ differently?

The predictions for this study were that /u/ would be more likely to show imitation in this experiment because /u/ shows more sociolinguistic F2 variation, and factors which show sociolinguistic variation tend to allow more imitation (Babel, 2010). Note we did not compare the two English-like vowels in the pre-exposure phase, so we are only comparing pre-test and post-exposure productions to answer this question. The results of the current study show that for overall imitation, the vowel /u/ was imitated more closely, but only for the formant that shows the sociolinguistic variation (F2). We saw this difference between F2 for /i/ and /u/ both in the overall convergence measure, as well as in the measure comparing the degree of imitation.

As an interesting side note, we also observed a significant interaction between F2 vowel imitation and AQ score: subjects with higher AQ scores showed less F2 imitation for /u/. This is likely also due to the social nature of F2 imitation, but further analysis would need to look more closely into this question.

Chapter 6

Effects of modality and register on imitation by children

Chapter 5 established that register and modality affect how English-learning adults imitate English-like and foreign sounds. However, it is well established that children and adults learn language in drastically different ways, and at different rates (Penfield & Roberts, 1959, Lenneberg, 1967). In addition to differences in neural and behavioral plasticity, there is also a well-documented developmental difference between children and adults in their reliance on the visual cues to speech. There is some evidence that children rely less on the visual cues of speech: evident in the McGurk effect (McGurk & MacDonald, 1976, Massaro 1984, Massaro et al. 1986, Desjardins et al. 1997) and in studies on lipreading (Massaro, 1987).

These findings bring up a set of questions. Do children just pay less attention to the visual cues at this point in development? Or, if the task at hand could be eased by using the visual cues, will they recognize this and devote greater attention to this aspect of the input? In this chapter, we test how children aged 4 to 6 perform on the same tasks as in Chapter 5. Four- to 6- year olds are at a critical point in the integration of visual cues: they have just entered the age at which they integrate visual cues less often, and they are also at a period in which cross-linguistically, children are diverging in the use of visual cues, developing their adult visual speech systems. Given that English-speaking children have been shown to integrate audio and visual speech cues to a lesser extent than adults, it is possible that they will be less influenced by visual cues in an imitation task. The expectation would then be that unlike adults, children may not derive any benefit from audiovisual as opposed to auditory exposure. Finally, these experiments will test

whether child-directed speech could facilitate imitation either by increasing children's attention or by maximizing differences.

The preceding chapter established that adult subjects' imitation of speech can be affected by the modality of exposure and the register of exposure. This chapter will look at whether this holds for child subjects as well. While imitative behaviors have been frequently studied with regards to social behavior and for speech behavior in young infants, little work has looked at imitation of speech by children. Additionally, although children's decreased reliance on the visual cues is well documented within perception literature, no research has evaluated whether children will use visual cues in imitation tasks. This chapter will seek to address these gaps.

Methods

Subjects

The subjects were monolingual English-speaking children whose parents all reported in a language questionnaire that their child has had no extensive exposure to a language other than English, and no familiarity with French, and also that they have no history of hearing problems. Subjects were recruited from the UCLA Language Acquisition Lab. Group A-A (auditory exposure, adult-directed register) consists of 9 subjects (females = 6; mean age: 4.67 years; range: 4.07-5.69 years). Group AV-A (audiovisual exposure, adult-directed register) consists of 12 subjects (females = 8; mean age: 4.97 years; range: 4.15-5.97 years). Group A-C (auditory exposure, child-directed register) consists of only 4 subjects (females = 2; mean age: 4.50 years; range: 4.12-5.14 years). We tested an additional 28 subjects, but their data was not usable due to

fussing out ($n = 4$), prior exposure to other languages ($n = 6$), and for methodological issues with the recording equipment ($n = 18$). Due to all of the additional subjects that were tested, but not useful, we could not complete enough testing to run the audiovisual exposure, child-directed register condition, and we will save that for future testing.

Speaker

The speaker was the same as for the adult experiments described in Ch. 4.

Stimuli

The stimuli were the same as for the adult experiments described in Ch. 4.

Procedure

The procedure for this experiment was nearly the same as for the adult experiments described in Ch. 4, with a couple of exceptions. The first of these differences is that the experiment was controlled by a research assistant, rather than by the subject. The research assistant led the subject through the experiment, asking the child to tell her parent what word she heard. This was done because not all children were familiar with computer controls, and this also ensured that the kids completed the experiment appropriately. Additionally, between the different phases of the experiment, the children received a sticker to note their completion of

each phase. We did this in order to ensure children's willingness to participate and focus on the study.

An additional difference between the procedure for the child subjects and for the adult subjects is that, although we attempted to select pretest words that children would know, there were a number of child subjects who did not know the words, and they had to have the research assistant alert them to what the words were (particularly for 'coop', 'geese', and 'dude'). In this sense, their pre-test words were not true pronunciations since they were essentially shadowing the research assistant. We did not analyze how this may have affected our results.

Analysis and coding

The analysis and coding for this experiment was identical to Chapter 5, with a few exceptions. One of these differences was that we had one fewer analysis variable, and one fewer interaction for the current study compared to the study in Chapter 5. Although there have been recent developments within the realm of AQ testing on children, we did not perform any AQ testing of the children in our study. Therefore, we have no AQ main effect or interactions for this experiment. Additionally, since we only collected data in the auditory, not audiovisual, modality for child subjects in the child-directed register thus far, we could not include the interaction of Register X Modality in our analysis.

In the acoustic analysis of the data, there were also a couple differences between Chapter 5 and the present experiment. The child data showed much more variation in the data analysis than for the adult subjects, and the analysis program was less reliable in tracking the formants appropriately (there were many cases in which the program was tracking the wrong formant, i.e.,

the second formant instead of the first). We found that for the child data, the Praat algorithm (Boersma & Weenink, 2013) was much better at tracking the child formant data, and so we used this algorithm to calculate formant frequencies rather than the Snack Sound Toolkit (Sjölander, 2004) used on the adult data, and the Snack Sound Toolkit was only used in case of discrepancies. The STRAIGHT algorithm (Kawahara et al. 1999) was still the algorithm used to calculate f_0 values. In order to ensure that our measurements were accurate, we compared our measurements with other studies looking at child formant values (Vorperian & Kent, 2007, Hasek, Singh, & Murry, 1980, Busby, & Plant, 1995). In cases where the measurement was outside the range of formant values given by previous studies, we looked at the measurements of both algorithms (and all formants) and selected the value that represented the correct formant tracking.

There were also two behaviors to note that we did not look at in our analysis. The first was that child subjects directed attention to the computer screen far less in the exposure phase than adult subjects. To get a more accurate representation of the role of the visual cues in this exposure phase, we should perform calculations of the time children looked at the screen and were actually exposed to the visual cues. While we have this data from our video recordings, we did not look at this in the analysis. The second behavior that we did not analyze was substitution behavior by children. There were many cases in which the child subjects did not attempt to imitate the word that they heard; rather, they substituted the heard vowel for a similar vowel in their language (usually /ɪ/, /ɛ/, or schwa). However, there is no way to confirm exactly which instances were “substitutions” versus poor imitations, and thus we did not analyze these tokens separately.

Finally, unlike for the adult data in Chapter 5, we will not analyze whether the place of articulation of the onset consonant affected imitation, or whether child subjects generalized their imitation across all words rather than just the words they heard in exposure. Although these are interesting questions, we wanted the focus of this to be on the other factors within the experiment, as these questions are not relevant until we have a more baseline understanding of child imitation.

Results

For the first subsection below, we will compare the pre-test results to the post-exposure task results, in order to analyze how far subjects modified their natural pronunciations of the English-like vowels for the experiment.¹⁹ This is, in a sense, our only measure looking at absolute “convergence” because subjects already have an established pronunciation of the English-like vowels, and in this measure, we are testing how the subjects modify their natural pronunciation after exposure. In this section, we analyze how the experimental factors affected convergence on the English-like vowels, first looking at answering the question of “is there convergence?” and then moving on to an analysis of how the experimental factors affect this convergence. We begin with global measures of convergence, and then look at the more fine-tuned acoustic measurements.

In the second analysis section, we will evaluate how the exposure session affected implicit imitation between the two different sessions of the experiment according to the

¹⁹ In order to get a perfect sense of how the subjects modified their natural pronunciations, we would have had to ask subjects to repeat a recording of the pre-test word list after the experiment was over

experimental variables. Here, we are interested in looking at what factors affected imitation between the two conditions. We will separate this data out by gender, looking at male and female performance separately. Like in the previous analysis section, we will first look at the question of “is there convergence?” before we move on to looking at which variations in the experiment affected subject performance.

In both analysis sections, we separate the variables into three subsections: global measures, global phonetic measures, and individual phonetic measures. Global measures included duration and f0, global phonetic measures included the two Euclidean Distance measurements, and individual phonetic measures included the three formants. We separated out the variables this way in order to conceptualize the analysis in terms of types of variables, and to reconcile our results with other studies looking at convergence, especially those using perceptual measures.

Pre-test vs. post-exposure comparison

In this section, we will compare the means of the pretest tokens to the means of the post-exposure tokens for the vowels /i/ and /u/. As in the previous chapter, we are comparing the means for each vowel. Since there are varying predictions about convergence on the English-like vowels, we looked at /i/ and /u/ separately in this portion of the analysis.

Within each subsection of analysis, we report two separate analyses. For the first analysis, the variable for analysis was the difference between the model talker and the participant, in the pre-test and post-exposure exposure condition for each dependent variable [duration, f0, Euclidean Distance (F1+F2), Euclidean Distance (F1+F2+F3), F1, F2, and F3,],

and added a factor of Repetition (pre-test or post-exposure) to differentiate the two productions. We ran mixed effects models for each vowel separately with the fixed effect factor of Repetition, and the random effect of Subject for the entire data set (this allowed each subject to have a unique pronunciation), as well as within subsets of data based on Gender, Register, and Modality. A significant effect of repetition would mean that subjects either converged or diverged between their pre-test and post-exposure productions. This was identical to Chapter 5.

Following the analysis of whether there is convergence, in each subsection, we will also take a more in-depth look at how all of the experimental factors affect convergence. For this analysis, we do complex mixed effect models. The experimental design for this part of the analysis was a 2 (Gender: male or female) X 2 (Modality: audio or visual) X 2 (Register: child-directed or adult directed speech) X 2 (Vowel: /i/ or /u/) factorial design. We also included the random effect of Subject to allow subjects to differ in their overall degree of convergence. For this portion, the dependent variable is the measure of *convergence* for each of the phonetic factors (the difference between the pretest difference and the post-exposure difference, see the coding section above for details). The 7 dependent variables were duration convergence, f0 convergence, F1 convergence, F2 convergence, F3 convergence, Euclidean Distance (F1+F2), and Euclidean Distance (F1+F2+F3). We will provide full models for each variable, including all main effects and two-way interactions, following Harrell (2001), Jaeger & Snider (2013), and Jaeger (2011)²⁰, with the exception of the interaction of Register X Modality, which was not included because there was not data from all four subgroups. All modeling was done using the lme4 package in R (R Development Core Team, 2008). To determine the unique contribution of

²⁰ We report only two-way interactions because a full set of interactions would include over 80 different interactions, making the model too complex for analysis. Additionally, many of the complex 5-, 6-way interactions are difficult to interpret.

each variable, we compared models in a subset relationship using likelihood ratio tests. To confirm significance and to interpret pairwise comparisons, we subjected the data to a repeated measures analysis of variance, using Tukey's HSD test, also in R for all factors.

Note that we could see an influence of a particular factor in the first analysis section and not the second, or vice versa. While the first analysis looks at whether convergence is significantly different from zero, the second analysis takes into account relative degrees of convergence. Subjects may show differences in degree of convergence across conditions, with only one or even neither being significantly different from zero. Then, convergence would not be significantly different from zero in either condition; however, the two conditions could be significantly different from each other. Alternatively, the degree of convergence across conditions might be the same for a particular measure; however, due to within group variation, only one result might be significantly different from zero.

In model summary tables for each effect, we report the parameter estimate, the standard error, and two tests of significance: Wald's Z statistic, which tests whether coefficients are significantly different from zero, given the estimated standard error, as well as the χ^2 over the change in data likelihood, $\Delta(-2\lambda)$, associated with the removal of the factor or interaction from the final model. For the likelihood ratio test for the main factors, we tested with and without the factor *and its interactions*. Degrees of freedom are reported for these tests. Finally, we report the pseudo r-squared as a measure of effect size for that variable. Significant main effects and interactions were identified based on the Wald's Z statistical tests.

Is there convergence between the pre-test and post-exposure productions?

Overall, the output for the mixed effects models for the Repetition factor for child data is given in Table 6.1. Note that each row in the table represents the output for the factor of Repetition (pre-test or post-exposure) in separate mixed effect models for each factor for each subset of the data labeled in the “Variable” column.

	Variable	Vowel	Estimate	Std. Error	t-value	p-value	Significant?
Global Measures	Duration	/i/	-1.84	16.17	-0.11	0.910	
		/u/	25.63	21.87	1.17	0.247	
	f0	/i/	10.66	7.28	1.47	0.149	
		/u/	14.72	6.02	2.45	0.018	*, divergence
Global Phonetic Measures	Euclidean Distance F1+F2+F3	/i/	-0.01	0.13	-0.08	0.938	
		/u/	-0.93	0.15	-6.01	0.000	*
	Euclidean Distance F1+F2	/i/	0.17	0.15	1.09	0.280	
		/u/	-1.14	0.20	-5.81	0.000	*
Individual Phonetic Measures	F1	/i/	0.13	0.12	1.04	0.304	
		/u/	-0.34	0.08	-4.26	0.000	*
	F2	/i/	0.13	0.26	0.52	0.607	
		/u/	-1.10	0.21	-5.35	0.000	*
	F3	/i/	-0.13	0.06	-1.98	0.054	
		/u/	-0.11	0.06	-1.79	0.080	

Table 6.1. Results from subgroup analyses for child data, looking at whether there was any convergence in the comparison of pre-test and post-exposure productions.

There were significant overall findings of convergence for /u/ on every measure except duration and F3. However, for these variables, there are significant convergence findings for some main subsets of data, such as for the child-directed register overall (significant for both vowels for duration, and for /i/ for F3). An inspection of the means revealed that the significant finding for f0 was a finding of divergence rather than convergence. All other overall findings were findings of significance.

We will now begin with analyses of each variable individually, in order to determine not just whether there was convergence, but what experimental factors affected convergence on each variable. In order to better understand trends in the data, we break up the experimental variables into three categories: global measures, global phonetic measures, and individual phonetic measures.

Global Measures:

Duration:

For the duration measure, we did not see any overall findings of convergence for child participants. Looking at further subsets of data, we saw convergence for /i/ in both registers, and the audiovisual modality. For /u/, we saw convergence only in the child-directed register and the audiovisual modality.

		Overall	Register		Modality		Gender	
			Adult-Directed	Child-Directed	Auditory	Audiovisual	Males	Females
Duration	/i/	✗	✓	✓	✗	✓	✗	✗
	/u/	✗	✗	✓	✗	✓	✗	✗

Table 6.2. Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for duration in child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for duration is presented in Table 6.3. Table 6.3 also presents results from likelihood ratio tests

comparing the full model with models excluding each factor or interaction. Overall, this model accounts for 78.64% of the variance in this data set.

<i>Children Pretest Duration</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	118.63	42.65	2.78	0.008*	29.47	3	0.000	0.0180
Modality	38.07	28.63	1.33	0.191	3.32	3	0.344	-0.0028
Gender	54.55	35.99	1.52	0.138	4.56	4	0.336	0.0004
Vowel	-4.60	19.14	-0.24	0.811	7.89	4	0.096	0.0667
RegisterXGender	3.66	59.29	0.06	0.951	0.00	1	1.000	0.0042
ModalityXGender	-66.61	45.28	-1.47	0.149	2.65	1	0.104	-0.0016
RegisterXVowel	78.06	32.03	2.44	0.019*	6.30	1	0.012	0.0648
ModalityXVowel	0.80	23.34	0.03	0.973	0.07	1	0.787	-0.0035
GenderXVowel	-9.32	22.24	-0.42	0.677	0.28	1	0.597	-0.0015

Table 6.3. Summary of results from the mixed effects model for duration for child participants in pretest tokens compared to post-exposure production tokens.

Overall, the model showed a significant main effect of Register and a significant interaction of Register X Vowel. Post-hoc testing confirms the significant effect of register ($p < 0.001$), favoring increased imitation in the child-directed register, and of the interaction of Register X Vowel ($p = 0.011$). Looking at the pairwise interaction of Register X Vowel, post-hoc testing reveals that within either register, the two vowels were not imitated significantly differently (child was near significant: $p = 0.057$, adult: $p = 0.925$), but between registers, there was more imitation in the child-directed register for each vowel, and there was more imitation for /u/ than /i/ (all $p < 0.001$).

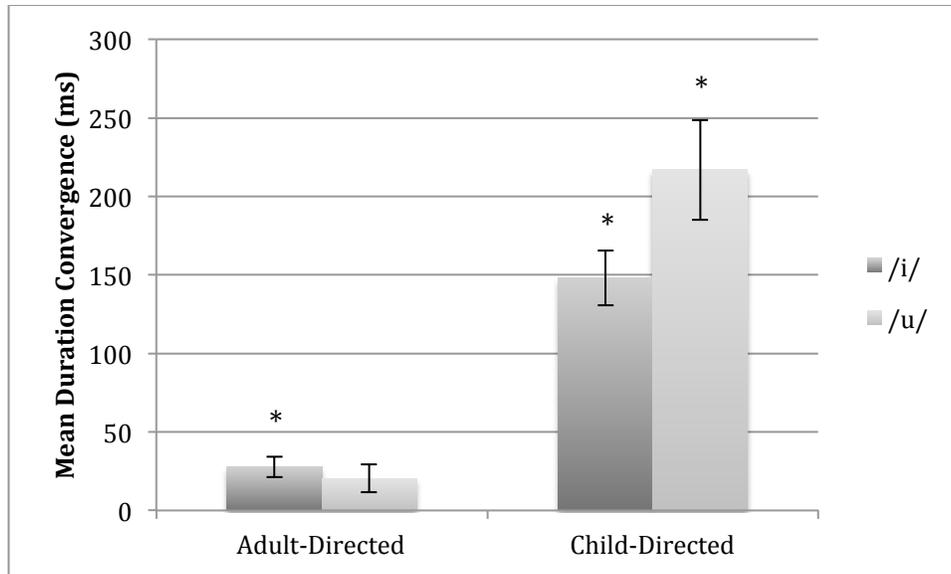


Figure 6.1. Comparison of pre-test and post-exposure phonetic distance in each vowel by register for child subjects in the duration measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

f_0 :

Recall that for f_0 , we saw overall significant divergence for the vowel /u/. None of the other subgroupings (by register, modality, or gender) showed significant findings of convergence or divergence.

The final output for the mixed effects model including all factors and interactions for fundamental frequency is presented in Table 6.4. Table 6.4 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Overall, this model accounts for 87.55% of the variance in this data set.

The mixed effects model for f_0 showed no significant effects or interactions.

<i>Children Pretest F0</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-28.58	28.35	-1.01	0.319	1.65	3	0.648	0.0012
Modality	3.40	18.82	0.18	0.858	1.65	3	0.648	0.0003
Gender	19.25	24.39	0.79	0.435	3.67	4	0.453	0.0073
Vowel	-3.19	6.53	-0.49	0.628	3.62	4	0.460	0.0161
RegisterXGender	11.10	41.73	0.27	0.792	0.05	1	0.831	0.0001
ModalityXGender	-32.58	31.87	-1.02	0.313	1.29	1	0.256	-0.0004
RegisterXVowel	5.68	10.92	0.52	0.606	0.39	1	0.531	0.0015
ModalityXVowel	2.95	7.96	0.37	0.712	0.24	1	0.628	0.0007
GenderXVowel	-8.96	7.58	-1.18	0.244	1.68	1	0.195	0.0079

Table 6.4. Summary of results from the mixed effects model for f0 for child participants in pretest tokens compared to post-exposure production tokens.

Summary: Global Measures

We saw clear register effects for the child participants on the global measure of duration for both of the English-like vowels in the pre-test/post-exposure comparison. We saw a clear register effect on both vowels (significantly more convergence for the child-directed register) but the register effects were maximally evident for /u/, in which we saw convergence only in the child-directed register (there was also a register/vowel interaction which showed greater register differences for /u/). For modality, we saw significant convergence only in the audiovisual modality for both vowels; the results for the auditory modality were not significant.

For f0, however, we did not see any effects of modality or gender, and overall, there was significant divergence to the model talker on this measure.

Global Phonetic Measures

Euclidean Distance (F1 + F2 + F3):

In our analysis of global convergence, recall that we saw that child subjects in all conditions converged on the Euclidean Distance (F1 + F2 + F3) measure in the experiment for /u/ but not for /i/. For /i/ we only saw convergence for child-directed speech. For /u/, we saw convergence overall, and we saw convergence in both genders, both modalities, and both registers.

		Overall	Register		Modality		Gender	
			Adult-Directed	Child-Directed	Auditory	Audiovisual	M	F
Euclidean Distance (F1+F2+F3)	/i/	✗	✗	✓	✗	✗	✗	✗
	/u/	✓	✓	✓	✓	✓	✓	✓

Table 6.5. Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for Euclidean Distance (F1+F2+F3) in child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for the Euclidean Distance (F1 + F2 + F3) is presented in Table 6.6. Table 6.6 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction.

<i>Euclidean Distance</i> <i>F1+F2+F3</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-0.80	0.53	-1.51	0.140	5.68	3	0.128	0.0471
Modality	1.14	0.36	3.16	0.003*	14.96	3	0.002	-0.0871
Gender	-0.26	0.44	-0.59	0.557	7.41	4	0.116	0.0590
Vowel	0.02	0.35	0.07	0.947	9.18	4	0.057	0.1920
RegisterXGender	1.15	0.67	1.70	0.097	3.49	1	0.062	-0.0295
ModalityXGender	-0.01	0.52	-0.02	0.988	0.00	1	1.000	0.0249
RegisterXVowel	0.75	0.57	1.33	0.191	2.04	1	0.153	0.0531
ModalityXVowel	-0.52	0.42	-1.25	0.218	1.95	1	0.163	0.0152
GenderXVowel	0.54	0.39	1.38	0.176	2.18	1	0.140	0.0558

Table 6.6. Summary of results from the mixed effects model for Euclidean Distance (F1+F2+F3) for child participants in pretest tokens compared to post-exposure production tokens.

Overall, there was a significant main effect of Modality, but no significant interactions. Post-hoc testing did not confirm the significant result ($p = 0.0845$), but we still plotted this data below in Figure 6.2.

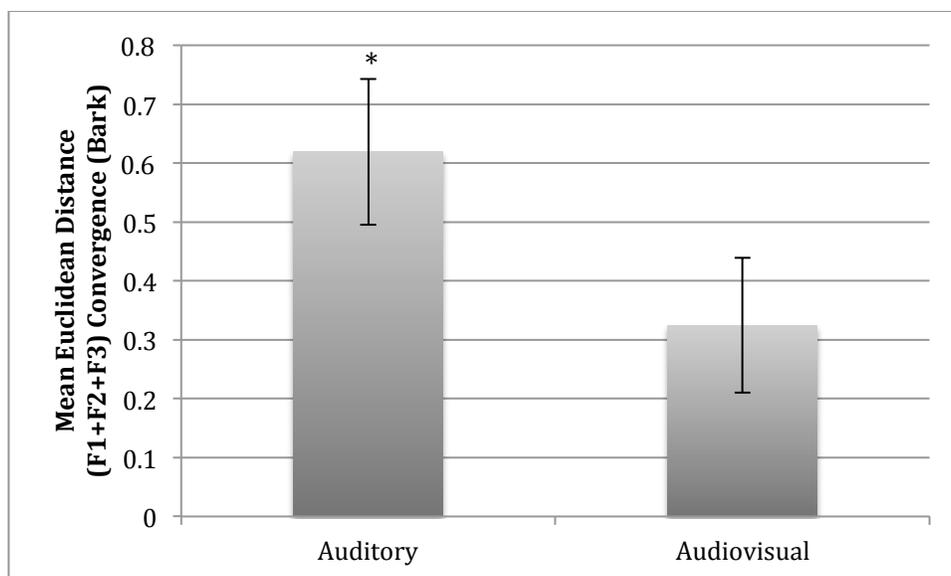


Figure 6.2. Comparison of pre-test and post-exposure phonetic distance in each modality for child subjects in the Euclidean Distance (F1+F2+F3) measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Euclidean Distance (F1 + F2):

In our analysis of global convergence, recall that we saw that child subjects in all conditions converged overall on the Euclidean Distance (F1 + F2) measure in the experiment for /u/ but not for /i/. Further analysis revealed no significant convergence findings for subsets based on gender, modality, or register for /i/. For /u/, like for the global phonetic measure including F3, we saw convergence in all main conditions.

The final output for the mixed effects model including all factors and interactions for the Euclidean Distance (F1 + F2) measure is presented in Table 6.7. Table 6.7 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Overall, this model accounts for 46.54% of the variance in this data set.

<i>Euclidean Distance F1+F2</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-0.88	0.67	-1.31	0.199	4.45	3	0.217	0.0547
Modality	1.38	0.46	3.01	0.005*	15.53	3	0.001	-0.0983
Gender	-0.19	0.56	-0.34	0.739	6.55	4	0.162	0.0490
Vowel	-0.23	0.40	-0.57	0.571	7.24	4	0.124	0.1175
RegisterXGender	1.16	0.88	1.32	0.195	2.14	1	0.143	-0.0026
ModalityXGender	0.12	0.67	0.18	0.861	0.00	1	1.000	0.0169
RegisterXVowel	0.93	0.66	1.41	0.168	2.32	1	0.128	0.0407
ModalityXVowel	-0.33	0.48	-0.68	0.503	0.61	1	0.435	-0.0015
GenderXVowel	0.64	0.46	1.40	0.170	2.29	1	0.130	0.0401

Table 6.7. Summary of results from the mixed effects model for Euclidean Distance (F1+F2) for child participants in pretest tokens compared to post-exposure production tokens.

Overall, there was a significant main effect of Modality, but no significant interactions. Like in the Euclidean Distance measure including F3, this did not come out as significant in post-hoc testing ($p = 0.084$).

Summary: Global Phonetic Measures

The global phonetic measures provided minimal or no evidence for register or modality effects. There was only one finding which suggested a difference in imitation across registers, and that was a significant finding of convergence in the child-directed, and not adult-directed, register for /i/ on the global phonetic measure including F3. As for a modality effect, for both measures, the larger mixed effects models reported a modality finding, which were not significant in post-hoc testing. Both of these reported increased imitation on the global phonetic

measures in the *auditory* modality. We will look to the individual phonetic measures in the following section to explain these findings.

The most robustly observed difference in the global phonetic measures was a difference in imitation of the two vowels; /i/ was significantly imitated in fewer conditions than /u/. For the Euclidean Distance (F1+F2) measure, there were no significant convergence findings for /i/, but convergence findings for /u/ in all conditions. For the measure including F3, there was only one significant finding of convergence for /i/, and that was for the child-directed register, and once again, /u/ was significant in all conditions. However, the mixed effect models did not identify a significant effect of vowel for either measure, so while these vowel differences are evident in significant findings of convergence, there is no evidence for a significantly different degree of convergence.

Individual Phonetic Measures

F1:

We found a significant result of convergence overall for /u/, and not for /i/ for the F1 measure for child-subjects. Looking at further subsets, we saw no significant results of convergence for any subset for /i/, but for /u/, we saw significant convergence findings for both genders and both registers, but neither modality.

The final output for the mixed effects model including all factors and interactions for the first formant is presented in Table 6.8. Table 6.8 also presents results from likelihood ratio tests

comparing the full model with models excluding each factor or interaction. Overall there were no significant main effects or interactions.

<i>Children Pretest F1</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-0.35	0.36	-0.97	0.340	3.55	3	0.315	0.0688
Modality	0.14	0.25	0.58	0.568	1.74	3	0.629	0.0100
Gender	0.54	0.29	1.83	0.074	10.12	4	0.039	-0.0625
Vowel	0.41	0.25	1.65	0.107	14.03	4	0.007	0.1687
RegisterXGender	0.43	0.44	0.97	0.340	1.15	1	0.283	0.0119
ModalityXGender	-0.37	0.34	-1.10	0.276	1.51	1	0.220	0.0081
RegisterXVowel	0.55	0.41	1.34	0.189	2.11	1	0.147	0.0382
ModalityXVowel	0.08	0.30	0.26	0.797	0.15	1	0.697	-0.0194
GenderXVowel	-0.18	0.29	-0.64	0.529	0.55	1	0.459	-0.0075

Table 6.8. Summary of results from the mixed effects model for F1 for child participants in pretest tokens compared to post-exposure production tokens.

F2:

For the F2 measurement, we only found a significant result of convergence overall for /u/ and not for /i/. Looking at further subsets, we found a significant result for /i/ only in the audiovisual modality, but that was a result of divergence. For /u/, we found significant convergence not only overall, but for all the major subsets of data, divided either by modality, gender, or register.

	Overall	Register		Modality		Gender	
		Adult-Directed	Child-Directed	Auditory	Audiovisual	Males	Females
F2	/i/	✗	✗	✗	D	✗	✗
	/u/	✓	✓	✓	✓	✓	✓

Table 6.9. Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for F2 in child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for the second formant is presented in Table 6.10. Table 6.10 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction.

<i>Children Pretest F2</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	0.60	0.64	0.93	0.358	2.26	3	0.520	0.0215
Modality	-0.32	0.44	-0.72	0.474	0.97	3	0.808	0.0093
Gender	0.48	0.52	0.93	0.357	3.87	4	0.424	0.0364
Vowel	1.31	0.46	2.87	0.007*	22.92	4	0.000	0.1600
RegisterXGender	-1.00	0.77	-1.29	0.204	2.04	1	0.153	0.0194
ModalityXGender	0.08	0.59	0.13	0.896	0.02	1	0.883	0.0002
RegisterXVowel	0.02	0.76	0.03	0.975	0.00	1	0.971	0.0000
ModalityXVowel	0.11	0.56	0.19	0.847	0.05	1	0.828	0.0005
GenderXVowel	-0.47	0.53	-0.88	0.382	0.97	1	0.325	0.0093

Table 6.10. Summary of results from the mixed effects model for F2 for child participants in pretest tokens compared to post-exposure production tokens.

Overall, there was a significant main effect of Vowel, which is unsurprising considering we found convergence across the board for /u/, and only in the audiovisual modality for /i/. Post-hoc testing confirmed that this difference in vowel imitation was significant ($p < 0.001$) and the results are shown below in Figure 6.3.

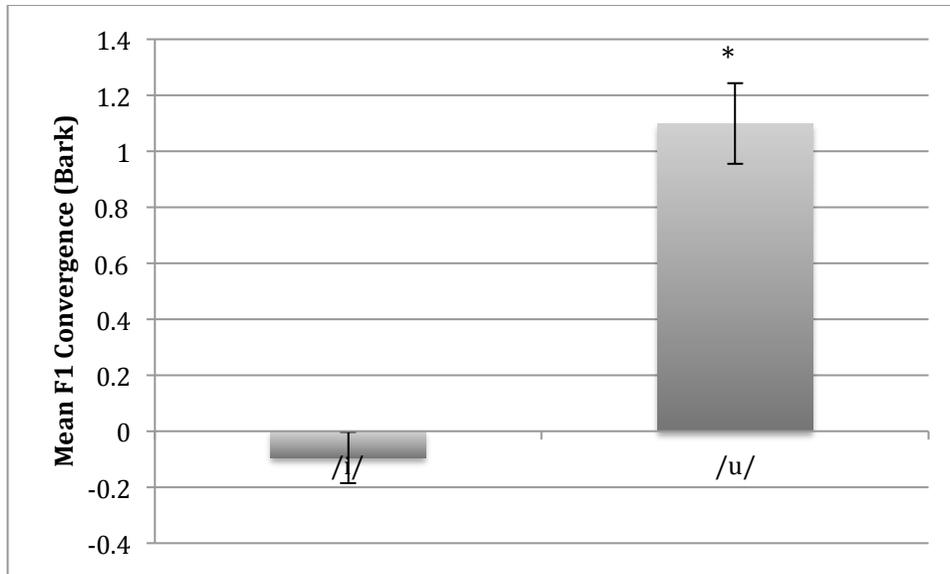


Figure 6.3. Comparison of pre-test and post-exposure phonetic distance in each vowel for child subjects in the F2 measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

F3:

For F3, we found no significant results overall for /i/ or /u/, but we found significant results of particular subsets of data for /i/ (but not for /u/). For /i/, we found significant convergence overall in the male participants, and in the child-directed register.

	Overall	Register		Modality		Gender	
		Adult-Directed	Child-Directed	Auditory	Audiovisual	Males	Females
F3	/i/	×	✓	×	×	✓	×
	/u/	×	×	×	×	×	×

Table 6.11. Results from analyses looking at whether there was any overall convergence in the comparison of pre-test and post-exposure productions for F3 in child subjects. The full statistical output is shown in the appendix.

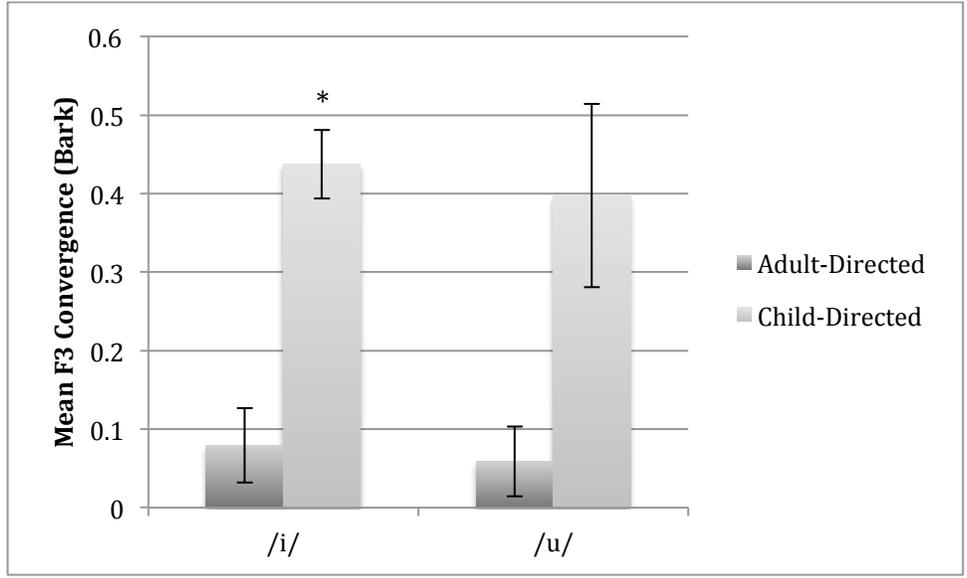


Figure 6.4. Comparison of pre-test and post-exposure phonetic distance in each vowel by register for child subjects in the F3 measurement. Significance stars represent significant convergence or divergence (difference from zero) with a *p*-value of 0.05 or smaller.

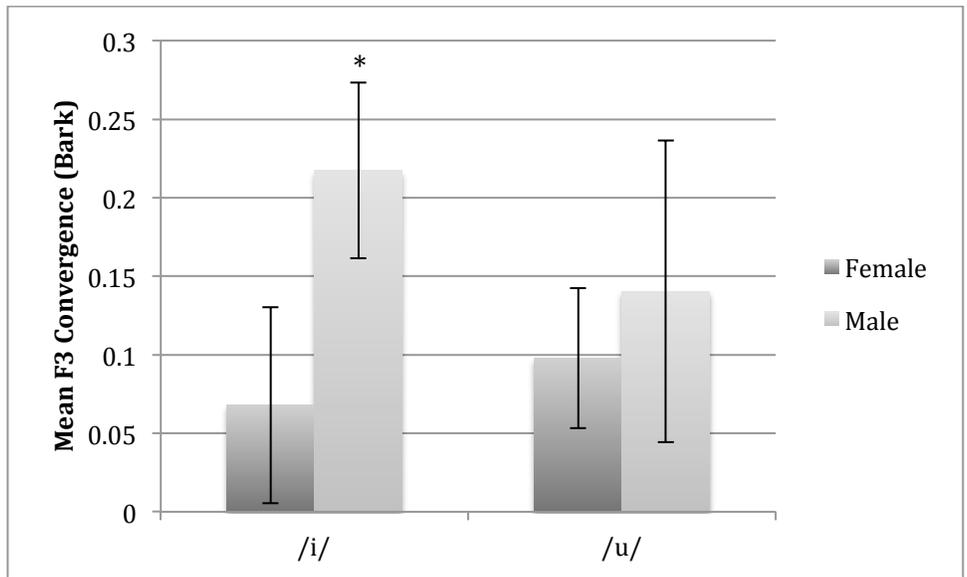


Figure 6.5. Comparison of pre-test and post-exposure phonetic distance in each vowel by gender for child subjects in the F3 measurement. Significance stars represent significant convergence or divergence (difference from zero) with a *p*-value of 0.05 or smaller.

The final output for the mixed effects model including all factors and interactions for the third formant is presented in Table 6.12. Table 6.12 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Overall, the mixed effects model showed no significant main effects or interactions.

<i>Children Pretest F3</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	0.33	0.22	1.49	0.145	6.49	3	0.090	-0.0529
Modality	0.16	0.15	1.09	0.282	2.33	3	0.507	0.0263
Gender	0.16	0.18	0.92	0.364	1.99	4	0.738	0.0243
Vowel	0.14	0.15	0.91	0.368	2.40	4	0.662	-0.0182
RegisterXGender	0.07	0.27	0.26	0.797	0.05	1	0.821	0.0120
ModalityXGender	-0.15	0.20	-0.73	0.471	0.64	1	0.423	0.0128
RegisterXVowel	-0.10	0.26	-0.38	0.707	0.24	1	0.623	-0.0135
ModalityXVowel	-0.22	0.19	-1.19	0.241	1.70	1	0.192	0.0182
GenderXVowel	-0.10	0.18	-0.53	0.597	0.41	1	0.523	-0.0095

Table 6.12. Summary of results from the mixed effects model for F3 for child participants in pretest tokens compared to post-exposure production tokens.

Summary: Individual Phonetic Measures

The individual phonetic measures showed little evidence of register or modality effects. For F1, we did not see differences in significant convergence according to register or modality, and no results in the mixed effects models that point towards differing imitation according to these factors. For F2, the only evidence of a register or modality effect was that there was a significant finding of divergence in the audiovisual, and not auditory, modality for /i/ (/u/ had

significant convergence findings for both modalities). For F3, there was another incomplete effect, but this time for register; there was a significant convergence finding in the child-directed but not adult-directed register only for /i/ (this time /u/ showed no significant convergence findings in either register). While the register effect is hardly substantial, it is in the expected direction. The modality effect, however, is not. Recall the finding of increased imitation in the auditory modality for the global phonetic measures (even though it did not remain significant in post-hoc testing). In order to find out if it is just the F2 measure that is motivating this finding (since we saw no evidence for it in the analysis of the individual measures themselves), we plotted the data for convergence in each modality by vowel in the figure below.

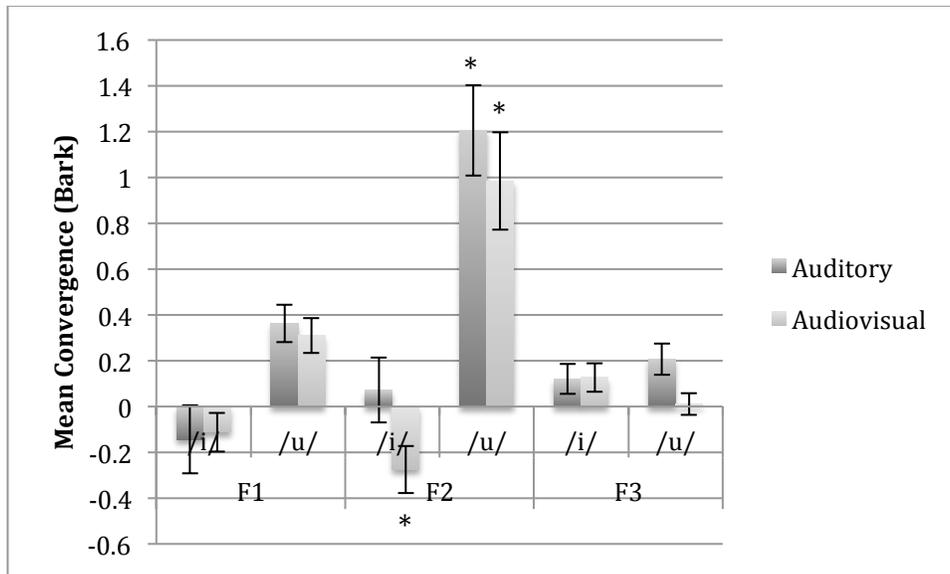


Figure 6.6. Comparison of pre-test and post-exposure phonetic distance in each vowel by modality for child subjects in each of the individual phonetic measurements. Significance stars represent significant convergence or divergence (difference from zero) with a *p*-value of 0.05 or smaller.

Of note is that we see more convergence for /u/ in the auditory modality on each measure, and for /i/ in F2 (making the auditory > audiovisual effect complete for F2). For F1 and F3, however, we see more convergence (or less divergence) for the audiovisual modality (although these do not appear to be nearly significant) for /i/. From this analysis we would assume that the results for the global measures are driven by F2 and by the vowel /u/.

Finally, recall in the section on global phonetic measures, we saw many more convergence findings for /u/ than /i/. In the individual phonetic measures, we saw evidence for this vowel difference in F1 and F2, but not F3. For both F1 and F2, we saw a number of significant convergence findings for /u/ but none for /i/ (the only significant finding was a significant finding of divergence). Additionally, for F2, we found a significant result for vowel in the mixed effects model, favoring larger amounts of imitation for /u/ than /i/. F3 was the only measure that we saw more favorable convergence for /i/, as we found a significant result overall for male /i/ convergence, and for convergence in the child-directed register, and no findings of convergence to /u/ anywhere.

Interim summary: English-like vowels, pre-test vs. post-exposure convergence

Overall, we saw some evidence of register and modality effects, but these effects depended on the measure of analysis. The register effects were all in the expected direction, whereas the modality effects depended on the type of measure being analyzed. These results will be discussed in the subsections below, following the discussion of the direction of convergence.

Direction of Convergence

The sections above have given information about whether there was significant convergence or divergence exhibited by child subjects in the pre-test/post-exposure comparison. However, there is information missing when just asking “is there convergence?”, namely, how subjects are actually changing their productions to converge with the model talker. In this section, we show graphs that can indicate the *direction* of convergence.

The graphs below summarize the results for production in each of our experimental measures according to register and modality, separated by vowel. In each of these graphs, rather than indicating the *amount* of convergence (as in the graphs previously shown), we plot mean values in the pretest and post-exposure productions.

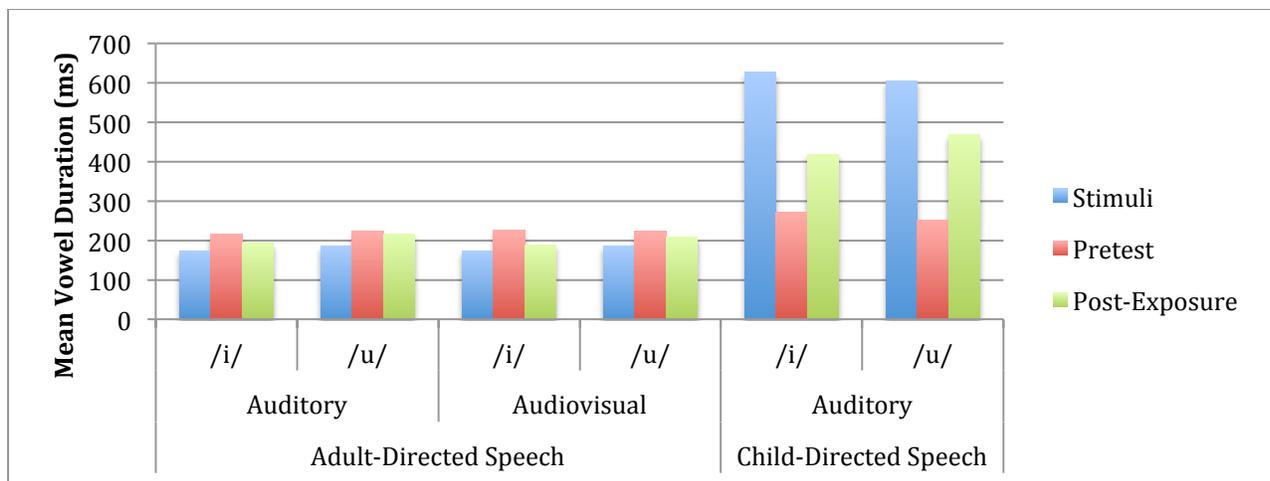


Figure 6.7. Mean duration of vowels produced by subjects in in the pretest and post-exposure phases, compared with the mean duration of the stimuli, separated by vowel, modality, and register.

In the duration graph above, we can see that in the adult-directed speech condition, child subjects shortened their vowel duration, imitating the model talker. In the child-directed speech condition, child subjects increased their vowel duration, once again, imitating the model talker.

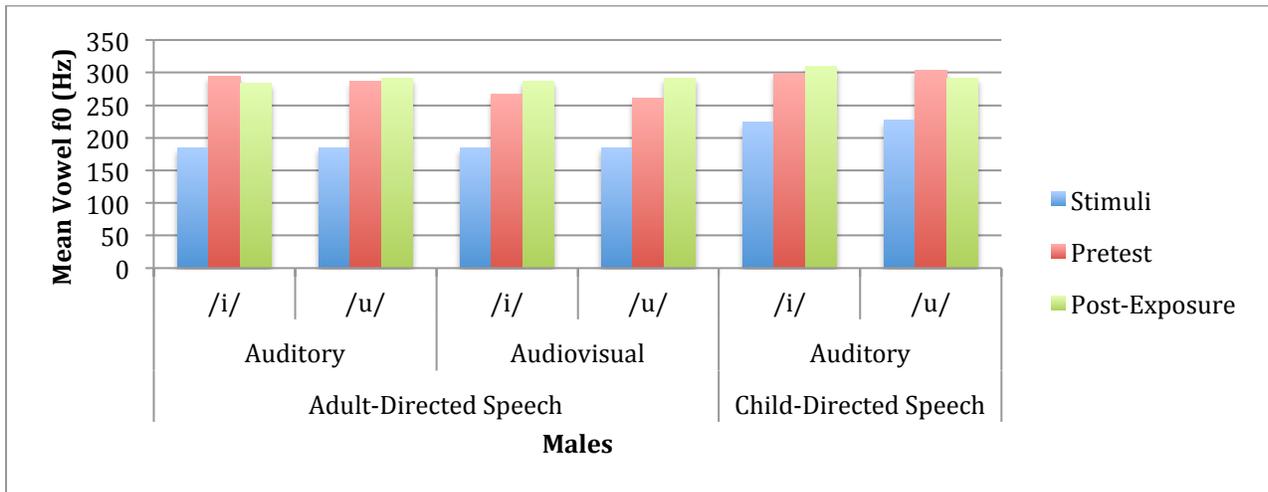


Figure 6.8. Mean f0 of vowels produced by male child subjects in in the pretest and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel, modality, and register.

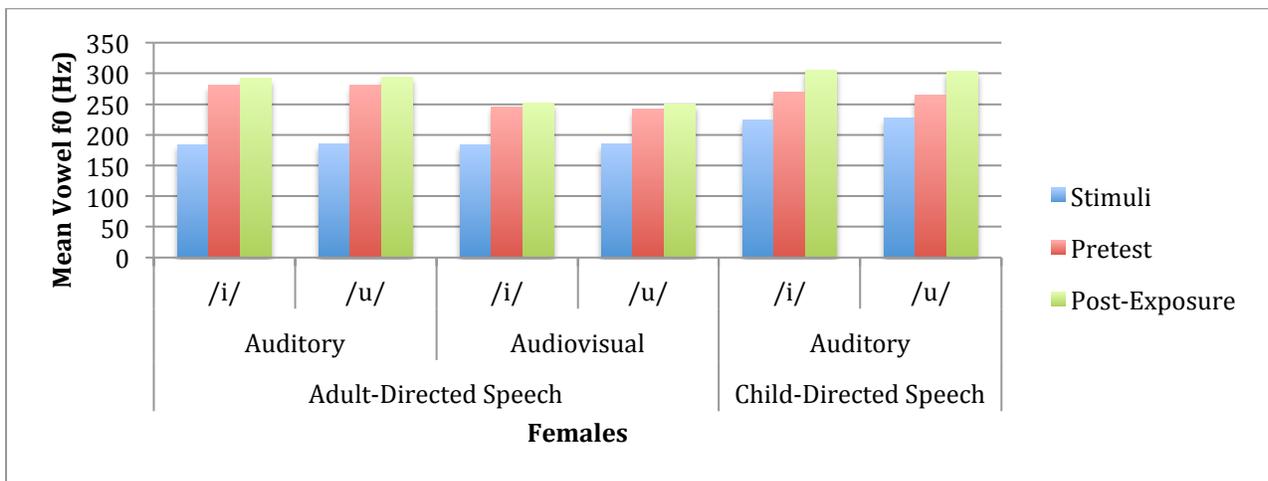
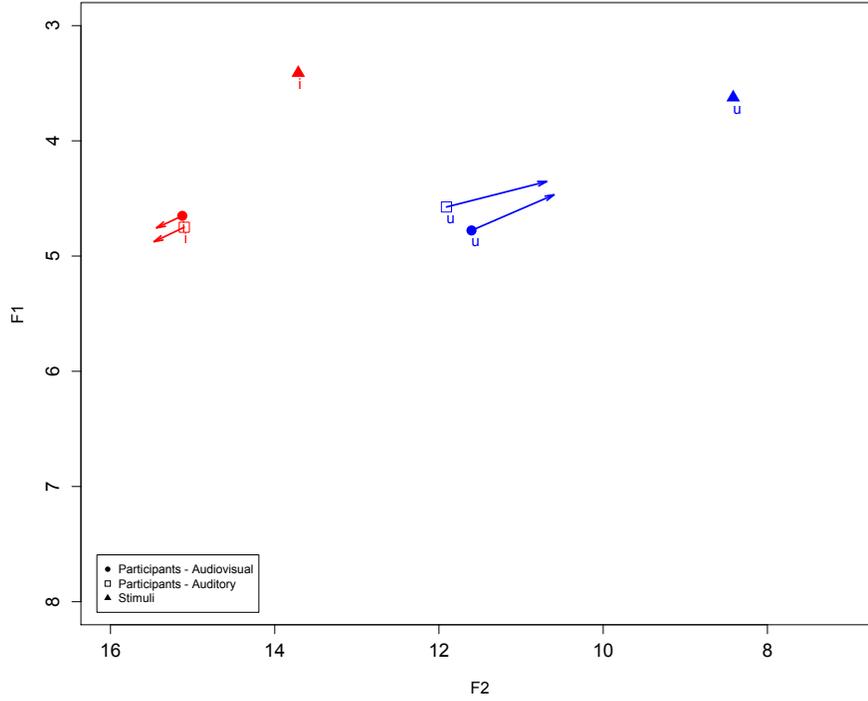


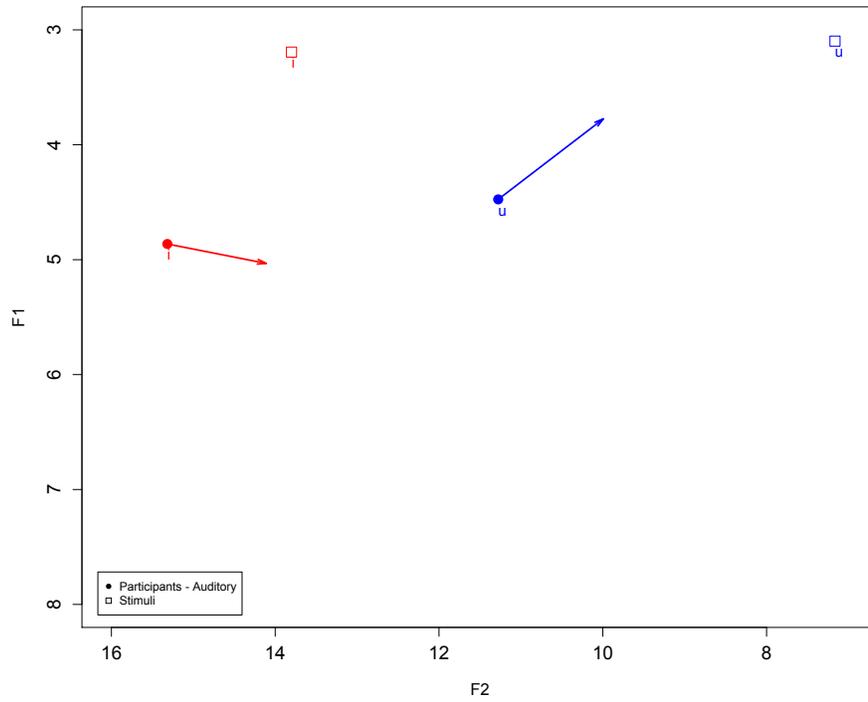
Figure 6.9. Mean f0 of vowels produced by female child subjects in in the pretest and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel, modality, and register.

The graphs above show the direction of convergence to f_0 for male and female children separately. For male talkers, we saw that subjects raised their f_0 after continued exposure to the model talker, except for /u/ in the child-directed speech:auditory condition and for /i/ in the adult-directed speech:auditory condition. Female child subjects raised their f_0 after continued exposure to the model talker for all vowels and all conditions. This pattern is especially interesting, considering in all cases, the f_0 of the model talker was initially lower than their initial f_0 , and so after continued exposure, the child subjects *diverged* in their productions. However, this seems to be similar to the pattern seen in the previous chapter for many of the adult female f_0 results. It is possible that child subjects recognized that the speaker had a higher than average f_0 for his speech, and in turn they raised their own f_0 . In order to determine the validity of this hypothesis, imitation of speakers with lower average f_0 s is needed.

F1-F2 Imitation: Adult-Directed Speech



F1-F2 Imitation: Child-Directed Speech



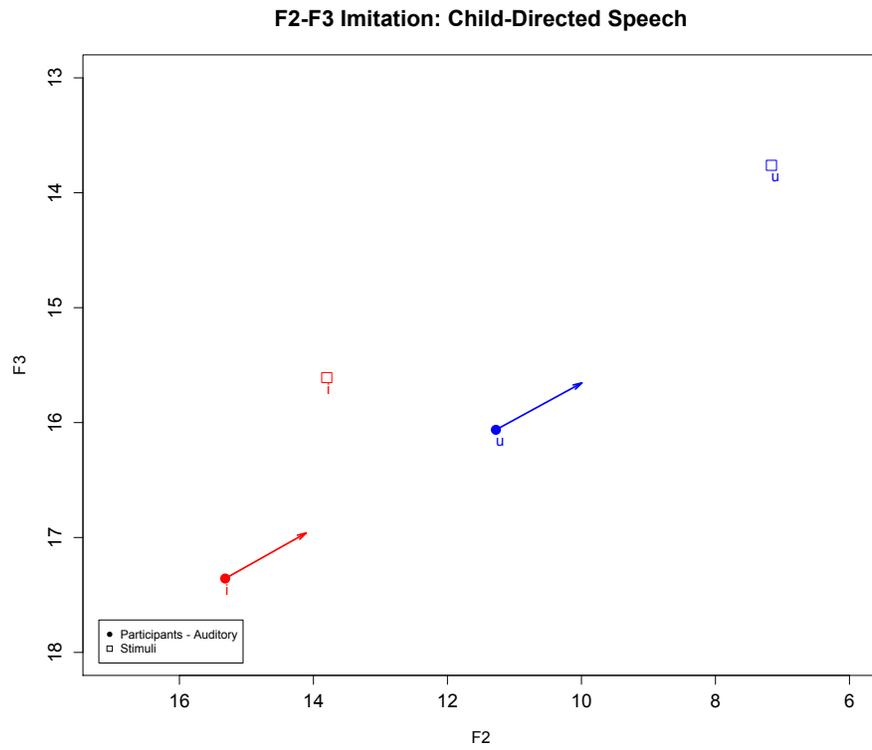
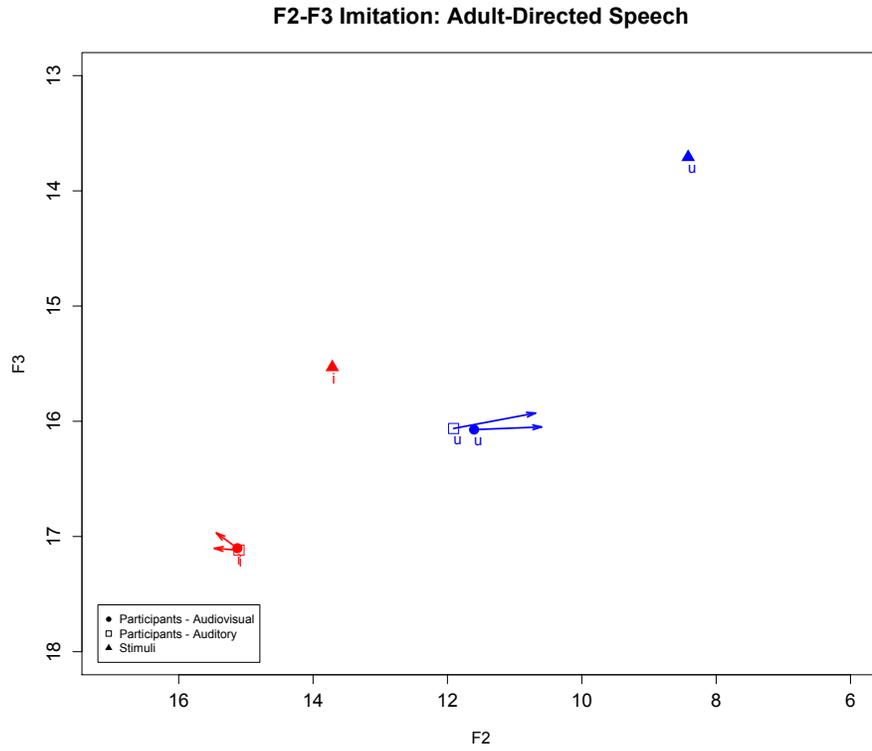


Figure 6.10. Formant plots of vowels produced by subjects in the pretest and post-exposure phases, compared with the mean formant values of the stimuli, separated by vowel, modality, and register.

The graphs above show the direction of convergence for child subjects in the phonetic measures. Looking first at the F1/F2 adult-directed speech graph, we see clear differences in imitation between /i/ and /u/; subjects converged in both F1 and F2 for /u/ (with larger convergence in F2 than F1) raising and backing their productions, but for /i/, subjects diverged in both dimensions, lowering their F1 and raising their F2 values. For child-directed speech, we see a large amount of convergence on F1 and F2 for /u/, and convergence on F2 for /i/, making their production further back, but slight divergence on F1, lowering rather than raising their production. For F2/F3 in adult-directed speech, we saw little convergence on F3 for /i/, and some convergence on F3 for /u/, but only in the audiovisual condition. For F2/F3 in child-directed speech however, we see both vowels are imitated correctly in both dimensions, raising their F3 values, and lowering their F2 values (and to about the same degree).

Register

The figure below shows the convergence results for each measure separated by register. We can see that for all measures, there is more convergence in the child-directed register compared to the adult-directed register (however, these results were not all significantly different), with the exception of f0. For f0, there was divergence in both registers for f0, and there was more divergence for the child-directed register.

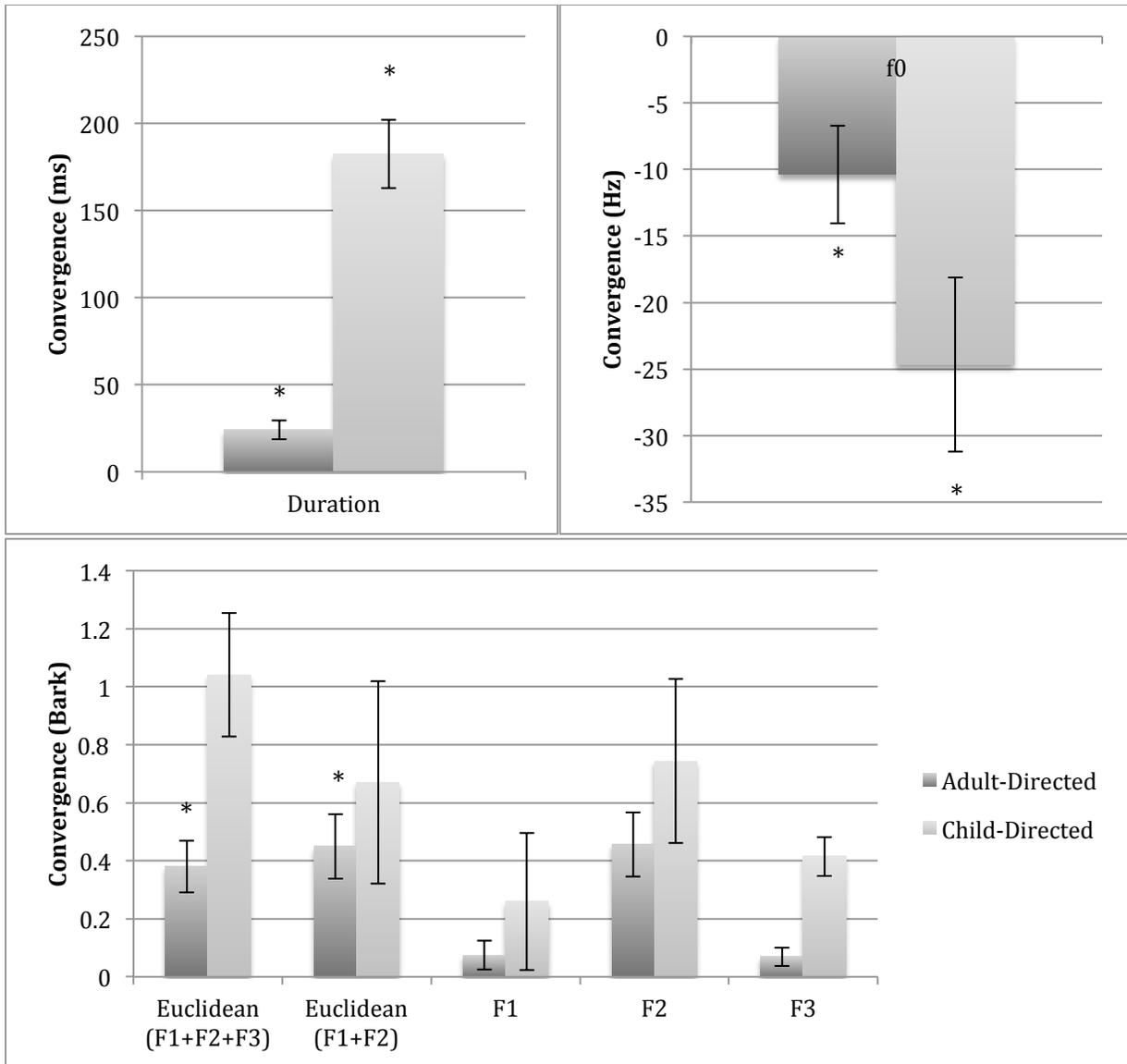


Figure 6.11. Results for convergence according to register for each measure in the pre-test/post-exposure comparison for child participants. The dark bars represent results for the adult-directed register and the light bars represent results for the child-directed register. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Looking at the significant results that showed register effects, we saw the strongest register effects for duration. We saw a significant difference in imitation by register for this measure, as

well as differences in significant convergence (there was only significant convergence to the child-directed, not adult-directed, register for /u/). For both the overall Euclidean Distance (F1+F2+F3) measure and for the F3 measure, we found significant convergence for /i/ in the child-directed, but not adult-directed register (F3 is likely driving this result for the global phonetic measure). Looking at the relative amounts (Figure 6.11 above) we also see greater amounts of convergence, but these were not statistically significant. We saw no evidence of register effects for f0, Euclidean Distance (F1+F2), F1, or F2. Overall, the differences in convergence based on register were evidence for increased convergence in the child-directed register. These differences were evident in duration and F3.

Modality

The findings based on modality are inconclusive, but when present point towards increased imitation in the auditory modality. The graph below plots all of the data separated by modality. In all measures except f0, there is a larger mean convergence value for the auditory compared to the audiovisual modality (and for f0, the results are extremely similar).

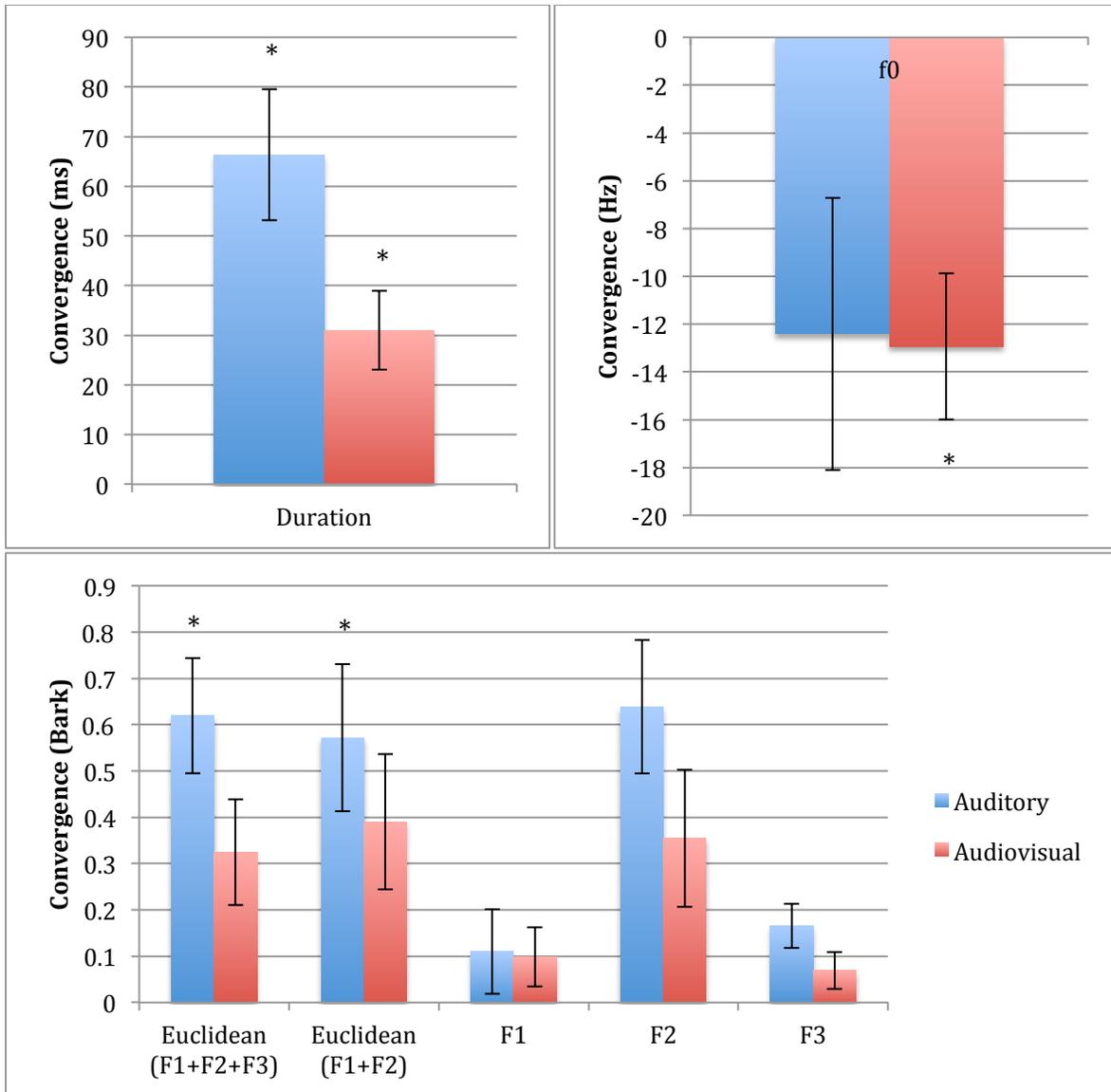


Figure 6.12. Results for convergence according to modality for each measure in the pre-test/post-exposure comparison for child participants. The blue bars represent results for the auditory modality and the red bars represent results for the audiovisual modality. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

The global phonetic measures had a nearly significant (significant in the modeling, but not post-hoc testing) finding of greater imitation in the auditory modality. Looking at the phonetic

measures individually, they all seem to trend towards increased imitation in the auditory modality, so it appears that these results are significant only when combined. One other interesting finding to note is that there is only a significant finding of convergence in duration in the audiovisual modality, even though we can see the amount of convergence appears much greater in the auditory modality. The difference here can likely be attributed to variability within the group.

Overall, it appears that the addition of the visual cues did not aid child subjects in their implicit imitation of the English-like vowels (when comparing pre-test tokens with post-exposure tokens). In fact, the visual cues may have slightly hindered imitation.

Vowel

One other finding in the pre-test/post-exposure comparison is that child subjects were much more inclined to significantly imitate /u/ than /i/. There were more significant convergence findings overall in this comparison for /u/ than /i/ (as a rudimentary measure of comparison, there were six findings of convergence in this section for /i/, and twenty-eight for /u/). Additionally, for F2, there was a finding of significantly greater convergence for /u/ than /i/. This follows along with the predictions about F2 imitation in /u/ compared to /i/ based on the sociophonetic status of /u/ within California English. While these predictions were not specifically made for child subjects, it is interesting that they are also susceptible to this increased F2 /u/ imitation, like adults.

Did the nature of the exposure session affect convergence within the experiment?

In this section of the experiment, we will look at whether child subjects converged to the model talker within the experiment. For this section, we will look at the subjects' productions in the pre-exposure and post-exposure phases, and compare the phonetic distance of the two productions with the productions of the model talker. For both analyses, we included two random factors in all models: Word and Subject. This allowed each subject to differ in his or her overall degree of convergence, and, since we were comparing at the word-level for this analysis, each word to have a unique pronunciation.

Like in the pre-test/post-exposure comparison, we separated the analysis by variable, and for each variable, we report two difference analyses. Like in the previous section all statistical analyses were done with mixed effects models, using the lme4 package in R (R Development Core Team, 2008). These models have the advantage of including random effects in addition to fixed effects. . In order to further interpret patterns, we subjected the data to a repeated measures analysis of variance, using Tukey's HSD test, also in R.

Like for the pre-test/post-exposure comparison, each section will have two types of analyses. The first portion of the analysis looks at overall convergence between the two experimental sessions. For this part of the analysis, the variable was the phonetic *difference* (for each phonetic measure) between the model talker and the participant, with a factor of Repetition (for this section, *pre-exposure* or post-exposure) to differentiate the two productions. We ran mixed effect models for each type of vowel separately (English-like or foreign vowels) with the fixed effect of Repetition and the random effects of Subject and Word.

For the second part of the analysis, we looked at how the experimental factors affected the convergence measure in the experiment. We model how each of 7 dependent variables (duration convergence, f0 convergence, F1 convergence, F2 convergence, F3 convergence, Euclidean distance with just F1 and F2, and Euclidean distance with F1, F2, and F3) vary according to the experimental design factors. The experimental design was a 2 (Gender: male or female) X 2 (Modality: audio or visual) X 2 (Register: child-directed or adult directed speech) X 2 (Vowel Type: English-like or foreign) X 2 (Carrier Type: CVC or CV) factorial design.²¹ The between-subject variables were Subject, Gender, Modality, and Register, and the within-subject variables were Vowel Type, Carrier Type, and Word. We provide full models for each variable, including main effects and two-way interactions (resulting in nine possible fixed effects, and two random effects) with the exception of the interaction of Register X Modality, which was not included because there was not data from all four subgroups.

The number of variables for analysis makes statistical analysis very complicated. To simplify the model, we analyzed the effects of convergence separately for male and female subjects (see also Babel, 2009) given that gender differences in convergence are well established in the literature on convergence in adults (Pardo, 2006).

While for the first part of the analysis, we look at the data all together, for the more complex mixed effect models, we will present male and female data separately. We first present the male data, and following that we will have an interim summary before moving on to the

²¹ There were a few other factors not considered in the modeling: the place of articulation of the onset consonant, the specific carrier ('d_', 'g_k', etc.), the final consonant (/g/, /k/, none. Carrier type, CV or CVC, was used instead.), the voicing of the consonants (although this should actually affect duration measurements, it is not intrinsically interesting because duration is known to vary with voicing of surrounding consonants), and, finally, the word itself. The combination of carrier type, vowel, and place of articulation of the onset consonant (analyzed separately) were thought to capture the variation that could potentially be seen in the Word factor.

female data. At the end of this section, we will have another discussion, comparing the female performance to the male performance.

Is there convergence between the pre-exposure and post-exposure productions?

Once again, before we get into the analysis of child participants, we wanted to ensure that these children did in fact imitate the speech they heard. In Table 6.13, we show the output for mixed effects models for the Repetition factor with all data included in the analysis for the pre-exposure/post-exposure comparison. Each row in the table represents the output for the factor of Repetition (pre-test or post-exposure) in separate mixed effect models for each factor for each vowel.

	Variable	Vowel Type	Estimate	Std. Error	t-value	p-value	Significant
Global Measures	Duration	English-like	8.17	2.72	3.01	0.0026	*, divergence
		Foreign	7.49	3.47	2.16	0.0308	*, divergence
	f0	English-like	4.51	0.92	4.88	$p < 0.0001$	*, divergence
		Foreign	3.95	1.20	3.28	0.0011	*, divergence
Global Phonetic Measures	Euclidean Distance F1+F2+F3	English-like	-0.10	0.03	-3.06	0.0023	*
		Foreign	-0.14	0.04	-3.71	0.0002	*
	Euclidean Distance F1+F2	English-like	-0.09	0.03	-2.70	0.0070	*
		Foreign	-0.13	0.04	-3.09	0.0021	*
Individual Phonetic Measures	F1	English-like	0.05	0.03	1.62	0.1055	ns
		Foreign	0.12	0.04	2.96	0.0031	*
	F2	English-like	0.01	0.04	0.34	0.7331	ns
		Foreign	0.08	0.04	1.95	0.0513	ns
	F3	English-like	0.02	0.02	1.14	0.2541	ns
		Foreign	0.03	0.02	1.19	0.2324	ns

Table 6.13. Results from subgroup analyses for child subjects, looking at whether were significant findings of convergence across all measures in the comparison of pre-exposure and post-exposure productions.

As we can see from the table above, there were significant findings for both vowel types for the global measures and global phonetic measures. For the individual phonetic measures, for F1 there was a significant finding for only one vowel type, and no significant overall findings for F2 and F3. For F2 and F3, there were subsets in which we did see significant convergence, so we will conduct further analysis of these variables. Also of note was that for both of the global measures, we saw significant findings of divergence, not convergence.

We will now take a more in-depth look at the results, as well as how each of the experimental factors affected convergence in the sub-sections below. We will begin with an analysis of the male child data.

Male Child Participants:

The following sections will present the data from male child participants in the pre-exposure/post-exposure comparison. We will first discuss the global measures, followed by the global phonetic measures. Finally, we will discuss the individual phonetic measures. At the end of this section, before moving on to female child participants, we will summarize the results for this subject group.

The numbers of participants for these groups was as follows: the auditory exposure, adult-directed register group consisted of 3 male participants, the audiovisual exposure, adult-directed register group consisted of 4 male participants, and the auditory exposure, child-directed register group consisted of 2 male participants.

Global Measures:

Duration:

Recall from the preceding subsection that we found significant divergence in duration overall for both vowel types. Looking at just the males specifically, this divergence finding was only significant for English-like vowels. It appears that this divergence in English-like vowels is being driven by the results for the child-directed register and the auditory modality (for the audiovisual modality and the adult-directed register there were no significant findings). For foreign vowels, there were no significant findings of divergence or convergence.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
Duration	English-like	D	×	D	D	×
	Foreign	×	×	×	×	×

Table 6.14. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for duration in male child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for duration for the male participants is presented in Table 6.15. Table 6.15 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effects of Subject [$\chi^2(1) = 0.00, p = 1$] and Word [$\chi^2(1) = 0.00, p = 1$] did not significantly reduce model fit.

<i>Duration Males</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-76.65	26.53	-2.89	0.004*	17.91	3	0.00	0.124
Modality	3.25	22.18	0.15	0.884	0.00	3	1.00	0.117
Carrier Type	18.03	20.18	0.89	0.372	2.36	4	0.67	0.115
Vowel Type	14.50	20.49	0.71	0.480	7.37	4	0.12	0.110
Register X Carrier Type	-26.38	26.31	-1.00	0.316	1.03	1	0.31	0.062
Modality X Carrier Type	-7.16	21.89	-0.33	0.744	0.12	1	0.73	0.064
Register X Vowel Type	55.95	25.46	2.20	0.028*	3.72	1	0.05	0.111
Modality X Vowel Type	5.56	21.09	0.26	0.792	0.00	1	1.00	0.120
Carrier Type X Vowel Type	-30.83	19.55	-1.58	0.115	2.52	1	0.11	0.060

Table 6.15. Summary of results from the mixed effects model for duration in pre-exposure compared to post-exposure productions for male participants.

Overall, there was a significant main effect of register and a significant interaction of Register X Vowel Type. Post-hoc testing confirmed the main effect of register ($p < 0.001$) and the interaction of register and vowel type ($p = 0.019$) were significant. There was significantly less imitation in the child-directed register (in fact, there was divergence) for both English-like and foreign vowels (English-like vowels: $p < 0.001$, foreign vowels: $p = 0.026$). There were no significant differences between the two vowel types in a single register (adult-directed: $p = 0.993$, child-directed: $p = 0.058$).

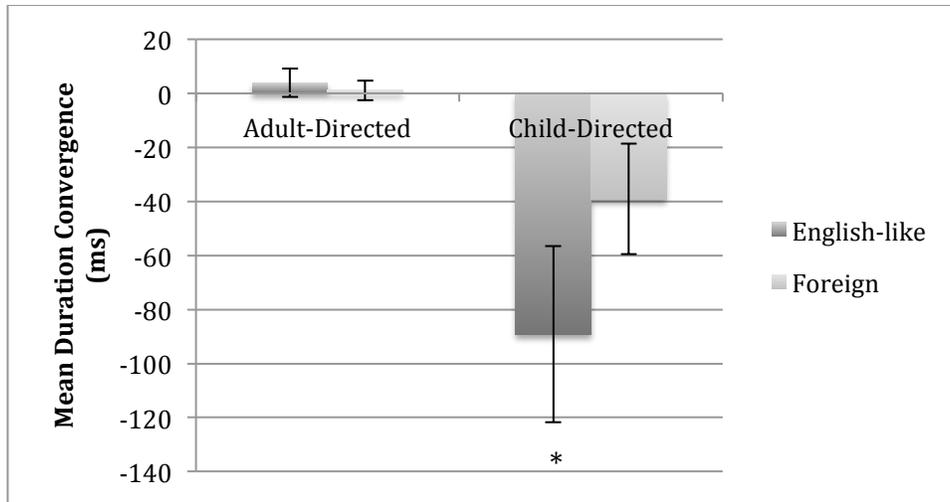


Figure 6.13. Comparison of pre-exposure and post-exposure phonetic distance in each vowel type by register for male child subjects in the duration measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

f0:

There was divergence overall for both vowel types. Looking at the two registers separately, there were no findings of convergence or divergence for the child-directed register. For the adult-directed register, there was significant divergence for both vowel types. Looking at the two modalities separately, there was significant divergence for both modalities and both vowel types for the male participants.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
f0	English-like	D	D	*	D	D
	Foreign	D	D	*	D	D

Table 6.16. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for f_0 in male child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for f0 for the male participants is presented in Table 6.17. Table 6.17 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Overall, this model accounts for 0.43% of the variance in this data set. Removing the random effects of Subject [$\chi^2(1) = 0.00, p = 1$] and Word [$\chi^2(1) = 0.00, p = 1$] did not significantly reduce model fit.

<i>f0 Males</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Partial Pseudo $-R^2$</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	30.07	7.94	3.79	0.000*	15.53	3	0.001417	0.067
Modality	15.89	6.64	2.39	0.017*	9.29	3	0.02562	0.062
Carrier Type	1.43	5.74	0.25	0.803	3.30	4	0.5088	0.061
Vowel Type	5.07	5.83	0.87	0.385	4.40	4	0.3549	0.060
Register X Carrier Type	-11.82	7.49	-1.58	0.115	2.54	1	0.1111	0.062
Modality X Carrier Type	-0.22	6.23	-0.04	0.971	0.01	1	0.939	0.066
Register X Vowel Type	-4.03	7.25	-0.56	0.579	0.32	1	0.5742	0.065
Modality X Vowel Type	-12.04	6.00	-2.01	0.045*	4.11	1	0.04256	0.060
Carrier Type X Vowel Type	0.85	5.56	0.15	0.879	0.03	1	0.8553	0.066

Table 6.17. Summary of results from the mixed effects model for F0 in pre-exposure compared to post-exposure productions for male participants.

Overall, there were significant main effects of Register and Modality, and a significant interaction of Modality X Vowel Type. Post-hoc testing confirmed the significant main effects and the significant interaction. Starting first with the register effects, there was significantly more convergence in the child-directed register.

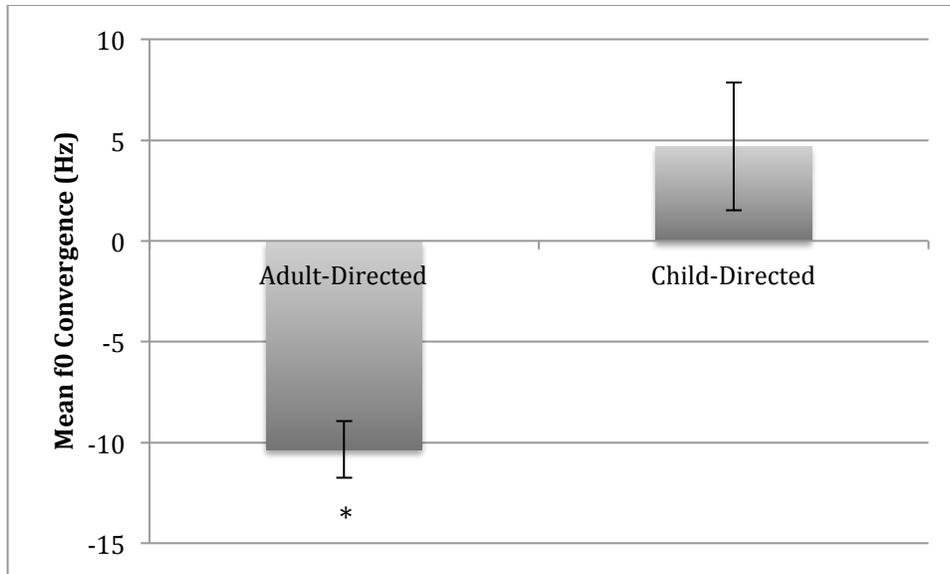


Figure 6.14. Results for convergence in f0 by male child participants by register of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

There was also significantly more overall divergence in the auditory modality. Looking at the pairwise comparisons of the Modality X Vowel Type interaction, there was a significant difference only between English-like vowels in the two modalities ($p = 0.009$), favoring significantly more divergence in the auditory modality. The differences between the foreign vowels were not significant ($p = 0.890$).

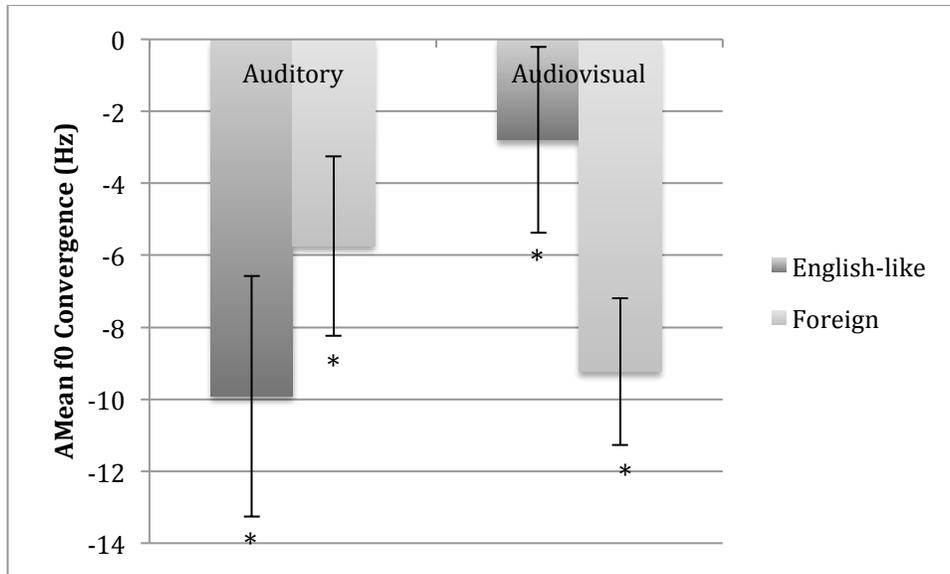


Figure 6.15. Results for convergence in f0 by male child participants by vowel type and modality of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Summary: Global Measures

The results for the global measures are rather indirect in the effects they show. For both global measures, there is significant divergence overall for both vowel types. Indirectly, we see effects of register and modality, but rather than results of greater convergence, these are results of greater divergence. To make this portion of the analysis more complicated, the two global measures show opposite patterning, at least with respect to register. For duration, there is divergence in the child-directed register (English-like vowels were the only type to show any effect) and not in the adult-directed register, and additional support for this is in the mixed effects modeling, which shows significantly more divergence in the child-directed register (for both vowel types). However, for f0, we see divergence only in the adult-directed register and an

effect favoring better imitation of child-directed speech in the mixed effects modeling. In fact, the means reveal that there is convergence to f_0 in the child-directed register, but divergence in the adult-directed register.

The results with respect to modality are slightly clearer, and they favor better imitation in the audiovisual register, at least for English-like vowels. Looking at just English-like vowels, for duration, we see significant divergence in the auditory, but not audiovisual modality and for f_0 , we see significantly more divergence in the auditory modality. For foreign vowels, we see no modality effects.

Global Phonetic Measures

Euclidean Distance ($F1 + F2 + F3$):

Recall for the overall Euclidean Distance ($F1+F2+F3$) measure, we saw convergence for both vowel types overall. For male participants, we only saw overall convergence for foreign vowels. For English-like vowels, we found no significant convergence findings for any subsets for this vowel type, and additionally, we found a significant finding of divergence for the child-directed register. For the foreign vowels, in addition to an overall finding of convergence, we saw a convergence only in the child-directed register and in the auditory modality.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
Euclidean Distance F1+F2+F3	English-like	✗	✗	D	✗	✗
	Foreign	✓	✗	✓	✓	✗

Table 6.18. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1+F2+F3) in male child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for Euclidean Distance (F1+F2+F3) for the male participants is presented in Table 6.19. Table 6.19 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effects of Subject [$\chi^2(1) = 0.00, p = 0.99$] and Word [$\chi^2(1) = 0.00, p = 1.00$] did not significantly reduce model fit.

Euclidean Distance F1+F2+F3 Males	Parameter estimates		Wald's test		$\Delta(-2\Delta)$ -test			Partial Pseudo- R^2
	Odds (B)	S.E.	Z	p_z	c^2	df	p	
Register	-0.27	0.28	-0.97	0.333	8.24	3	0.04139	0.028
Modality	0.02	0.24	0.07	0.948	4.10	3	0.251	0.029
Carrier Type	0.05	0.21	0.25	0.803	5.16	4	0.2713	0.031
Vowel Type	-0.22	0.22	-1.03	0.305	10.40	4	0.03428	0.021
Register X Carrier Type	0.18	0.28	0.65	0.514	0.43	1	0.5139	0.038
Modality X Carrier Type	-0.33	0.23	-1.39	0.164	2.03	1	0.1546	0.036
Register X Vowel Type	0.64	0.26	2.41	0.016*	5.81	1	0.01598	0.028
Modality X Vowel Type	0.35	0.22	1.57	0.116	2.47	1	0.1157	0.034
Carrier Type X Vowel Type	0.16	0.20	0.79	0.431	0.63	1	0.4264	0.038

Table 6.19. Summary of results from the mixed effects model for Euclidean Distance (F1+F2+F3) in pre-exposure compared to post-exposure productions for male child participants.

In the mixed effects model, there was a significant interaction of Register X Vowel Type. Post-hoc tests revealed significant pairwise comparisons of foreign and English-like vowels in the child-directed register, $p = 0.046$, in addition to foreign vowels between the two registers, $p = 0.043$ (and of the child:foreign – adult:English-like comparison, $p = 0.010$). Participants in the child-directed register imitated foreign vowels significantly more than English-like vowels in the child-directed register, or either vowel type in the adult-directed register.

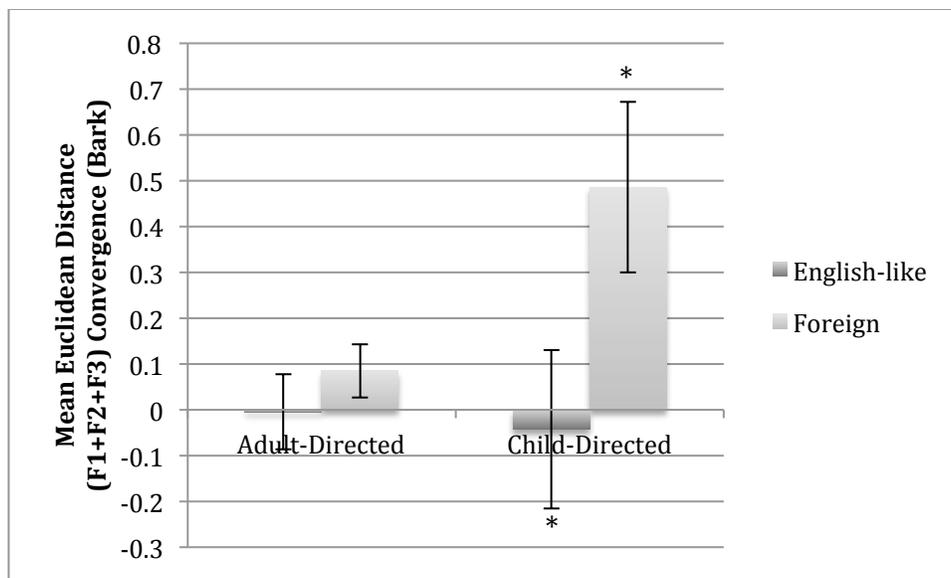


Figure 6.16. Results for convergence in Euclidean Distance ($F1 + F2 + F3$) by male child participants for each vowel type by register of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Euclidean Distance ($F1 + F2$):

For the Euclidean Distance ($F1+F2$) measure, the only significant finding we had for male participants was a finding of convergence to foreign vowels in the audiovisual modality.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
Euclidean Distance F1+F2	English-like	✗	✗	✗	✗	✗
	Foreign	✗	✗	✗	✗	✓

Table 6.20. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1+F2) in male child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for Euclidean Distance (F1+F2) for the male participants is presented in Table 6.21. Table 6.21 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effects revealed no significant differences in the model fit [Subject: $\chi^2(1) = 0.00, p = 1.0$; Word: $\chi^2(1) = 0.00, p = 1.0$]. Overall, there were no significant main effects or interactions.

Euclidean Distance F1+F2 Males	Parameter estimates		Wald's test		$\Delta(-2\Delta)$ -test			Partial Pseudo-R ²
	Odds (B)	S.E.	Z	p _z	c ²	df	p	
Register	-0.15	0.30	-0.52	0.607	6.78	3	0.07912	0.011
Modality	0.16	0.25	0.66	0.509	4.75	3	0.1907	0.023
Carrier Type	0.08	0.23	0.34	0.733	2.56	4	0.6345	0.023
Vowel Type	-0.14	0.24	-0.59	0.558	6.50	4	0.1649	0.027
Register X Carrier Type	0.13	0.30	0.43	0.667	0.20	1	0.6571	0.032
Modality X Carrier Type	-0.28	0.25	-1.11	0.268	1.24	1	0.265	0.030
Register X Vowel Type	0.54	0.29	1.88	0.061	3.57	1	0.05898	0.025
Modality X Vowel Type	0.31	0.24	1.30	0.194	1.71	1	0.1905	0.029
Carrier Type X Vowel Type	0.08	0.23	0.36	0.721	0.11	1	0.7455	0.029

Table 6.21. Summary of results from the mixed effects model for Euclidean Distance (F1+F2) in pre-exposure compared to post-exposure productions for male participants.

Summary: Global Phonetic Measures

Between the two global phonetic measures, we saw a number of differences, which we can attribute to the presence or absence of F3 (without F3, there was no longer significant overall foreign vowel convergence or divergence in child-directed speech, etc.). These F3 effects will be discussed in the section on individual phonetic measures.

First examining register, there were only observable register effects for the measure including F3. For the overall global phonetic measure, the results showed significant divergence for English-like vowels in the child-directed register, but no effects for the adult-directed register. This is in contrast to what is seen for the foreign vowels, which is significant convergence in the child-directed register and once again no effects for the adult-directed register. The mixed effects model supports the difference only for foreign vowels: there was significantly more convergence for foreign vowels in the child-directed register. No register effects were seen in the Euclidean Distance (F1+F2) measure.

When we look at the modality effects, we see no more clarity, and directly opposing results, but only for foreign vowels (there were no differences for English-like vowels). For foreign vowels in the Euclidean Distance (F1+F2+F3) measure, we see convergence only in the auditory modality, and no effect for the audiovisual modality. For foreign vowels in the measure without F3, we see convergence in the audiovisual modality, but no effect in the auditory modality. It appears as though the addition of F3 largely skews the results, so we will pay close attention to this measure in the following subsection, but these results suggest that the audiovisual modality will play a role in F3 imitation.

Individual Phonetic Measures:

F1:

Recall for F1, we only found a significant overall convergence finding for foreign vowels. However, for male participants, overall, there was only a significant convergence finding for English-like vowels. Looking at the subsets divided by register, we found that males only significantly converged to English-like vowels in the child-directed register. The modality subsets were somewhat clearer: males converged to both vowel types in the audiovisual modality (and not the auditory).

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
F1	English-like	✓	✗	✓	✗	✓
	Foreign	✗	✗	✗	✗	✓

Table 6.22. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F1 in male child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for the first formant for the male participants is presented in Table 6.23. Table 6.23 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effects revealed no significant differences in model fit [Subject: $\chi^2(1) = 0.91, p = 0.340$; Word: $\chi^2(1) = 0.00, p = 0.991$].

<i>F1 Males</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Partial Pseudo -R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	0.07	0.32	0.22	0.824	3.89	3	0.2735	0.112
Modality	0.11	0.26	0.41	0.682	12.66	3	0.005446	0.106
Carrier Type	0.05	0.21	0.24	0.807	2.51	4	0.6433	0.050
Vowel Type	-0.32	0.21	-1.52	0.129	8.62	4	0.07138	0.042
Register X Carrier Type	0.15	0.27	0.54	0.593	0.30	1	0.5837	0.053
Modality X Carrier Type	-0.24	0.23	-1.04	0.301	1.08	1	0.299	0.052
Register X Vowel Type	0.31	0.26	1.15	0.249	1.36	1	0.2432	0.052
Modality X Vowel Type	0.63	0.22	2.89	0.004*	8.48	1	0.0036	0.042
Carrier Type X Vowel Type	0.00	0.20	0.01	0.992	0.00	1	0.9463	0.054

Table 6.23. Summary of results from the mixed effects model for F1 in pre-exposure compared to post-exposure productions for male participants.

Overall, there were no significant main effects but there was a significant interaction of Modality X Vowel Type. Post-hoc tests revealed that the only significant pairwise comparison was between foreign vowels in the two modalities; there was significantly more convergence on foreign vowels in the audiovisual modality ($p = 0.006$).

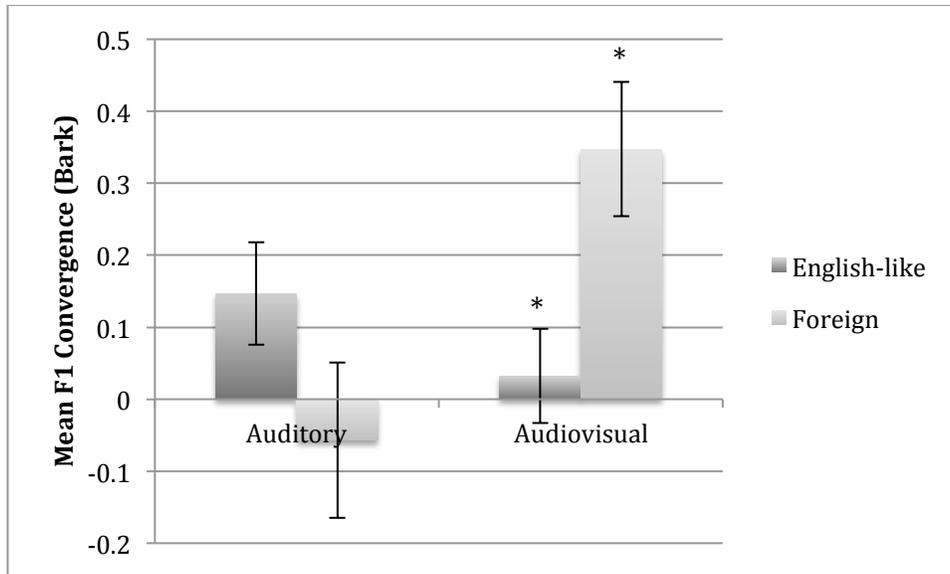


Figure 6.17. Results for convergence in F1 by male child participants to each vowel type by modality of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

F2:

None of the male subsets showed any significant convergence (or divergence) findings, and therefore, we will not conduct further analysis on this measure.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
F2	English-like	*	*	*	*	*
	Foreign	*	*	*	*	*

Table 6.24. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F2 in male child subjects. The full statistical output is shown in the appendix.

F3:

Recall that there were no overall convergence findings for F3, but we mentioned that particular subsets of the data did show significant convergence findings. One of these subsets was that male participants showed significant convergence to F3 for foreign vowels. Looking further into the male data, we found convergence for both vowel types in the child-directed register and in the auditory modality. We saw no significant F3 results for males in either the audiovisual modality or adult-directed register.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
F3	English-like	✗	✗	✓	✓	✗
	Foreign	✓	✗	✓	✓	✗

Table 6.25. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F3 in male child subjects. The full statistical output is shown in the appendix.

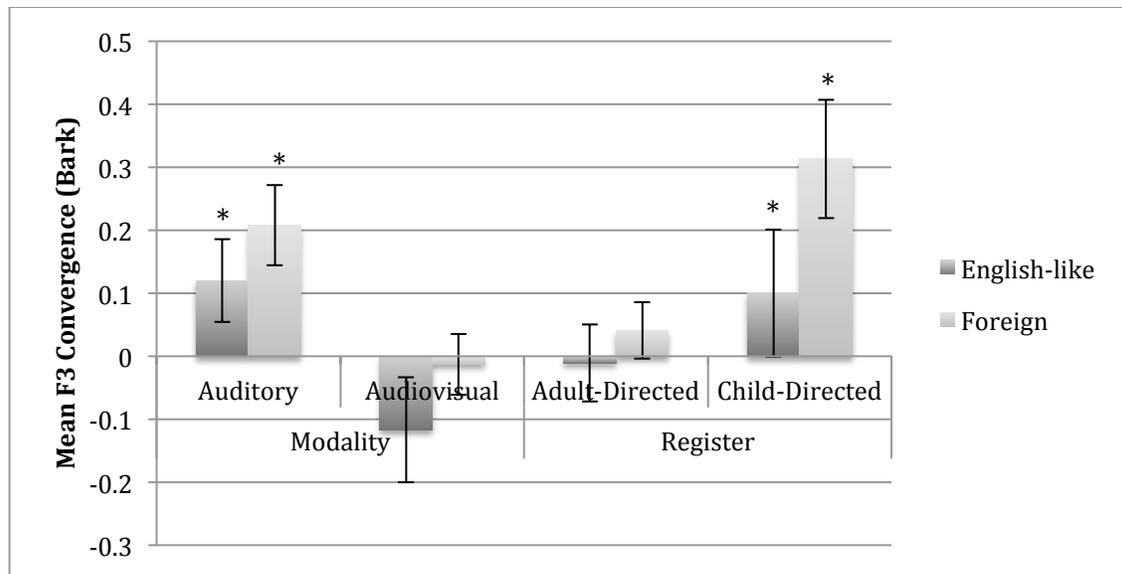


Figure 6.18. Results for convergence in F3 by male child participants to each vowel type by modality and register of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

The final output for the mixed effects model including all factors and interactions for the third formant for the male participants is presented in Table 6.26. Table 6.26 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effects of Subject [$\chi^2(1) = 0.00, p = 0.98$] and Word [$\chi^2(1) = 0.00, p = 1$] did not significantly reduce model fit. Overall, there were no significant main effects or interactions.

<i>F3 Males</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-0.17	0.20	-0.88	0.379	3.47	3	0.3248	0.045
Modality	-0.15	0.17	-0.92	0.356	5.39	3	0.1451	0.050
Carrier Type	0.02	0.14	0.11	0.911	7.19	4	0.1263	0.039
Vowel Type	-0.10	0.15	-0.67	0.503	4.76	4	0.3123	0.042
Register X Carrier Type	0.22	0.19	1.17	0.242	1.42	1	0.233	0.048
Modality X Carrier Type	-0.14	0.16	-0.92	0.357	0.84	1	0.3594	0.049
Register X Vowel Type	0.22	0.18	1.21	0.226	1.52	1	0.2177	0.048
Modality X Vowel Type	0.11	0.15	0.72	0.475	0.54	1	0.4633	0.050
Carrier Type X Vowel Type	0.15	0.14	1.07	0.285	0.54	1	0.4633	0.048

Table 6.26. Summary of results from the mixed effects model for F3 in pre-exposure compared to post-exposure productions for male participants.

Summary: Individual Phonetic Measures

Our analysis of the individual phonetic measures will include just F1 and F3, not F2. There were no significant effects of convergence or divergence for F2 for male participants, so we will leave that measure out. For both F1 and F3, we saw register effects favoring imitation in the child-, but not adult-, directed register. However, we saw opposite modality effects for these two measures.

First, looking at the clearer register effects, for F1, we only saw significant convergence in the child-directed register (just for English-like vowels). For F3, we found significant convergence effects for both vowel types only in the child-directed register. We found no significant differences in the mixed effect models, however.

Next, looking at the modality effects, we found results favoring increased imitation in the audiovisual modality for F1, but results favoring increased imitation in the auditory modality for

F3 (note that this is directly contrastive with the hypotheses for which formant will be better imitated in each modality, which will be discussed in the discussion at the end of this chapter). For F1, we found significant convergence results only in the audiovisual modality for both vowel types, and additionally, we found significantly more imitation in the audiovisual modality than the auditory modality for foreign vowels. For F3, however, we found convergence in the auditory modality, but not the audiovisual modality, for both vowel types.

Summary: Male Child Participants

In the data for male participants, we saw a difference in convergence based on measure type. In this pre-exposure/post-exposure comparison, we only saw convergence for the global phonetic measures and the individual phonetic measures. In the global measures, there were significant findings of divergence only (there was also one finding of divergence in male data for the overall Euclidean Distance measure).

While there were many different register and modality effects, no clear pattern emerged. The table below displays a summary of the results for the male participants. First taking a look at the register results, we can see a split in the results for the global measures, with results favoring increased convergence for the adult-directed register in duration, and the child-directed register for f_0 . For the global phonetic measures, there are differences in which register is favored based on vowel type, favoring adult-directed speech for English-like vowels and child-directed speech for foreign vowels (these results are only for the measure including F3). For the individual phonetic measures, however, the only results that significantly show a preference all favor increased imitation in the child-directed register.

	Measure	Vowel Type	Register Results	Modality Results
Global Measures	Duration	English-like	ADS +	AV +
		Foreign	ADS +	none
	f0	English-like	CDS +	AV +
		Foreign	CDS +	none
Global Phonetic Measures	Euclidean Distance F1+F2+F3	English-like	ADS +	none
		Foreign	CDS +	Aud +
	Euclidean Distance F1+F2	English-like	none	none
		Foreign	none	AV +
Individual Phonetic Measures	F1	English-like	CDS +	AV +
		Foreign	none	AV +
	F2	English-like		
		Foreign		
	F3	English-like	CDS +	Aud +
		Foreign	CDS +	Aud +

Table 6.27. Summary of register and modality findings by vowel type in each measure for male child participants in the pre-exposure/post-exposure comparison.

As for the modality results, the results are much more clear. While there are not very many results showing modality differences, they can more easily be summarized. For all measures except F3 and the Euclidean Distance measure including F3, there is an advantage to the audiovisual modality. For the F3 measure and the Euclidean Distance Measure with F3, there is an advantage to the auditory modality. Since F3 is a cue to vowel rounding, which is visually salient, it is surprising that we find the results patterning this way, as we would expect an advantage to the addition of the visual modality for this cue. We will discuss possible explanations for this in the overall summary at the end of this chapter.

Female Child Participants:

The following sections will present the data from female child participants in the pre-exposure/post-exposure comparison. First there will be a discussion of the more global measures of convergence (the Euclidean distance measures) followed by the formant measures, and then finally, the f_0 and duration measures. At the end of this section, before moving on a comparison between male and female child participants, we will summarize the results for this subject group.

The numbers of participants for these groups was as follows: the auditory exposure, adult-directed register group consisted of 6 female participants, the audiovisual exposure, adult-directed register group consisted of 8 female participants, and the auditory exposure, child-directed register group consisted of 2 female participants.

Global Measures

Duration:

For female participants, there was significant divergence on English-like vowels overall. For the results by modality, we found that females significantly diverged in the auditory modality, but showed no significant convergence or divergence for the audiovisual modality. For register, we saw significant divergence in the adult-directed register, but only for English-like vowels. There were no results for English-like vowels in the child-directed register, or for foreign vowels in either register.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
Duration	English-like	D	D	×	D	×
	Foreign	×	×	×	D	×

Table 6.28. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for duration in female child subjects. The full statistical output is shown in the appendix.

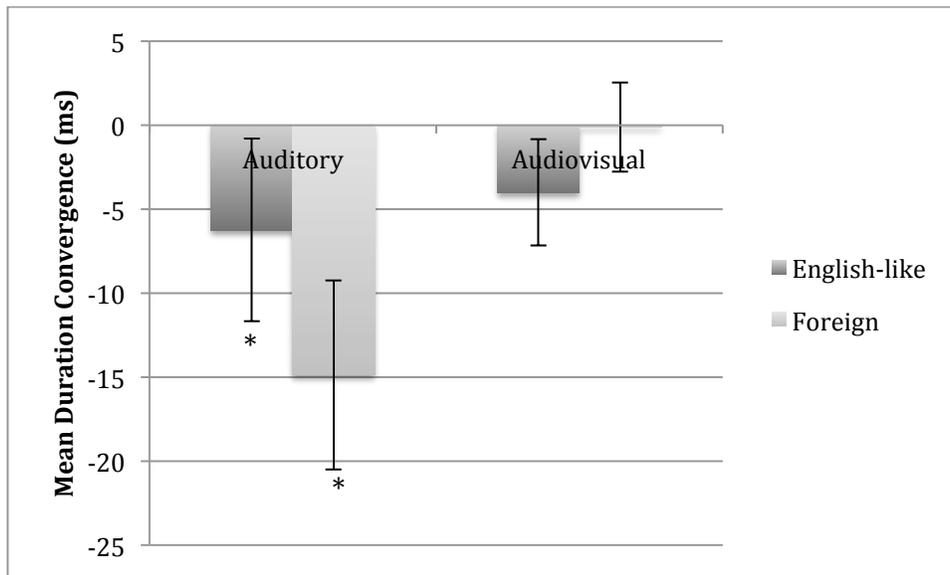


Figure 6.19. Results for convergence in duration by female child participants to each vowel type by modality of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

The final output for the mixed effects model including all factors and interactions for duration for the female participants is presented in Table 6.29. Table 6.29 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or

interaction. Removing the random effect of Subject, but not Word reduced model fit significantly

[Subject: $\chi^2(1) = 0.74, p = 0.39$; Word: $\chi^2(1) = 0.00, p = 1.0$].

<i>Duration Females</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-19.94	16.53	-1.21	0.228	16.80	3	0.00	0.029
Modality	1.27	10.98	0.12	0.908	0.63	3	0.89	0.037
Carrier Type	1.24	9.65	0.13	0.898	7.15	4	0.13	0.031
Vowel Type	4.54	9.83	0.46	0.644	9.49	4	0.05	0.030
Register X Carrier Type	36.06	15.35	2.35	0.019*	5.53	1	0.02	0.033
Modality X Carrier Type	2.96	10.22	0.29	0.772	0.09	1	0.77	0.038
Register X Vowel Type	-39.16	14.68	-2.67	0.008*	7.12	1	0.01	0.031
Modality X Vowel Type	2.34	9.84	0.24	0.812	0.06	1	0.81	0.038
Carrier Type X Vowel Type	-4.41	9.63	-0.46	0.647	0.21	1	0.65	0.037

Table 6.29. Summary of results from the mixed effects model for duration in pre-exposure compared to post-exposure productions for female participants.

Overall there were no significant main effects but there were significant interactions of Register X Carrier Type and Register X Vowel Type. Post-hoc testing revealed significantly more divergence for CV words in the child-directed register, $p < 0.001$, and within the child-directed register, significantly more divergence for CV words than CVC words, $p = 0.048$ (as well as a significant difference between adult:CVC and child:CV, $p < 0.001$).

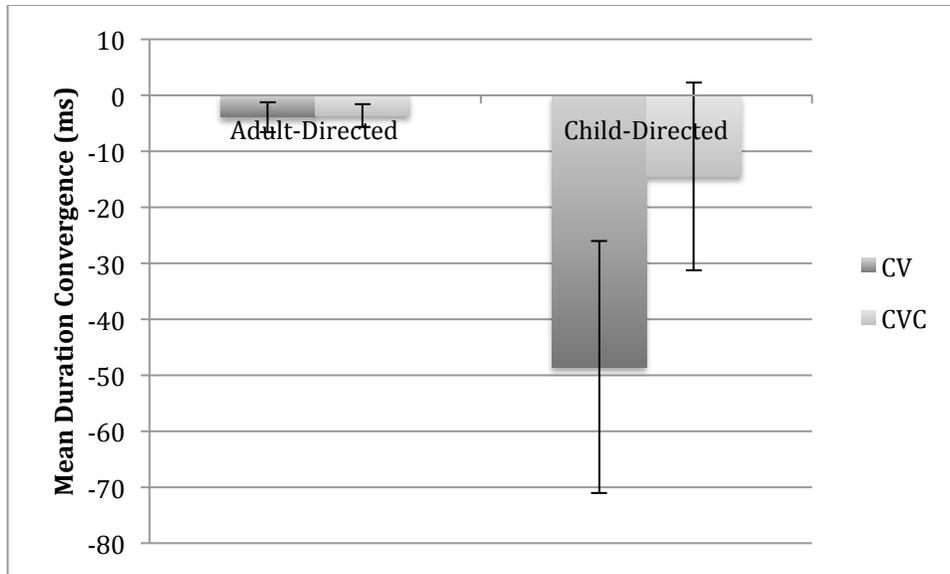


Figure 6.20. Results for convergence in duration by female child participants to each carrier type by register of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

As for the interaction of Register X Vowel Type, post-hoc testing revealed there was significantly more divergence for foreign vowels in the child-directed register, compared to English-like vowels in the same register ($p = 0.017$), foreign vowels in the adult-directed register ($p < 0.001$), and English-like vowels in the adult-directed register ($p < 0.001$).



Figure 6.21. Results for convergence in duration by female child participants to each vowel type by register of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

f0:

For f_0 , female participants showed divergence for English-like vowels overall, and for the child-directed register. They also showed divergence (like when combining both genders) to the auditory, not audiovisual, modality.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
f0	English-like	D	x	D	D	x
	Foreign	x	x	x	D	x

Table 6.30. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for f_0 in female child subjects. The full statistical output is shown in the appendix.

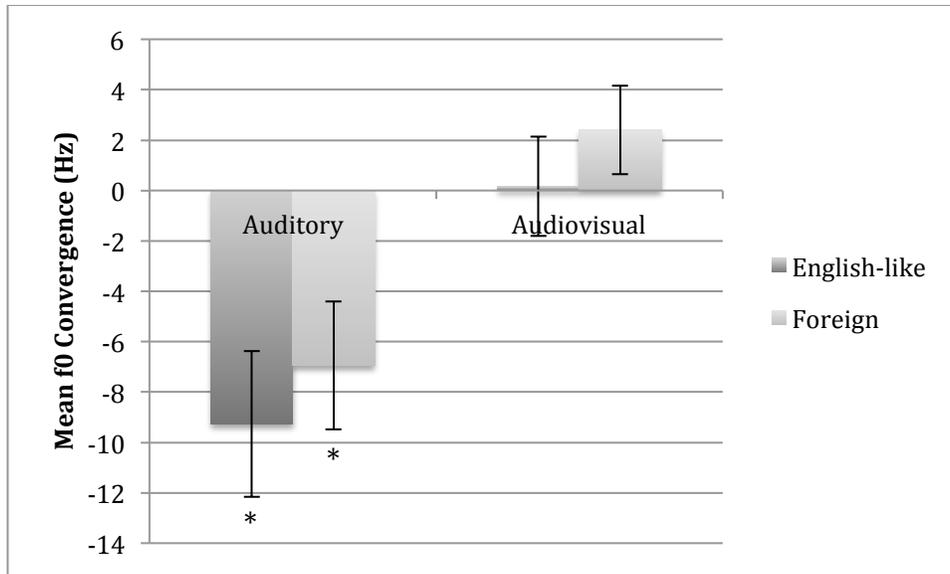


Figure 6.22. Results for convergence in f0 by female child participants to each vowel type by modality of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

The final output for the mixed effects model including all factors and interactions for f0 for the female participants is presented in Table 6.31. Table 6.31 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effect of Subject [$\chi^2(1) = 83.86, p < 0.01$] but not Word [$\chi^2(1) = 0.00, p = 1.0$] significantly reduced model fit. For the fundamental frequency variable, no significant main effects or interactions were observed.

<i>f</i> 0 Females	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>Δ(-2Δ)-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-9.22	13.48	-0.68	0.494	1.27	3	0.74	0.134
Modality	8.07	8.93	0.90	0.367	1.49	3	0.69	0.135
Carrier Type	0.25	4.85	0.05	0.959	1.12	4	0.89	0.132
Vowel Type	3.03	4.94	0.61	0.539	2.22	4	0.70	0.133
Register X Carrier Type	6.42	7.58	0.85	0.397	0.74	1	0.39	0.135
Modality X Carrier Type	0.45	5.05	0.09	0.929	0.01	1	0.91	0.135
Register X Vowel Type	6.02	7.27	0.83	0.407	0.68	1	0.41	0.134
Modality X Vowel Type	1.13	4.86	0.23	0.816	0.06	1	0.80	0.135
Carrier Type X Vowel Type	-2.79	4.89	-0.57	0.568	0.31	1	0.58	0.134

Table 6.31. Summary of results from the mixed effects model for F0 in pre-exposure compared to post-exposure productions for female participants.

Summary: Global Measures

The global measures in the pre-exposure/post-exposure comparison for female subjects both showed significant divergence findings overall to each vowel type. For duration, there seems to be a difference in register effects depending on vowel type; while the only significant finding in the convergence analysis was for divergence in the adult-directed register (English-like vowels only), the mixed effects models showed significantly more divergence in the child-directed register, although this was evident in the CV syllable types and in foreign vowels only. For f0, however, the pattern was more consistent, favoring increased imitation in adult-directed speech; the only significant divergence (or convergence) finding was for divergence to speech in the child-directed register for English-like vowels.

Both global measures showed a slight preference for the audiovisual modality. For the duration measure, there was significant divergence to both vowel types only in the auditory

modality. In the f0 measure, there was also significant divergence to both types only in the auditory modality, but also, an analysis of the means showed that overall, the mean trend for f0 in the audiovisual modality was that of convergence (although this was not significantly greater than zero).

Global Phonetic Measures

Euclidean Distance (F1 + F2 + F3):

Looking at the female participants, we saw significant findings of convergence for both vowel types. Female participants showed significant convergence to English-like vowels in the child-directed register, and to foreign vowels in the adult-directed register. As for modality, females only showed significant convergence to the auditory modality.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
Euclidean Distance F1+F2+F3	English-like	✓	✗	✓	✓	✗
	Foreign	✓	✓	✗	✓	✗

Table 6.32. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1 + F2 + F3) in female child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for Euclidean Distance (F1+F2+F3) for the female participants is presented in Table 6.33. Table 6.33 also presents results from likelihood ratio tests comparing the full model with models

excluding each factor or interaction. Removing the random effect of Subject [Subject: $\chi^2(1) = 29.79, p < 0.01$] but not Word [$\chi^2(1) = 2.19, p = 0.139$] significantly reduced model fit.

<i>Euclidean Distance F1+F2+F3 Females</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2A)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	1.24	0.48	2.60	0.010*	7.20	3	0.07	0.156
Modality	0.16	0.29	0.56	0.576	1.18	3	0.76	0.155
Carrier Type	0.41	0.17	2.50	0.013*	6.87	4	0.14	0.160
Vowel Type	0.44	0.17	2.51	0.012*	6.56	4	0.16	0.159
Register X Carrier Type	-0.32	0.34	-0.95	0.341	0.89	1	0.34	0.154
Modality X Carrier Type	-0.13	0.16	-0.79	0.428	0.64	1	0.42	0.156
Register X Vowel Type	-0.05	0.32	-0.15	0.883	0.06	1	0.80	0.156
Modality X Vowel Type	-0.13	0.15	-0.86	0.393	0.73	1	0.39	0.155
Carrier Type X Vowel Type	-0.35	0.18	-1.93	0.054	3.74	1	0.05	0.160

Table 6.33. Summary of results from the mixed effects model for Euclidean Distance (F1+F2+F3) in pre-exposure compared to post-exposure productions for female participants.

Overall, there were significant main effects of Register, Carrier Type, and Vowel Type, but no significant interactions. Post-hoc testing only confirmed a significant main effect of register ($p < 0.001$), favoring more imitation in the child-directed register (Figure 6.23).

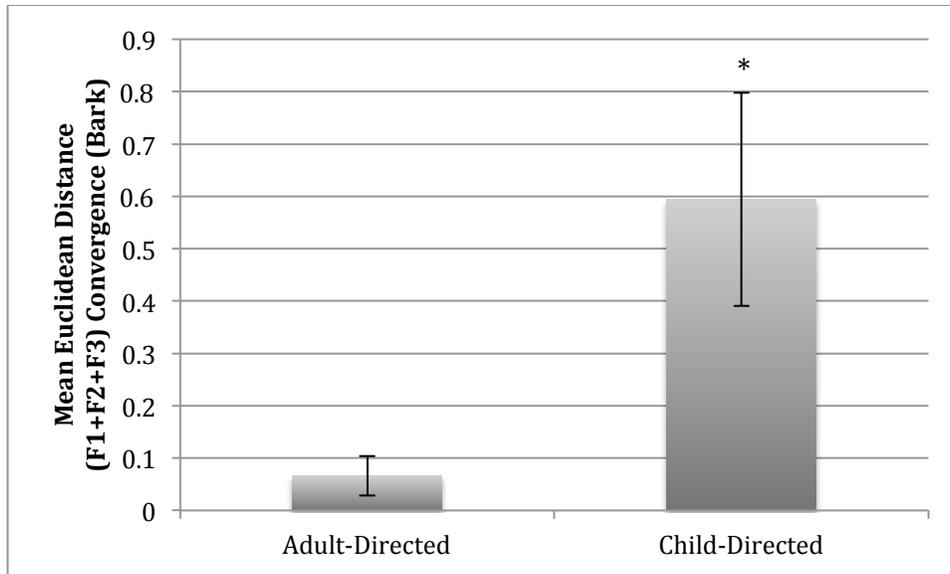


Figure 6.23. Comparison of pre-exposure and post-exposure phonetic distance in each register for female child subjects in the Euclidean distance convergence (F1/F2/F3) measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Euclidean Distance (F1 + F2):

For the Euclidean Distance (F1+F2) measure, female participants showed convergence overall to both vowel types, in the audiovisual modality to both vowel types. In the child-directed register, female participants only showed significant convergence to English-like vowels. In the adult-directed register, they converged to both vowel types.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
Euclidean Distance F1+F2	English-like	✓	✓	✓	✗	✓
	Foreign	✓	✓	✗	✗	✓

Table 6.34. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for Euclidean Distance (F1 + F2) in female child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for Euclidean Distance (F1+F2) for the female participants is presented in Table 6.35. Table 6.35 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Overall, this model accounts for 0.22% of the variance in this data set. Removing the random effects of Subject [Subject: $\chi^2(1) = 3.15, p = 0.07$] and Word [$\chi^2(1) = 0.00, p = 1.0$] did not significantly reduce model fit.

<i>Euclidean Distance F1+F2 Females</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2A)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	1.19	0.34	3.50	0.001*	14.32	3	0.00	0.083
Modality	0.15	0.20	0.75	0.451	0.90	3	0.82	0.080
Carrier Type	0.33	0.17	1.93	0.054	5.89	4	0.21	0.086
Vowel Type	0.48	0.17	2.77	0.006*	12.96	4	0.01	0.084
Register X Carrier Type	-0.35	0.30	-1.16	0.246	1.36	1	0.24	0.046
Modality X Carrier Type	0.05	0.18	0.25	0.800	0.07	1	0.79	0.047
Register X Vowel Type	-0.91	0.29	-3.18	0.002*	9.16	1	0.00	0.078
Modality X Vowel Type	-0.21	0.17	-1.24	0.217	0.74	1	0.39	0.080
Carrier Type X Vowel Type	-0.34	0.17	-2.00	0.046*	2.46	1	0.12	0.087

Table 6.35. Summary of results from the mixed effects model for Euclidean Distance (F1+F2) in pre-exposure compared to post-exposure productions for female participants.

Overall, there were significant main effects of Register and Vowel Type, and significant interactions of Register X Vowel Type and Carrier Type X Vowel Type (although the latter interaction was not significant in likelihood ratio tests). Post-hoc testing revealed that the only main effect that was significant was the effect of register ($p = 0.012$); there was significantly more imitation in the child-directed register. Post-hoc testing on the interaction of Register X Vowel Type revealed that the register differences were evident only for English-like vowels ($p < 0.001$). Post-hoc testing also revealed that English-like vowels in the child-directed register were also imitated more than foreign vowels in the same register ($p = 0.049$) and foreign vowels in the adult-directed register ($p = 0.006$).

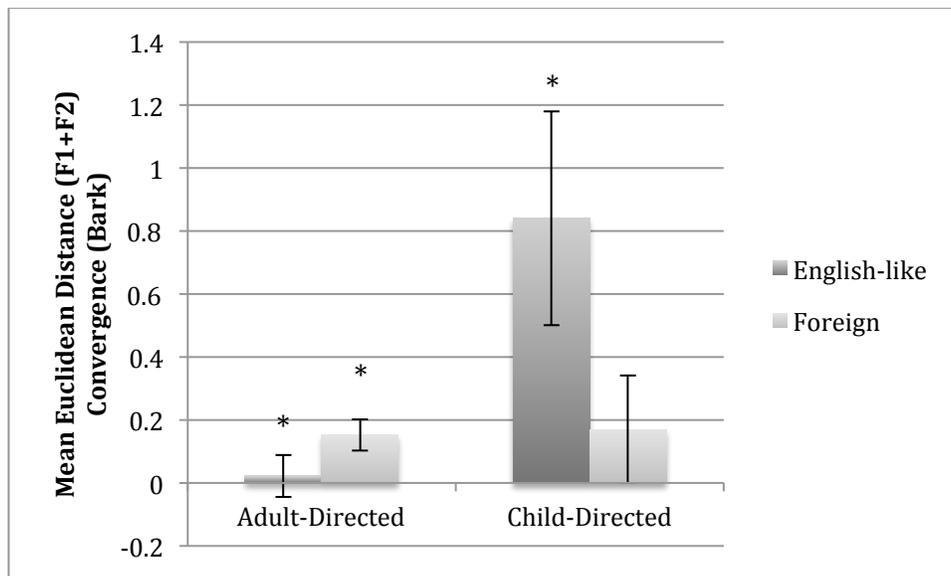


Figure 6.24. Comparison of pre-exposure and post-exposure phonetic distance in each vowel type by register for female child subjects in the Euclidean distance convergence (F1/F2) measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

As for the interaction of Carrier Type X Vowel Type, post-hoc testing showed no significant pairwise effects for this comparison.

Summary: Global Phonetic Measures

The two global phonetic measures showed the same pattern for register and an opposite pattern for modality in female participants. In both measures, there were overall findings of convergence to both vowel types, which is in contrast to what we saw for the global measures. Starting with register-related findings, for both measures, female participants showed significant convergence in the child-directed register for English-like vowels (for the measure with just F1 and F2, there was also convergence to adult-directed speech for English-like vowels), and the adult-directed register for foreign vowels. Also, the mixed effects models showed register findings as well; for the Euclidean Distance (F1+F2+F3) measure, there was significantly more convergence in the child-directed register, and for the Euclidean Distance (F1+F2) measure, there was significantly more convergence for the child-directed register, but only for English-like vowels. There appears to be a significant correlation between register and vowel type for these results, favoring imitation of English-like vowels in child-directed speech and foreign vowels in adult-directed speech (possible reasoning for this will be discussed in overall summary sections).

Next, looking at the results by modality, we saw contrasting effects for the two global phonetic measures. For the overall measure including F3, we found convergence only in the auditory modality, but for the measure without F3, we found convergence only in the audiovisual modality. From these results, we can hypothesize that F3 will be much better imitated in the

auditory modality in the individual phonetic measure results, which we will look at in the next subsection.

Individual Phonetic Measures

F1:

Female participants showed significant convergence overall, for adult-directed speech, and in the audiovisual modality for foreign vowels. For English-like vowels, there were no significant findings of convergence for child participants but there was divergence in the child-directed register.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
F1	English-like	✗	✗	D	✗	✗
	Foreign	✓	✓	✗	✗	✓

Table 6.36. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F1 in female child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for the first formant for the female participants is presented in Table 6.37. Table 6.37 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effect of Subject [$\chi^2(1) = 35.36, p < 0.01$] but not Word [$\chi^2(1) = 0.00, p = 0.99$] significantly reduced model fit.

<i>F1 Females</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\lambda)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>	<i>df</i>	<i>p</i>	
Register	-0.34	0.33	-1.05	0.292	2.62	3	0.45	0.090
Modality	0.05	0.22	0.24	0.813	0.99	3	0.80	0.090
Carrier Type	0.10	0.14	0.73	0.468	2.27	4	0.69	0.089
Vowel Type	0.37	0.14	2.61	0.009*	16.31	4	0.00	0.078
Register X Carrier Type	-0.13	0.23	-0.55	0.581	0.30	1	0.58	0.090
Modality X Carrier Type	0.08	0.15	0.52	0.606	0.27	1	0.60	0.090
Register X Vowel Type	0.13	0.22	0.58	0.559	0.35	1	0.55	0.090
Modality X Vowel Type	-0.13	0.14	-0.93	0.354	0.87	1	0.35	0.090
Carrier Type X Vowel Type	-0.12	0.14	-0.83	0.407	0.70	1	0.40	0.090

Table 6.37. Summary of results from the mixed effects model for F1 in pre-exposure compared to post-exposure productions for female participants.

Overall, there was a significant main effect of Vowel Type but no significant interactions.

There was significantly more imitation for foreign vowels.

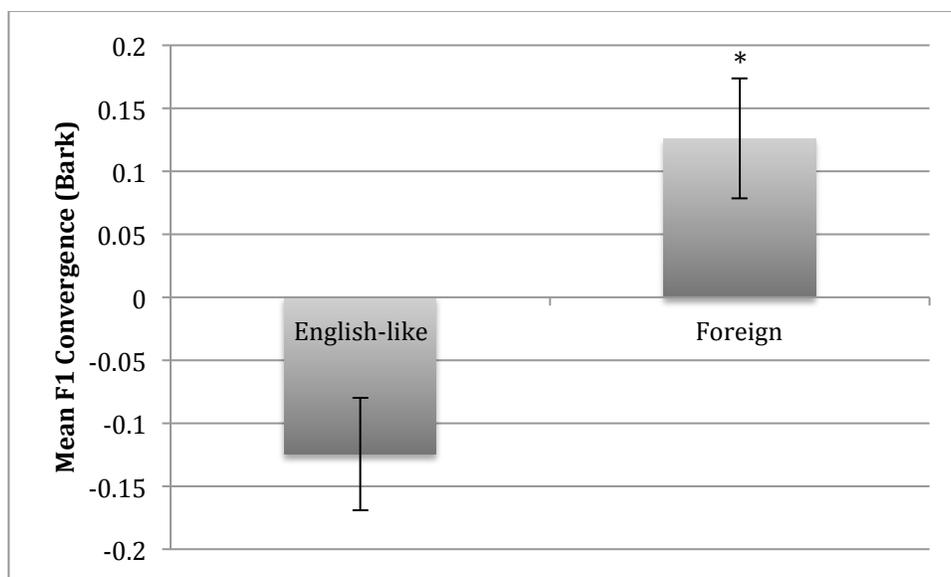


Figure 6.25. Comparison of pre-exposure and post-exposure phonetic distance for each type of vowel for female child subjects in the F1 measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

F2:

In the F2 measurement for female participants, there was only one significant finding of convergence: for the English-like vowels in the child-directed register.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
F2	English-like	×	×	✓	×	×
	Foreign	×	×	×	×	×

Table 6.38. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F2 in female child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for the second formant for the female participants is presented in Table 6.39. Table 6.39 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effect of Subject [$\chi^2(1) = 16.28, p < 0.01$] but not Word [$\chi^2(1) = 0.15, p = 0.70$] significantly reduced model fit.

<i>F2 Females</i>	<i>Parameter estimates</i>		<i>Wald's test</i>		<i>$\Delta(-2\Delta)$-test</i>			<i>Partial Pseudo-R²</i>
	<i>Odds (B)</i>	<i>S.E.</i>	<i>Z</i>	<i>p_z</i>	<i>c²</i>		<i>p</i>	
Register	1.15	0.39	2.95	0.003*	11.12	3	0.01	0.090
Modality	0.10	0.24	0.43	0.668	0.18	3	0.98	0.090
Carrier Type	0.34	0.17	1.97	0.049*	6.13	4	0.19	0.094
Vowel Type	0.36	0.18	2.00	0.046*	11.58	4	0.02	0.093
Register X Carrier Type	-0.17	0.30	-0.57	0.567	0.35	1	0.56	0.090
Modality X Carrier Type	-0.04	0.18	-0.20	0.841	0.05	1	0.83	0.090
Register X Vowel Type	-0.70	0.28	-2.46	0.014*	6.05	1	0.01	0.088
Modality X Vowel Type	-0.09	0.17	-0.55	0.581	0.30	1	0.58	0.090
Carrier Type X Vowel Type	-0.41	0.18	-2.28	0.023*	5.24	1	0.02	0.095

Table 6.39. Summary of results from the mixed effects model for F2 in pre-exposure compared to post-exposure productions for female participants.

Overall, there were significant main effects of Register, Carrier Type, and Vowel Type, and significant interactions of Register X Vowel Type, and Carrier Type X Vowel Type. Post-hoc testing revealed that the only main effect that remained significant was the effect of register; there was significantly more imitation in the child-directed register. Pairwise comparisons of the Register X Vowel Type interaction revealed that this register difference was significant only for English-like vowels ($p < 0.001$).

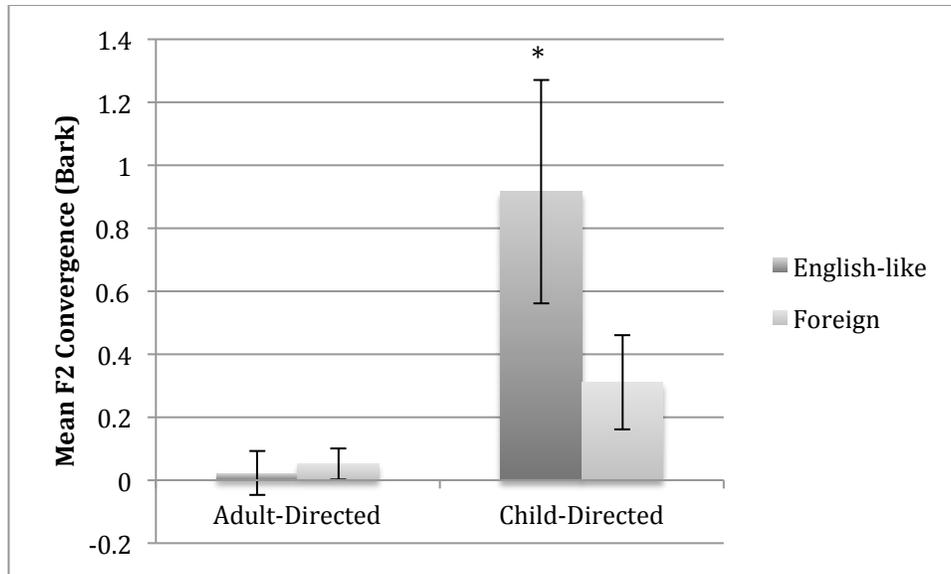


Figure 6.26. Comparison of pre-exposure and post-exposure phonetic distance for each type of vowel by register for female subjects in the F2 measurement. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

As for the interactions, post-hoc testing showed no significant pairwise effects for the Carrier Type X Vowel Type comparison.

F3:

The results for female participants in the F3 measure showed one significant result, and it was a result of divergence. Females significantly diverged in their production of F3 for foreign vowels in the child-directed register.

		Overall	Adult-Directed	Child-Directed	Auditory	Audiovisual
F3	English-like	*	*	*	*	*
	Foreign	*	*	D	*	*

Table 6.40. Results from analyses looking at whether there was any overall convergence in the comparison of pre-exposure and post-exposure productions for F3 in female child subjects. The full statistical output is shown in the appendix.

The final output for the mixed effects model including all factors and interactions for the third formant for the female participants is presented in Table 6.41. Table 6.41 also presents results from likelihood ratio tests comparing the full model with models excluding each factor or interaction. Removing the random effects of Subject [Subject: $\chi^2(1) = 2.04, p = 0.15$] and Word [$\chi^2(1) = 0.00, p = 1.0$] did not significantly reduce model fit.

F3 Females	Parameter estimates		Wald's test		$\Delta(-2\Delta)$ -test			Partial Pseudo-R ²
	Odds (B)	S.E.	Z	p _z	c ²	df	p	
Register	0.11	0.20	0.55	0.582	9.67	3	0.02	0.058
Modality	0.10	0.11	0.91	0.364	6.59	3	0.09	0.041
Carrier Type	0.15	0.10	1.55	0.121	8.76	4	0.07	0.038
Vowel Type	0.04	0.10	0.37	0.711	9.76	4	0.04	0.044
Register X Carrier Type	0.15	0.20	0.78	0.434	0.62	1	0.43	0.046
Modality X Carrier Type	-0.23	0.10	-2.24	0.025*	5.00	1	0.03	0.042
Register X Vowel Type	-0.54	0.18	-3.00	0.003*	9.43	1	0.00	0.044
Modality X Vowel Type	-0.05	0.10	-0.54	0.588	0.30	1	0.59	0.046
Carrier Type X Vowel Type	0.04	0.10	0.41	0.683	0.17	1	0.68	0.047

Table 6.41. Summary of results from the mixed effects model for F3 in pre-exposure compared to post-exposure productions for female participants.

Overall, there were no significant main effects, but there were significant interactions of Modality X Carrier Type and Register X Vowel Type. For the Modality X Carrier Type interaction, there was a significant difference only between CVC and CV vowels in the auditory modality, favoring significantly more convergence on CVC words.

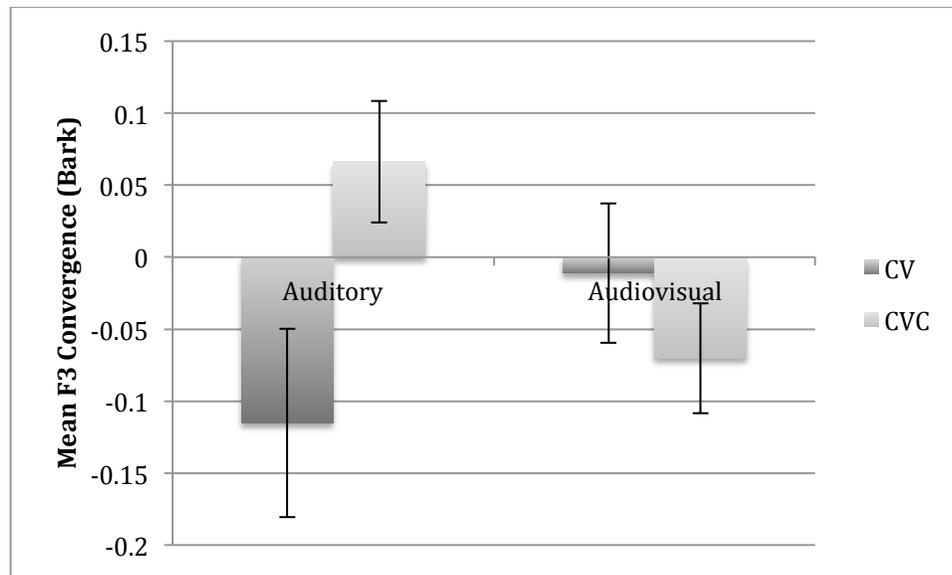


Figure 6.27. Results for convergence in F3 by female child participants to each carrier type by modality of exposure in the pre-exposure/post-exposure comparison.

For the Register X Vowel Type interaction, there was a significant register difference for foreign vowels ($p = 0.014$), favoring more divergence in the child-directed register. There was also a significant difference in the child-directed register between English-like and foreign vowels ($p = 0.008$); English-like vowels were imitated significantly more. Finally, there was also a significant pairwise difference between child:foreign and adult:English-like ($p = 0.035$).

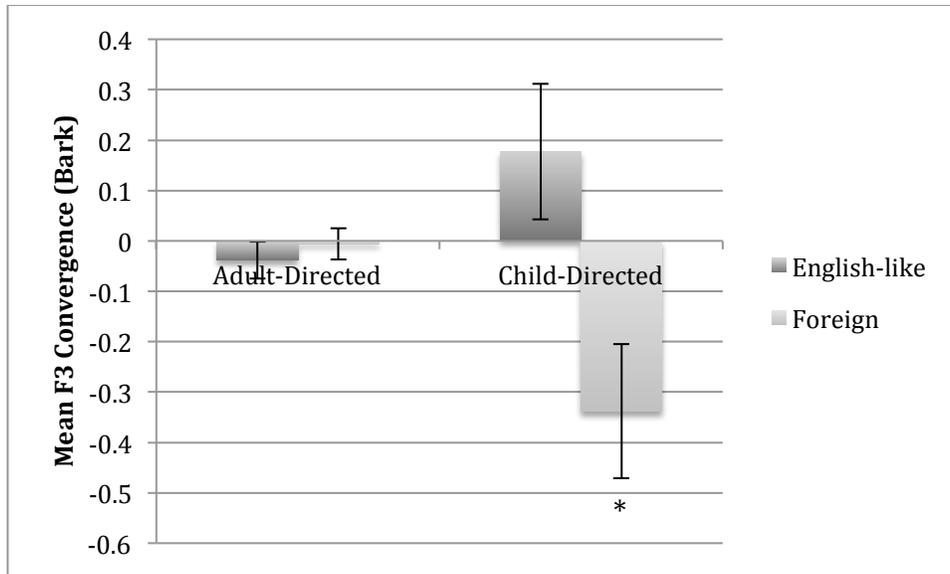


Figure 6.28. Results for convergence in F3 by female child participants to each vowel type by register of exposure in the pre-exposure/post-exposure comparison. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Summary: Individual Phonetic Measures

The individual phonetic measures did not show consistent patterning in the results. The only result that they were consistent on was that they all showed effects of register and vowel type, and in most cases, these two factors interacted.

Starting with modality, the only measure in which we saw modality effects was for F1. F2 and F3 showed no convergence in either the auditory or the audiovisual modality. The only effect for these two measures related to modality was that for F3, there was an interaction effect of modality and carrier type, showing that within the auditory modality, there was more imitation for CVC than CV words (The lack of significant results for modality in F3 is surprising, because with the global phonetic measures, we see that including F3 reverses the direction of patterning

for modality effects.). F1, however, did show a modality difference, but it was very slight; for F1, there was a significant convergence finding for foreign vowels only in the audiovisual modality. Looking at the means, overall, for F1, there was divergence (although it was not significant) in the auditory modality, but convergence in the audiovisual modality. The addition of the visual cues aided female subjects in their imitation of F1 in foreign vowels.

The register effects for the individual phonetic measures were very complex, and largely interacted with vowel type. Starting with F1, we saw a difference in convergence according to vowel type: convergence in the adult-directed register for foreign vowels, and divergence in the child-directed register for English-like vowels (and not surprisingly, considering this, overall more convergence for foreign than English-like vowels). Both of these F1 results favor increased imitation in the adult-directed register. For F2, we found the opposite pattern, which was results favoring increased imitation in the child-directed register. However, this was only for English-like vowels. For this vowel type, we found significant convergence to F2 only in the child-directed register. Mixed effect models showed additional support for this, with a finding of significantly more convergence in the child-directed register on English-like vowels. Finally, F3 additionally showed register differences, and like F1, they were in relation to foreign vowels. For foreign vowels, in the F3 measure, there was significant divergence only in the child-directed register. This measure also showed a correlation between vowel type and register in the mixed effects model: in the child-directed register, there was significantly more convergence for English-like vowels. Overall, to summarize the register effects seen in the individual phonetic measures, there appears to be a correlation with vowel type: we see increased convergence for English-like vowels in the child-directed register, and increased convergence to foreign vowels

in the adult-directed register (with the exception of divergence findings for English-like vowels in F1 in the child-directed register).

Summary: Female Child Participants

In the data for female child participants, we saw a difference in convergence based on measure type. In this pre-exposure/post-exposure comparison, we only saw convergence for the global phonetic measures and the individual phonetic measures. In the global measures, there were significant findings of divergence only. We also saw significant divergence in F3 for foreign vowels, and F1 for English-like vowels.

The table below summarizes all of the register and modality effects witnessed for female participants in the pre-exposure/post-exposure comparison. Starting first with the register results, we can see at first look that there is a lot of variation within the results. However, there is a clear trend, when looking at each vowel type separately. For English-like vowels, we notice that there are significant results for each measure, but for all measures except f0 and F2, these results point to an advantage of child-directed speech. This is in contrast to the results for foreign vowels, in which we see significant results for all measures except f0 and F2 (coincidentally, the same two measures in which we saw an opposite pattern for English-like vowels) and every one of these results points to an advantage of adult-directed speech. It is reasonable to conclude that for female participants, the child-directed register helped with imitation of English-like vowels, and the adult-directed register helped with imitation of foreign vowels. We will discuss possible explanations for this in the summary chapter.

	Measure	Vowel Type	Register Results	Modality Results
Global Measures	Duration	English-like	CDS +	AV +
		Foreign	ADS +	AV +
	f0	English-like	ADS +	AV +
		Foreign	none	AV +
Global Phonetic Measures	Euclidean Distance F1+F2+F3	English-like	CDS +	Aud +
		Foreign	ADS +	Aud +
	Euclidean Distance F1+F2	English-like	CDS +	AV +
		Foreign	ADS +	AV +
Individual Phonetic Measures	F1	English-like	ADS +	none
		Foreign	ADS +	AV +
	F2	English-like	CDS +	none
		Foreign	none	none
	F3	English-like	CDS +	none
		Foreign	ADS +	none

Table 6.42. Summary of register and modality findings by vowel type in each measure for female child participants in the pre-exposure/post-exposure comparison.

Next, looking at the modality results, we also see a somewhat clear pattern, although less complex than the results pattern for register. The modality results for female participants point to an advantage of the audiovisual modality for all measures except F3, with which there is an advantage to the auditory modality. While there were no significant results showing effects for F3, the only measures in which we saw an auditory advantage was the Euclidean Distance measure *including* F3. For the Euclidean Distance measure without F3, we saw an audiovisual advantage. Overall, it appears as though for female participants, there is a slight global advantage to the audiovisual modality in imitation, except for in the imitation of F3.

Discussion of male and female participants, Summary Pre-exposure/Post-Exposure

The male and female child participants both showed a number of significant results and some different patterning in their results. In the following chart, we show a comparison between the register effects and modality effects in the two genders. The column labeled “Match?” identifies whether the results for male and female participants were matching. From this chart, we notice one main finding, which is that for all cases in which there are results for both male and female child participants, the modality effects *all* match, but the register effects are mostly mismatching (only two of eight effects match). It appears that, while male and female participants were matched on their susceptibility to modality influences, they treated register information differently.

			Register Results			Modality Results		
	Measure	Vowel Type	M	F	Match?	M	F	Match?
Global Measures	Duration	English-like	ADS +	CDS +	No	AV +	AV +	Yes
		Foreign	ADS +	ADS +	Yes	none	AV +	-
	f0	English-like	CDS +	ADS +	No	AV +	AV +	Yes
		Foreign	CDS +	none	-	none	AV +	-
Global Phonetic Measures	Euclidean Distance F1+F2+F3	English-like	ADS +	CDS +	No	none	Aud +	-
		Foreign	CDS +	ADS +	No	Aud +	Aud +	Yes
	Euclidean Distance F1+F2	English-like	none	CDS +	-	none	AV +	-
		Foreign	none	ADS +	-	AV +	AV +	Yes
Individual Phonetic Measures	F1	English-like	CDS +	ADS +	No	AV +	none	-
		Foreign	none	ADS +	-	AV +	AV +	Yes
	F2	English-like		CDS +	-		none	-
		Foreign		none	-		none	-
	F3	English-like	CDS +	CDS +	Yes	Aud +	none	-
		Foreign	CDS +	ADS +	No	Aud +	none	-

Table 6.43. Summary of register and modality findings by vowel type in each measure for child participants in the pre-exposure/post-exposure comparison. Results are separated by gender, and the column “Match?” indicates

whether the findings for the male and female participants matched. A dash mark indicates that there was no result for one gender or the other.

Perhaps the most robust finding in this experiment is the finding that child subjects did not imitate the global measures in the pre-exposure/post-exposure comparison. While these measures have been said to be the most robustly imitated measures by adults (Mitterer, 2013), the child subjects significantly diverged on both of these measures. Instead, the imitation in these subjects in this comparison is found only in the phonetic measures (either global or individual measures).

Direction of convergence

The graphs below summarize the results for production in each of our experimental measures by child subjects according to register and modality, separated by vowel. In each of these graphs, rather than indicating the *amount* of convergence (as in the graphs previously shown), we plot mean values in the pre-exposure and post-exposure productions.

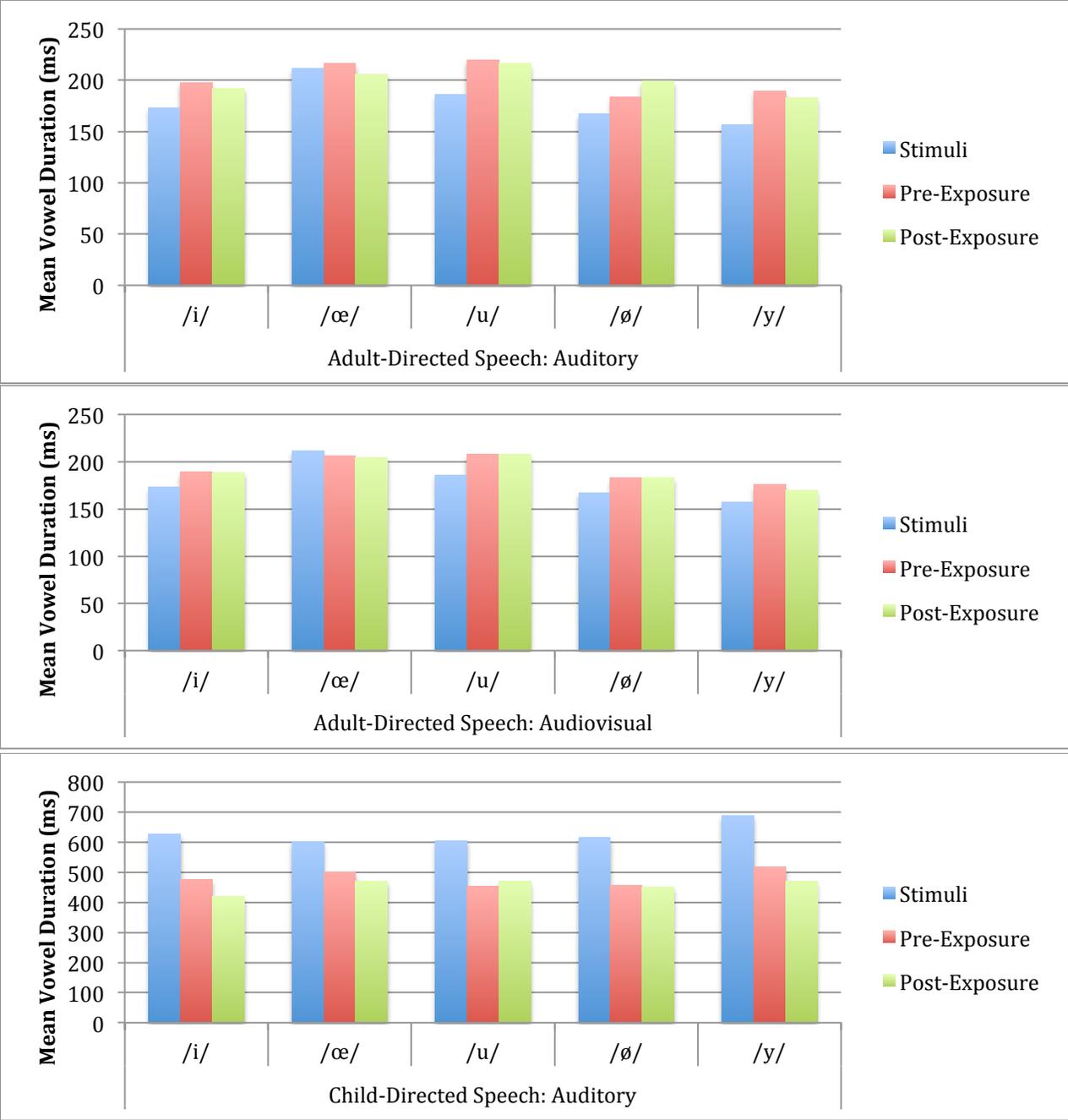


Figure 6.29. Mean duration of vowels produced by child subjects in the pre-exposure and post-exposure phases, compared with the mean duration of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.

The above graphs plot duration in each of the three conditions for child subjects. Starting first with the adult-directed speech:auditory condition, we see that subjects converged in duration to the model talker for all vowels except /ø/, shortening the duration of their productions across the two phases of the experiment. For the adult-directed speech:audiovisual condition, subjects converged in duration to the model talker uniformly, shortening their duration, and this resulted in convergence for all vowels except /œ/, in which shortening the duration resulted in divergence. For the child-directed speech:auditory condition, subjects shortened the duration of the vowels resulting in divergence, for all vowels except /u/, in which subjects converged by increasing the duration.

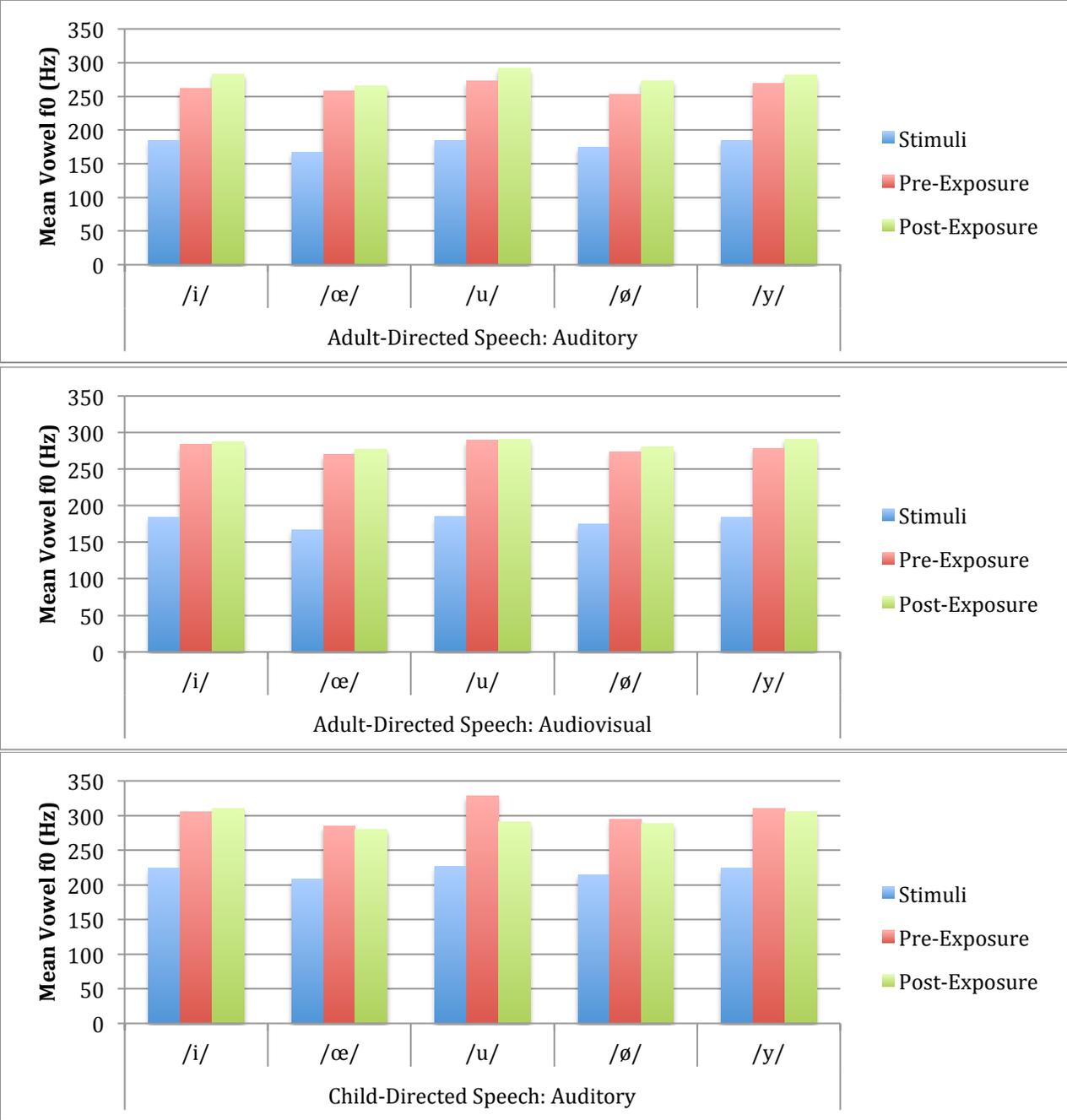


Figure 6.30. Mean f0 of vowels produced by male child subjects in the pre-exposure and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.

Male child subjects in both of the adult-directed speech conditions raised their f_0 after continued exposure to the model talker. In all cases, this resulted in divergence (but once again, may be explained by them normalizing the speech that they hear, a higher than average f_0). For the child-directed speech condition, we see convergence to the model talker (lowering of f_0) for all vowels except /i/. This is interesting because we would expect if child subjects were to raise their f_0 , it would be in response to the largely raised f_0 of child-directed speech.

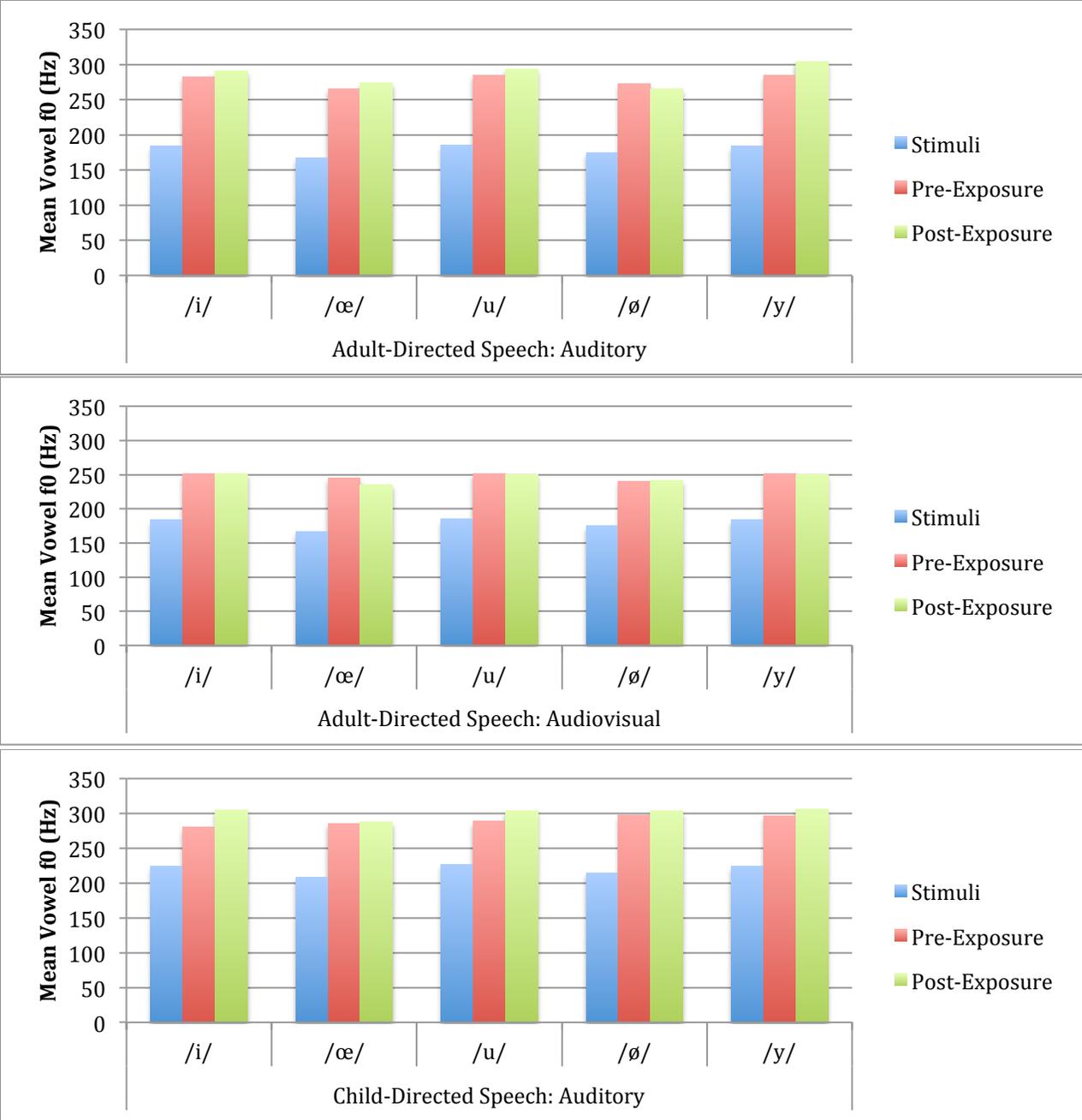
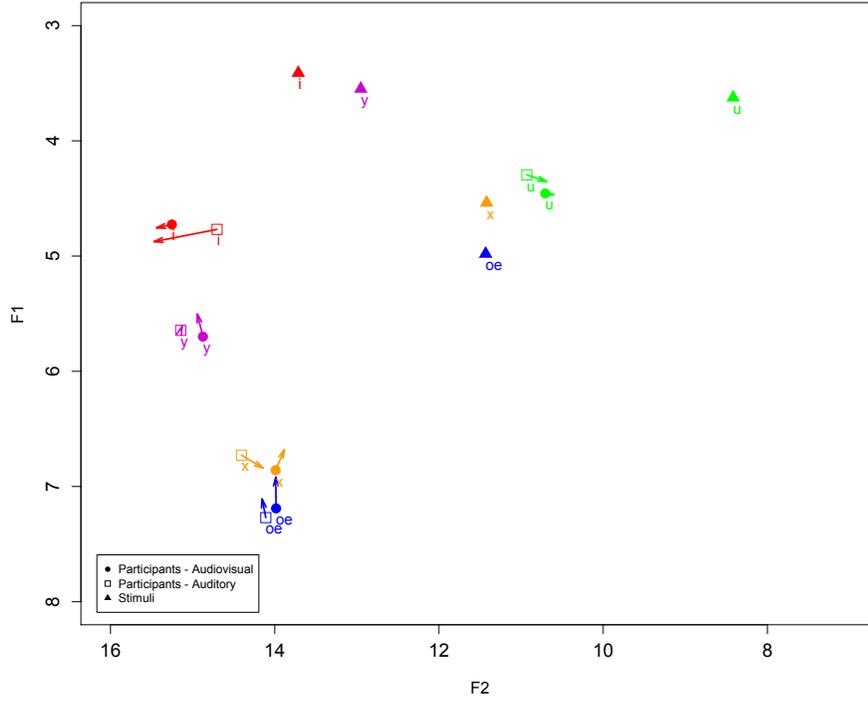


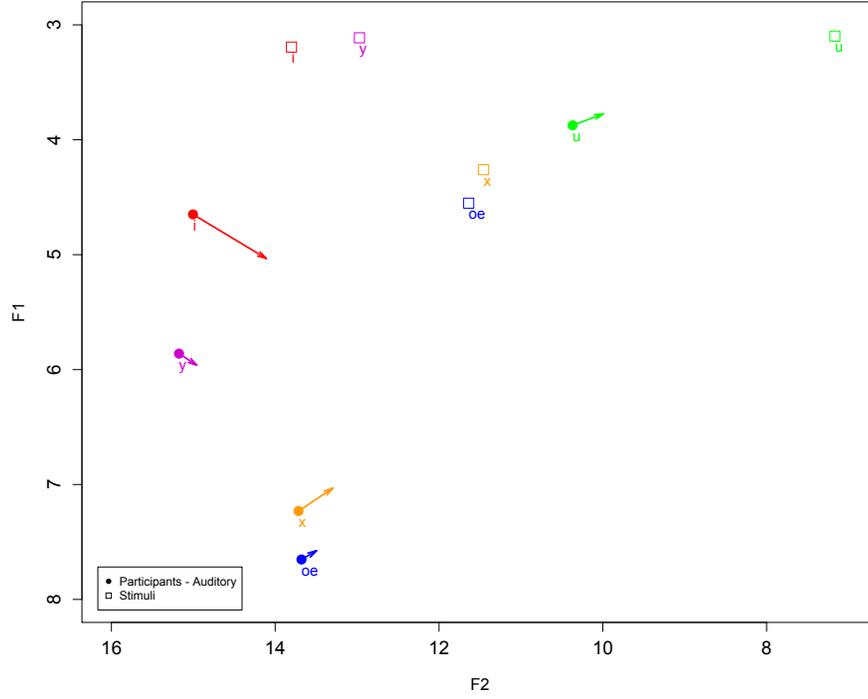
Figure 6.31. Mean f0 of vowels produced by female child subjects in the pre-exposure and post-exposure phases, compared with the mean f0 of the stimuli, separated by vowel. Separate plots are made for each modality and register combination.

Female child subjects showed varying productions depending on condition. In the adult-directed speech:auditory condition, female child subjects raised their f0 for all vowels except /ø/ (once again, this could be normalizing). For the adult-directed speech:audiovisual condition, they lower their f0 for all vowels, converging to the model talkers raw f0 values. For the child-directed speech:auditory condition, they raise their f0 after exposure to the model talker (another possible case of normalizing).

F1-F2 Imitation: Adult-Directed Speech



F1-F2 Imitation: Child-Directed Speech



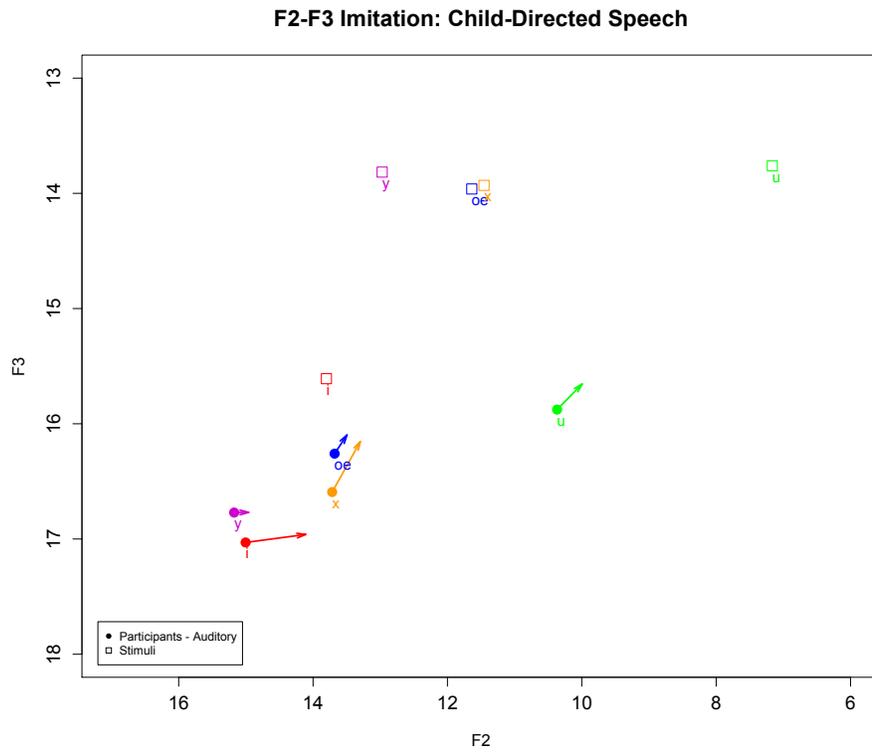
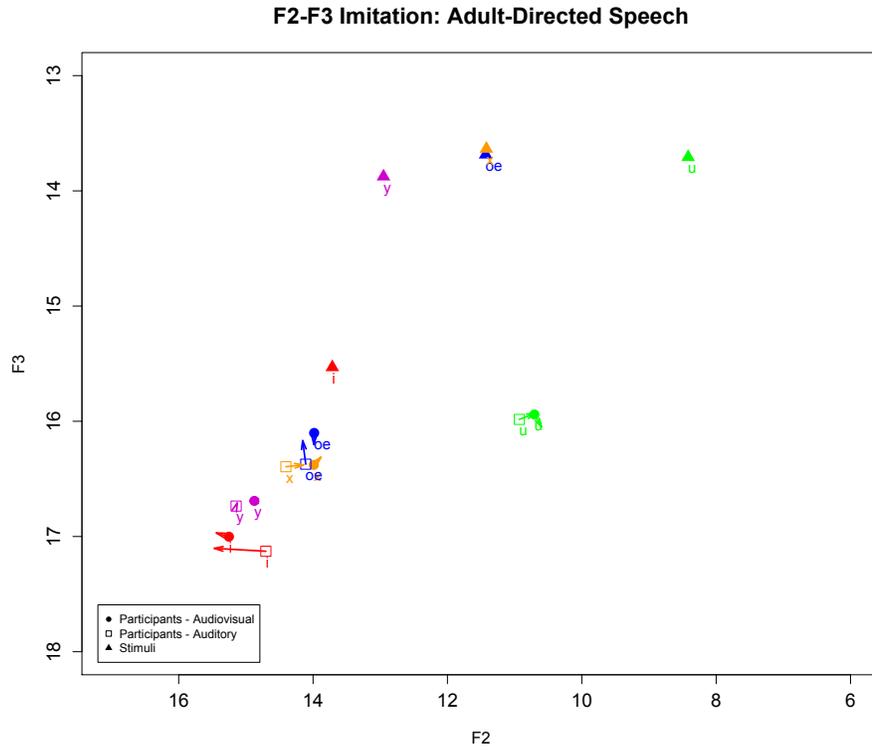


Figure 6.32. Formant plots of vowels produced by subjects in in the pre-exposure and post-exposure phases, compared with the mean formant values of the stimuli, separated by vowel, modality, and register.

Finally, looking at the direction of convergence on the phonetic measures for the vowels in the pre-exposure/post-exposure comparison, we will first analyze the adult-directed speech conditions. In the adult-directed speech conditions, for F1/F2, we saw that each of the vowels patterned slightly differently and, in many cases, the imitation of the vowels also differed according to modality. Both the English-like vowels /i/ and /u/ patterned similarly regardless of modality, although in different ways: converging by backing F2, but diverging by raising F1 for /u/ and diverging in F1 and F2 for /i/. For /æ/ and /y/, there was convergence in F1 but not F2 (although this was much less for /y/). For /ø/, there was a difference according to modality. In the auditory modality, participants converged on F2, backing their vowel, but diverged on F1, lowering their vowel production (raising F1 value). In the audiovisual modality, participants converged in both F1 and F2 for /ø/, raising the vowel height and backing the vowel, but the difference in F1 was greater. For F3 values of all vowels in adult-directed speech, the only case of a larger amount of F3 imitation came for /æ/ in the auditory domain, lowering F3 values in response to low F3 in the stimuli (we also saw some F3 imitation of other vowels, but it was very minimal). The vowel imitation in adult-directed speech did not show any consistent patterns.

Looking next at child-directed speech, for F1/F2 we saw the subjects converged in F1 and F2 for /u/, /ø/ and /æ/. For the other two vowels, /i/ and /y/, they converged only on F2, but diverged on F1. For F3 in child-directed speech, we once again saw that child subjects clearly imitated the stimuli for /u/, /ø/ and /æ/, raising their F3, but only slightly imitated F3 for /i/ and /y/. The high front vowels were not well imitated by the child subjects in child-directed speech compared to the other vowels.

Register

Overall, child subjects showed more convergence (or less divergence) in the child-directed register for all measures except duration and F1 (these results are displayed in Figure 6.33 below). Although the subsets of data separated by gender and vowel did not show such robust findings, pooling the male and female data, and the vowel data, together revealed that there are at least trends, not significant in all cases, for convergence in the child-directed register to be greater on all measures other than duration and F1.

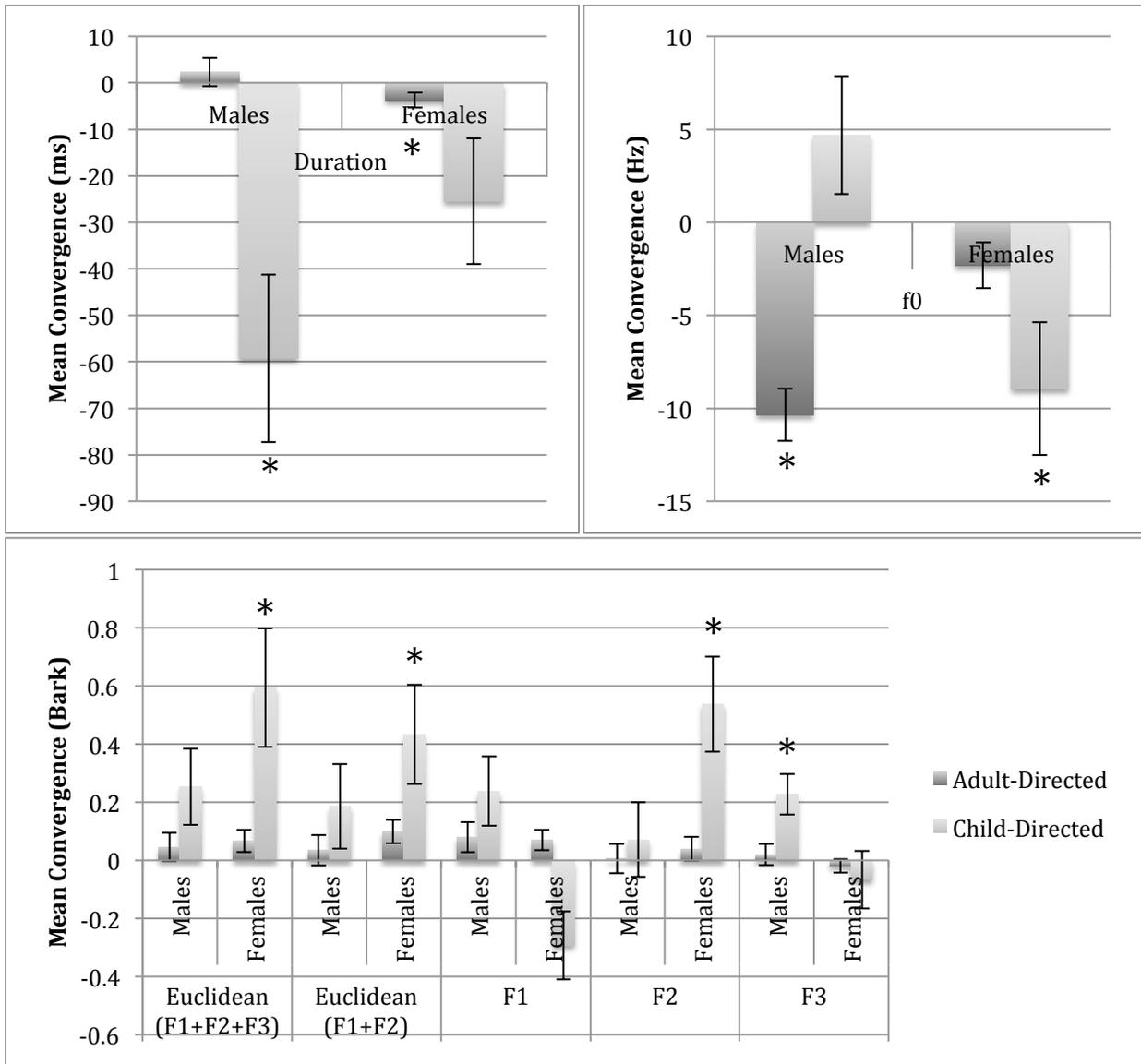


Figure 6.33. Results for convergence according to register for each measure in the pre-exposure/post-exposure comparison for child participants. The dark bars represent results for the adult-directed register and the light bars represent results for the child-directed register. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

Recall, in the male data, we saw convergence findings favoring adult-directed speech for duration and for English-like vowels in the overall Euclidean Distance (F1+F2+F3) global

phonetic measure. All other findings were favoring increased imitation in the child-directed register. However, for the female data, quite a different scenario was presented. For all measures except f_0 and F2, the female subjects converged more on adult-directed speech in foreign vowels, and on child-directed speech for English-like vowels.

Modality

The results separated by modality reveal that there is a tendency for divergence to be lesser, and convergence to be greater, in the audiovisual modality for the global measures of duration and f_0 . The individual phonetic measures only show a favorable finding for convergence in the audiovisual over auditory modality for F1, but this difference is robust enough to also create a trend for the global phonetic measure including F1 and F2 only to also show favorable imitation in the audiovisual over auditory modality. These modality findings are directly in contrast to the predictions for the effects of modality on particular formants, which would suggest that of all the formants, F1 would benefit the least from the addition of the visual modality, as openness is said to best be perceived auditorily (Traunmüller & Öhrström, 2007).

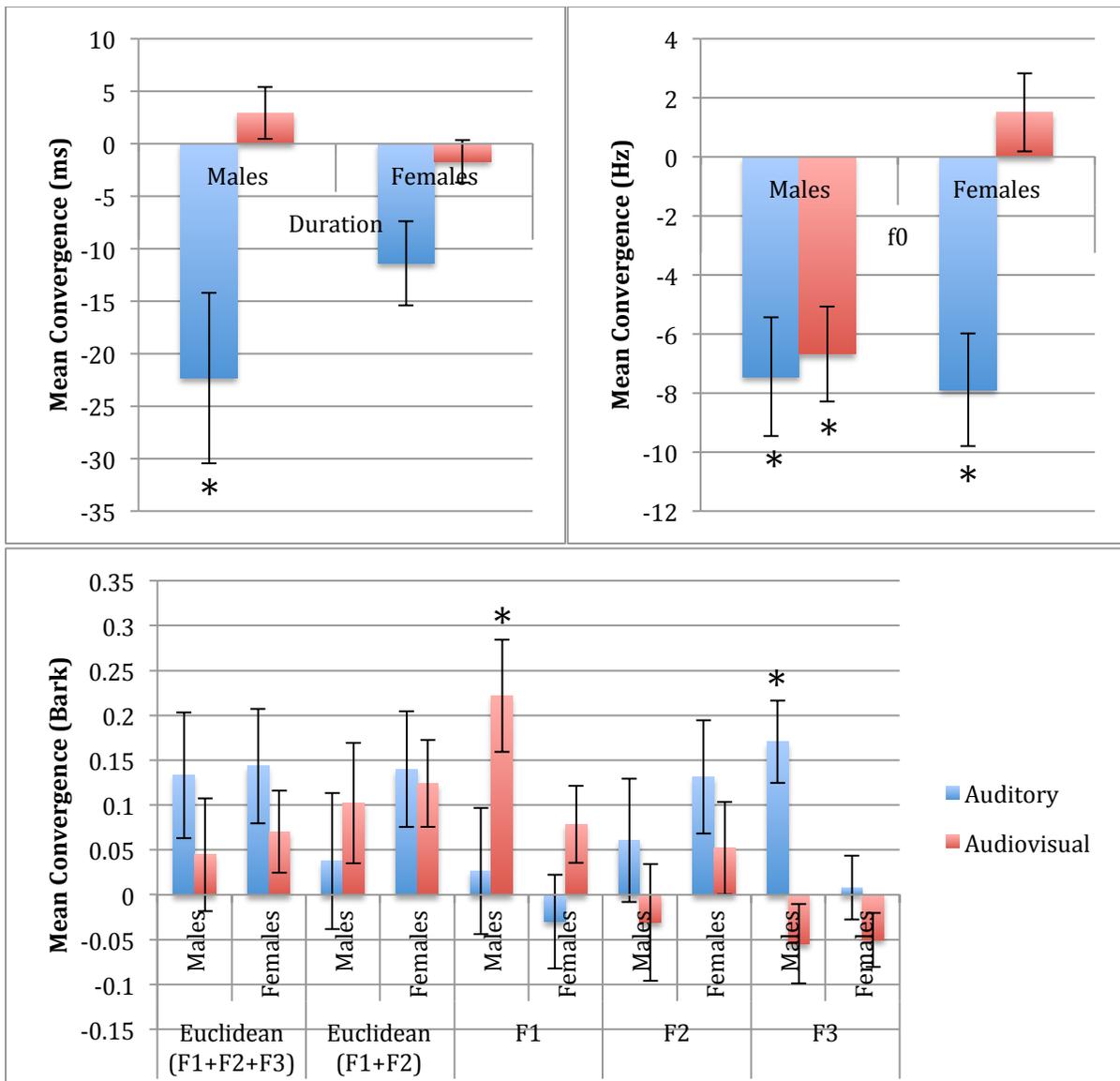


Figure 6.34. Results for convergence according to modality for each measure in the pre-exposure/post-exposure comparison for child participants. The blue bars represent results for the auditory modality and the red bars represent results for the audiovisual modality. Significance stars represent significant convergence or divergence (difference from zero) with a p -value of 0.05 or smaller.

In the results separated out by gender, while the male and female subjects both matched in their results by modality, there were still a number of differences. The female subjects showed

increased convergence to the audiovisual modality in both vowel types for both of the global measures, whereas the male subjects only showed this effect for English-like vowels. The male subjects showed significant findings of convergence in the auditory over audiovisual modality for both vowel types for F3, whereas female subjects showed no such differences. Finally, the other difference of note is that the male subjects showed no modality differences for convergence to the global phonetic measures for the English-like vowels, whereas females did.

Discussion and comparison between pre-test/post-exposure results and pre-exposure/post-exposure results

In this final section, we will look at the results from both of the analyses completed for this chapter. We will look at both the comparison of pre-test and post-exposure productions, and the comparison of pre-exposure and post-exposure productions in order to make generalizations about the imitative behavior of child subjects in this study. Specifically, we will evaluate whether the predictions and hypotheses for this study held for our experimental results. We will begin by reviewing these hypotheses, then we will look at whether they held in our results.

Review of predictions related only to child subjects

- Child subjects will imitate the speech that they hear. Previous research shows that child subjects converge in both timing measures as well as phonological measures (Street & Cappella, 1989, Eaton & Bernstein-Ratner, to appear, Ryalls & Pisoni, 1987, Nielsen, 2011b).

Prediction = Supported

- We had no specific predictions for register effects on child subjects, as compared to adult subjects, but we predicted an advantage of the child-directed register for imitation, as it provides longer durations of exposure to the vowels.

Prediction = Supported

- Child subjects will not show an increase in convergence with the audiovisual modality. Previous research shows that child subjects do not use the visual cues in the same way as adults (McGurk & MacDonald, 1976, Massaro, 1984, 1987, Desjardins et al. 1997), and at the age being tested, there is a decrease in reliance to these cues (Jerger et al. 2009).
 - Without prior research to corroborate our claim, we hypothesized that perhaps if the child subjects may use these cues in one condition of the experiment, it would be for foreign vowels.

Prediction = Not supported

Review of predictions related only to adult subjects

- Audiovisual speech was expected to facilitate convergence. Previous research using perceptual analysis measures showed increased rates for convergence when subjects were exposed to the audiovisual rather than auditory domain.
- Audiovisual speech was expected to facilitate convergence of visually salient cues.
 - In particular, a measure relating to vowel rounding – F3, which is visually salient, was expected to show increased convergence in the audiovisual domain, relative to cues for vowel height (F1) and backness (F2).
- Child-directed speech was expected to facilitate convergence, as it provides longer durations of vowel exposure, and it is a learning register.
- Imitation of the English-like vowel /u/ will be greater than imitation of the English-like vowel /i/ because of the variable pronunciations of /u/ in American English, attributed to sociolinguistic factors.
- Imitation can vary based on the type of measure being analyzed (global measures, global phonetic measures, individual phonetic measures). Previous research hypothesizes that global measures may be the most observable type of measure for studies on phonetic convergence (Mitterer, 2013).

Did child subjects show significant convergence?

Overall, the child subjects showed significant convergence findings in both the pre-test/post-exposure comparison and the pre-exposure/post-exposure comparison. Looking at just the overall findings, it appears as though for the pre-test/post-exposure comparison that the majority of findings were for the vowel /u/. Additionally, convergence was not evident in the f0 measure, rather, for this measure, there was significant divergence. In the pre-exposure/post-exposure comparison, looking once again at just overall findings of convergence, we once again see a number of significant convergence findings, although mostly just for the global phonetic measures. Furthermore, for this comparison, the global measures actually showed divergence across the board (the duration results are somewhat surprising, but it could just be explained by that subjects were attuning their attention to the phonetic detail evident in the formants, rather than the timing measures). It appears that in both the pre-exposure and pretest analyses, there is significant divergence for f0, demonstrating that child subjects did not want to imitate the pitch of their interlocutor. It is surprising that we observed significant divergence for duration in the pre-exposure/post-exposure comparison, because timing measures are well imitated in speech (Mitterer, 2013, Street & Cappella, 1989, Kim, 2011). In fact, the overall tendency of child subjects appeared to be imitation of the formant values, and the imitation effects were most visible on the global phonetic measures.

Did child-directed speech facilitate convergence?

Overall, in our study, we made the prediction that the child-directed speech register would facilitate imitation. A study in imitation in child subjects suggested increased levels of imitation for children rather than adults (no statistical comparison was made), and this difference was explained as likely due to the use of the child-directed register (Nielsen, 2011b). To review some of the information presented in Chapter 3, the child-directed speech register is slower and more exaggerated, allowing for both increased exposure and greater (perhaps clearer?) phonetic targets.

For the pre-test/post-exposure comparison, we saw a number of findings suggesting increased imitation of speech in the child-directed register. While we observed that for almost every measure (not for f_0) there was more overall convergence in the child-directed register not all findings were statistically significant (duration and F3 showed the strongest findings). However, there were no significant effects suggest register effects in the opposite direction.

For the pre-exposure/post-exposure comparison, like for the pretest comparison, the majority of the measures showed a mean value of convergence greater in the child-directed register than the adult-directed register, with the exception of duration and F1. However, looking at the male and female data separately, we found gender differences in imitation in the two registers. For the male data, there was increased convergence in the adult-directed register for two measures, one specific only to English-like vowels, but increased convergence in the child-directed register for all other measures. For the female data, however, we saw a trend (on most measures) for increased convergence in adult-directed speech for foreign vowels and child-directed speech for English-like vowels. This is perhaps the opposite pattern of what is expected;

the child-directed register by nature is a learning register, and so we would expect that it would be used more in learning to produce foreign vowels. Instead, female subjects were better at imitating foreign vowels in adult-directed speech. Future research would be needed to evaluate what is motivating this difference (perhaps sociolinguistic reasons?).

Overall, there is incomplete evidence for increased imitation of child subjects in the child-directed register, but there is also evidence that at least for the female child subjects, they may be systematically using register for imitation in different ways for different vowel types.

Did child subjects improve in imitation in the audiovisual modality?

Children are less susceptible to the visual cues of speech. They are suggested to not integrate auditory and visual speech cues in the same way as adults (McGurk & MacDonald, 1976) and they are also known to be poor lip readers (Massaro, 1987). They have been shown to have a U-shaped developmental function with regards to the visual cues to speech, declining in their use of these cues at around 3-4 years, and reincorporating them into their integrated processing systems at around 9-10 years (Jerger et al. 2009). We used subjects who were aged in the middle of this decline, when audiovisual speech integration is supposedly at the cusp of becoming the poorest. We hypothesized that subjects would be less likely to use the visual cues, but we allowed for that they *could* use them, as using the visual cues in an imitation task was not specifically addressed in the previous literature.

We found a difference in treatment of the two modalities between the pre-test/post-exposure comparison and the pre-exposure/post-exposure comparison. In the pretest analysis, we found that convergence appeared greater overall in the auditory modality, evident in the global

measures and in the global phonetic measures. However, for the pre-exposure analysis, we found an overall tendency of lesser divergence, and greater convergence in the audiovisual modality. This was trending in all of the measures, and significant for the four of the seven measures. It appears as though this difference is likely motivated by the vowels being imitated, with the English-like vowels in the pre-test comparison not showing as much change, and the foreign vowels being influenced more by the extra cues available in audiovisual speech.

Overall, there is some evidence that the audiovisual modality aids in the imitation of speech sounds by child subjects. This finding is surprising considering the literature on children's use (or lack thereof) of the visual cues in speech processing. While the findings are nowhere near robust (and for the pre-test comparison, there were a few effects suggesting the opposite effect), the fact that we find any evidence of an audiovisual advantage in children is intriguing.

Did audiovisual speech facilitate convergence more for foreign vowels than for English-like vowels?

We hypothesized that audiovisual speech cues could aid more in implicit imitation of foreign compared to English-like vowels. If child subjects were to use the visual cues, we thought that perhaps they would use them when they would provide the most potential advantage. The foreign vowels not only had visually salient cues corresponding with them, but because they were foreign, we thought that the subjects might attempt to use all possible information.

We found a number of differences between English-like and foreign vowels within the two separate analyses, but perhaps what is most significant is the trend for increased convergence to the auditory modality in the pre-test/post-exposure comparison (English-like vowels), and the trend for increased convergence to the audiovisual modality for the pre-exposure/post-exposure comparison (English-like and foreign vowels). This provides some evidence for this vowel difference in the use of the audiovisual modality.

Do modality effects selectively enhance uptake of particular cues that are salient in that modality?

Recall that we hypothesized that implicit imitation of specific phonetic measures could be affected by the modality of exposure. To recap, F1 is said to be most salient in the auditory modality, and F3 is most salient in the audiovisual modality (Traunmüller & Öhrström, 2007). We hypothesized that because of this difference, we would likely see the visual cues contribute more to imitation of F3 than to imitation of F1, and for F1 we would see little to no differences in imitation between the auditory and audiovisual modality.

The results for the individual formant measures in the pre-exposure/post-exposure analysis does show differences in imitation based on modality for particular formants, but these results are almost directly contradictory to what would be expected. First, we see an overall advantage of the audiovisual modality for F1. Next, in results separated by gender, we see that the male participants show a strong effect for the auditory over audiovisual modality for F3. These results are both directly contrastive to the hypotheses for these formants.

Did subjects imitate /i/ and /u/ differently?

Overall, in our study, we made the prediction that /u/ would show more implicit imitation than /i/ because /u/ shows more sociolinguistic variation, particularly in F2, and cues which show sociolinguistic variation tend to allow for greater degrees of imitation (Babel, 2010). In fact, we did observe this difference in the pre-test/post-exposure comparison. We not only saw a number more significant findings of convergence for /u/ than /i/ but for F2, we also found a significant result of greater imitation for /u/ than /i/. It appears that even children are sensitive to this social information being conveyed by F2 variation in American English, and their productions on this measure are especially pliable, just as in adult subjects.

Chapter 7

Discussion

The main goal of this dissertation was to determine the roles of the visual modality and child-directed register in imitation of vowel sounds. The previous two chapters presented results of studies on adults and children, looking at whether these two factors affected whether there was implicit imitation, as well as how they affected the degree of imitation. Before summarizing these results and discussing their implications for research on the visual cues, imitation, and language acquisition, I will first review the main research questions involved in these experiments, which can be broken into three separate questions.

7.1 Review of research questions

The main objectives of this study can be summarized in three main research questions, although there are other issues which we also attempted to evaluate, such as how subjects imitate English-like compared to foreign vowels, and how sociophonetic factors affect vowel imitation of /i/ and /u/. These three research questions all relate to how modality and register affect imitation, and they are described more fully below, along with how they related to our experimental design.

Research question #1: Is imitation of speech sounds increased with the addition of the visual cues? Does the visual modality increase convergence only for visually salient cues or globally across all acoustic characteristics of the sounds?

In this research question, we were interested in looking at the specific acoustic cues that show more implicit imitation with exposure to the visual modality. While it is well established that there is increased imitation of speech when visual cues are added (Miller et al. 2010, Dias & Rosenblum, 2011), there is no analysis of what measures show improved imitation. Are the cues that are imitated better those which are visually salient? Or is there a global increase overall on all measures? In order to answer this question, we compared auditory and audiovisual imitation using a number of different measures, of global features such as timing and pitch, and phonetic features such as vowel formants.

Research question #2: Will visual cues affect children's imitation of speech?

This research question is based on the literature suggesting that children do not attend as well to visual cues in speech perception tasks as adults (McGurk & MacDonald, 1976, Massaro, 1987). We were interested in whether children would continue to ignore these cues in an imitation task, in which they would need to produce speech based on what they perceive. In particular, we presented subjects with foreign vowel sounds with visually salient cues (vowel rounding), so that the use of the visual cues could help in the production of these sounds. We were interested in whether, in our task when the visual cues could aid in imitation, children might use these cues.

Research questions #3: Is imitation of speech sounds increased with the use of the child-directed register?

The final research question that we had was in relation to speech register. In particular, we were interested in whether child-directed speech would facilitate imitation, as it is not only a learning register with its maximized vowel space and increased segment durations (Kuhl et al. 1997). One previous study looked at imitation in child subjects, and this study found high levels of imitation (Nielsen, 2011b). Was this due to the child-directed speech, or was this a result of using child subjects? The nature of the task in this study was looking at not just imitation, but imitation of unfamiliar vowel sounds. It is logical that for these sounds in particular, the use of a learning register could facilitate improved production of the new speech sounds. Additionally, in our study, we were looking at imitation of adult and child subjects, and it has been suggested that children may be less likely to show imitation of an adult-directed register (Nielsen, 2011b).

7.2 General summary of results

Our first experiment looked at whether adults use the visual modality and the child-directed register to improve imitation of speech sounds. We found that exposure to the audiovisual modality indeed facilitated convergence, and this advantageous effect was evident in both global measures and phonetic measures. Audiovisual presentation enhanced overall imitation, not just for the visually salient cues. Additionally, modality effects were more evident in males than females.

The register effects were similar, although more robust. We found that the child-directed register was beneficial in imitation of phonetic measures and global measures, with the exception of f_0 in male participants. As for the role of whether the vowel was English-like or foreign, there was a tendency for the audiovisual modality and the child-directed register to aid more in imitation of foreign vowels than English-like vowels, but only for female participants. Overall, one final finding was that, when looking at all of the significant results, there was a trend that the female participants were more sensitive to register and the male participants were more sensitive to modality (females showed more significant advantages of child-directed speech, and males showed more significant advantages of the audiovisual modality).

Our second experiment extended the first study by testing child subjects aged 4-6-years. Somewhat surprisingly, we found a number of significant results favoring increased levels of imitation in the audiovisual modality for child participants. We saw significant results for both global measures and phonetic measures, although not consistently. There were a few results favoring increased imitation in the auditory domain, but for most of the results there was greater imitation in the audiovisual modality for child participants. It appears as though they are using the visual cues in some ways to aid in imitation in this task. This does not align with previous findings showing a lesser perceptual reliance on the visual cues in children in this age range.

As for register effects in the child subjects, child subjects were consistent in imitating only the phonetic measures (formant values) in the child-directed register. In fact, overall, children diverged on global measures in the child-directed register as well as audiovisual modality. Thus, imitation by children and adults is qualitatively different. Although adults converge towards non-criterial global information present in the child-directed register, children do not. Considering the features of child-directed speech, these are the cues that are used to

attract attention and evoke a positive affect. Children in our study ignored these cues in their imitation, and focused instead on the phonetic cues, which are comparably consistent across modalities.

Finally, when looking at the results based on familiarity with the vowel, it appeared as though for male adult participants, they were more inclined to show significant modality-related results for *English-like* vowels in the **global** measures, and significant modality-related results for *foreign* vowels in the **phonetic** measures. It seems that the male participants changed their phonetic pronunciation of an unfamiliar vowel after increased exposure, but for vowels that exist in their language, they simply changed their pronunciation of the global measures in an attempt to sound more like the talker, and did not want to change their phonetic pronunciation. For female adult participants, when separating out the results by familiarity, we saw a trend favoring increased imitation in the adult-directed register for foreign vowels, but favoring increased imitation in the child-directed register for English-like vowels. For the female participants, it seems that they were more likely to change their pronunciation of English-like vowels only when the pronunciation was exaggerated in child-directed speech (however, there is no coherent explanation for the tendency to imitate foreign vowels in the adult-directed register). Further testing would be needed to establish how robust these vowel familiarity trends are, as well as to figure out their exact role in imitation. Additionally, recall that we only partially filled out this subject group, and we need to further evaluate all of the child results after testing more subjects.

Overall, both the adult participants and the child participants modulated their imitation according to modality and register. The audiovisual modality and the child-directed register allowed for greater degrees of imitation for both child and adult participants. Overall, the contributions of register and modality appeared to be of a global manner, occurring in small

amounts on a number of different factors, rather than strongly appearing on any particular measures. The exceptions are the global cues in the child-directed register, which were not imitated by child subjects.

7.3 Implications for integration of visual cues in speech

The results of our study lead us to four main findings related to the use of the visual cues in imitation. First, as in previous studies (Dias & Rosenblum, 2011), we see greater imitation in the audiovisual compared to the auditory modality. Second, the results of our study lead us to believe that the facilitatory effect of visual cues is not specific to any one acoustic measure. These findings are contradictory to Traunmüller & Öhrström's (2007) information reliability hypothesis that claims that the uptake of information is maximal in the modality where the cue is most reliable. For their study, they defined this operationally as the visual modality for vowel rounding, and the auditory modality for vowel openness. In an effort to make our study on imitation as analogous as possible to their study, we chose to also vary our vowel stimuli with the features of rounding and openness. We did not find that vowel openness was better imitated in the auditory modality, nor did we find that that vowel rounding was better imitated in the visual modality.

Third, we saw that the visual modality also facilitated imitation by child subjects, although not as robustly. Recall, the children in this experiment were 4-6-years-old, at the age in which the decline in the use of the visual cues is beginning to increase (Jerger et al. 2009). At least in this task, children *are* using the visual information, and integrating that information into their linguistic system to aid them in imitating the speech around them. While this does show

that children use the visual cues in some speech related tasks, it is an open question why they are able to use the visual cues in the current study. Is it that this is a production task? Or is it that they are presented with the more difficult task of learning foreign vowels, and so they are using all the available cues?

Lastly, we see some evidence that there is more influence of the visual modality in male participants compared to female participants. Male participants more robustly showed an effect of modality in their imitation than females. This is surprising, considering that in the literature, if there was a modality-related perceptual difference related to gender, it was in favor of females showing increased use of the visual cues (Irwin et al. 2006, Johnson et al. 1988). Perhaps this is due to that it is an imitation task, and the female participants, who are more likely to imitate (Namy et al. 2002, Pardo, 2006, Pardo et al., 2010, Babel, 2009) are more focused on general imitation rather than the modality information?

7.4 Implications for research on language acquisition

One objective of this study was to determine whether visual cues and the child-directed register would aid in learning *new* sounds (which we tested by asking subjects to imitate *unfamiliar* sounds, whether foreign in pronunciation, i.e. familiar vowels with foreign pronunciations, or if the sounds themselves were not native to the subjects' language, i.e. foreign vowels). We hypothesized that the learning register, and the additional cues provided by the audiovisual modality, would be particularly helpful in learning to produce the unfamiliar vowel sounds. In Section 6.3 above, we discussed how the visual modality helped with imitation and we can abstract from that, since we were using foreign vowels, that the visual cues could aid in

learning a new sound. Overall, we found that the visual modality and the child-directed register aided imitation across the board, and facilitated production of these new sounds.

Considering next whether the “learning” register of child-directed speech would aid in subjects’ imitation of the foreign sounds, we found that imitation by both child and adult subjects in our experiment was improved when the speaker was using the child-directed speech register compared to the adult-directed register (and these findings were robust). One of the novel findings here was that females showed more significant increased imitation in the child-directed register compared to males. Further, there was some evidence that for females, this advantage was more evident for the foreign vowel sounds not in the phonetic repertoire of English. Also with regards to improved learning in the child-directed register, for child subjects, there was some support for suggesting that children were better at ignoring the “extraneous” information in the child-directed register (the global cues) and focused their attention instead on imitating the phonetic aspects of the speech signal (the vowel formants). Following from this outcome, the nature of convergence is shown to differ between adults and children; children appear to use register information and imitation in order to learn the phonetic properties of the speech around them, whereas for adults, they imitate any features in the speech they are exposed to. Imitation in children takes a different form than imitation in adults; perhaps convergence and imitation are used by children more for learning, whereas adults use imitation and convergence to continue to establish social identity.

7.5 Implications for research on imitation (not related to register or modality)

The previous research on imitation provided several additional predictions in our study. One of these predictions related to imitation of our two English-like vowels /u/ and /i/. In previous reports, differences in imitation based on the sociolinguistic status of the vowel have been reported, such that more variation in pronunciation allows for greater degrees of imitation (Trudgill, 1981, Babel, 2010). In our study, we saw sizeable differences in the imitation of /i/ and /u/ for F2, favoring increased imitation of /u/.²² This is not surprising, considering /u/ shows more sociolinguistic variation due to /u/ fronting in various American dialects (Clopper et al. 2005), and shows more acoustic variation between dialects of French and English (Strange et al. 2007), allowing for more imitation. Our results confirm that the vowel that is known to show more variation within English does show increased levels of convergence. This is particularly interesting when you consider that the target pronunciation was that of French vowels, and thus non-native and not part of the ongoing vowel shifts in American English.

An additional interesting, but not central, finding in the present study was we saw some evidence that AQ score affected the imitation results within our experiment (usually in interactions with other factors). There were a number of interactions of register, vowel, and modality with AQ score. For many of these factors, we can consider one of the two options the more “socially-driven” option. For example, consider the /i/ and /u/ differences discussed in the previous paragraph. Imitation of /u/ is conditioned by social variation, whereas imitation of /i/ is not. We found that there was a decline in /u/ imitation in adult participants with a higher AQ

²² This could also be attributed to quantal effects (the relationship between articulatory parameters and acoustic output not being linear), allowing for more variation for /u/ than /i/. We acknowledge this as a possible additional explanation.

score. The socially driven imitation decreased as AQ score got higher.²³ There were other similar findings of AQ score interacting with register and modality as well, and almost all of these can also be explained by social factors. While we cannot make a specific cause-effect connection, we can acknowledge this tendency, and suggest that future research include measures of AQ to better account for individual differences in the imitation behavior of adults.

7.6 Future directions for the current research

While this dissertation does address the main research questions of whether modality and register affect imitation, there are also a number of other studies that can be conducted based on the data that we have collected. An important follow-up to the current study will be an AXB perceptual analysis of the production data that we have collected. The existing research on the role of audiovisual cues in imitation has used perceptual measures, and an analysis of our data using perceptual measures will help us assess whether, using foreign vowels, or English-like vowels with foreign pronunciations, we are seeing the same pattern in the results. We found that the audiovisual modality increased imitation in a relatively global manner (the modality effects manifested across all of our speech measures), and a perceptual analysis will let us know whether the amounts of imitation were comparable to other studies.

Additional follow-up perceptual studies we think would be interesting are studies using the audiovisual recordings, or the visual component of the recordings, in an AXB task rather than just the auditory recordings. Perhaps the subjects are gesturally imitating the model talker when exposed to the AV cues, and this imitation is more evident in the visual modality (gestures)

²³ A poster presentation by Mielke, Nielsen, & Magloughlin (2012) at LabPhon 13, Stuttgart, suggests the same tendencies found in our data.

rather than in the auditory modality (acoustic characteristics). Along the same lines, we could also conduct an analysis of the video recordings for imitation of openness and rounding gestures, to establish whether subjects are imitating these movements, and learn how a visual analysis of imitation correlates with the acoustic findings from the current study. Since the visual cue of rounding is very evident visually, it is possible that subjects are imitating this gesturally, but not in the right manner to make it manifest acoustically as in the input.

With just the child data, there are a number more analyses we could complete with the data. First, we need to complete testing of the child subject groups (recall that we did not have data from child participants in the audiovisual exposure – child-directed register condition, and that we only had four participants for the auditory exposure – child-directed register condition). More participants might clarify the results and would give more power to the statistical tests. When testing the child participants we noticed a large tendency for child subjects to not want to attend to the screen during the exposure phase, even at the experimenter’s urging. Therefore, we think it would be worthwhile to see if child-subjects’ imitation in the audiovisual modality was correlated with time looking at the screen, and that could establish a more accurate connection between visual cue exposure and imitation. As we have video recordings of the subjects, we have all the necessary data to complete this analysis. Additionally of interest in the child-data is whether child subjects show differences in imitation by age. Recall, our child participants ranged in age from 4 to 6 years. Four years is towards the beginning of the known developmental decline in the use of the visual cues. In particular, does children’s use of the visual cues significantly decline for children towards the middle of the U-shaped developmental curve suggested by Jerger and colleagues (Jerger et al. 2009)?

Other analyses that we could complete on the current data are: statistical comparisons between child and adult participants, analysis of voice onset time in our stop consonants (correlated with register and modality), as well as an analysis of imitation correlations with our other experimental factors, such as the surrounding consonants. Recall that we only looked at influences of onset place of articulation and generalization in adult-participants, but we also have the ability to look at whether these factors affect the child results. Also, in addition to place of articulation, there were also voicing differences in the onset (and coda) consonants. We could also analyze whether particular foreign vowels showed more imitation (as we did find differences in the English-like vowels).

Another question of interest is whether child-directed speech facilitates imitation in adults because of its maximized vowel space, or due to the positive affect-response it elicits. Recall that child-directed speech differs from foreigner-directed speech in that child-directed speech shows high pitch, and lots of pitch variations, and these pitch qualities elicit positive feelings in their listeners. Both child-directed speech and foreigner-directed speech show maximized vowel spaces and longer durations of speech targets. If the increased imitation to child-directed speech is due to the positive affect, then foreigner-directed speech is not expected to facilitate imitation. However, if increased imitation in child-directed speech is due to exaggerated phonetic properties, then foreigner-directed speech should also show increased imitation compared to adult-directed speech.

These additional follow up studies are all justified by our findings that subjects exposed to speech in the audiovisual modality and the child-directed register increase imitation to speech sounds. In conclusion, although the facilitative effect of visual cues to imitation had been documented previously using perceptual measures, no acoustic analysis of audiovisual imitation

had been done. The current study contributes to that gap in the literature, and shows that imitation is increased globally, across different non-criterial global as well as phonetic measures, when speech presented in the audiovisual modality. We also saw some evidence of increased imitation in the audiovisual modality for children, which is especially important considering the literature that children do *not* use visual cues to speech at this age. Whether this finding is due to that this is a production task or that we presented children with foreign speech sounds should be evaluated in future studies. As for the finding of increased imitation in the child-directed register, there have been a number of studies that show an advantage of child-directed speech in acquisition, no studies have, to our knowledge, compared imitation in the two registers. And we saw an *advantage* to child-directed speech, not only for our child subjects but also for the adults. Also of particular interest is that child and adult subjects treat the phonetic cues and the global cues differently in imitation of the child-directed register; child subjects do not imitate the global factors, which obviously characterize child-directed speech. They ignore a lot of the information that they see as meant to make the speech more interesting, and hone in on the phonetic measures relevant to the speech contrasts. Adults, however, imitate all features of child-directed speech, global and phonetic features.

Appendix

The following charts represent the statistical output for the analysis of whether convergence is significantly different from zero throughout the study.

Adults:

Pre-test/post-exposure

			Estimate	Std. Error	t-value	p-value
Duration	/i/	All	71.86	21.18	3.39	0.001
		Female	70.72	26.03	2.72	0.008
		Male	73.96	37.13	1.99	0.052
		Adult	-77.02	9.53	-8.08	0.000
		Aud	48.32	30.65	1.58	0.120
		AV	94.75	29.20	3.25	0.002
		Child	233.88	18.57	12.59	0.000
		Female Child	217.02	23.20	9.35	0.000
		Female Adult	-75.58	13.80	-5.48	0.000
		Female Aud	57.45	40.44	1.42	0.163
		Female AV	83.99	33.49	2.51	0.016
		Male Child	269.12	23.91	11.26	0.000
		Male Adult	-79.38	9.38	-8.46	0.000
		Male Aud	30.82	46.43	0.66	0.514
	Male AV	113.79	56.59	2.01	0.056	
	/u/	All	111.35	21.20	5.25	0.000
		Female	109.91	26.17	4.20	0.000
		Male	114.00	36.88	3.09	0.003
		Adult	-40.95	6.68	-6.13	0.000
		Aud	87.29	30.02	2.91	0.000
		AV	134.75	29.84	4.52	0.000
		Child	277.09	17.51	15.82	0.000
		Female Child	264.41	21.33	12.40	0.000
		Female Adult	-44.59	9.08	-4.91	0.000
		Female Aud	96.48	40.22	2.40	0.021
		Female AV	123.34	34.17	3.61	0.001
		Male Child	303.60	29.08	10.44	0.000
		Male Adult	34.97	9.64	-3.63	0.001
Male Aud		69.66	43.34	1.61	0.122	
Male AV	154.93	57.88	2.68	0.013		
f0	/i/	All	20.84	2.72	7.67	0.000

		Female	18.40	3.01	6.11	0.000	
		Male	25.32	5.34	4.74	0.000	
		Adult	19.03	3.69	5.16	0.000	
		Aud	16.11	4.01	4.01	0.000	
		AV	25.43	3.56	7.15	0.000	
		Child	22.80	4.04	5.64	0.000	
		Female Child	18.49	3.43	5.40	0.000	
		Female Adult	18.31	5.04	3.64	0.001	
		Female Aud	14.17	4.23	3.35	0.002	
		Female AV	22.63	4.20	5.39	0.000	
		Male Child	31.81	10.03	3.17	0.005	
		Male Adult	20.22	5.36	3.77	0.001	
		Male Aud	19.83	8.65	2.29	0.032	
		Male AV	30.39	6.47	4.70	0.000	
	/u/	All	19.40	2.98	6.51	0.000	
		Female	15.76	3.22	4.90	0.000	
		Male	26.10	5.91	4.41	0.000	
		Adult	17.36	3.99	4.36	0.000	
		Aud	13.43	4.16	3.23	0.002	
		AV	25.20	4.09	6.17	0.000	
		Child	21.61	4.49	4.81	0.000	
		Female Child	16.02	3.56	4.51	0.000	
		Female Adult	15.49	5.44	2.85	0.007	
		Female Aud	8.84	3.76	2.35	0.023	
		Female AV	22.67	4.88	4.65	0.000	
		Male Child	33.31	11.29	2.95	0.008	
		Male Adult	20.44	5.72	3.57	0.001	
		Male Aud	22.22	9.55	2.33	0.030	
		Male AV	29.68	7.43	4.00	0.001	
Euclidean Distance (F1+F2+F3)	/i/	All	-0.18	0.04	-4.80	0.000	
		Female	-0.18	0.05	-3.70	0.000	
		Male	-0.18	0.06	-3.06	0.004	
		Adult	-0.18	0.05	-3.68	0.000	
		Aud	-0.18	0.05	-3.41	0.001	
		AV	-0.18	0.05	-3.34	0.001	
		Child	-0.19	0.06	-3.12	0.003	
		Female Child	-0.21	0.08	-2.73	0.009	
		Female Adult	-0.16	0.06	-2.46	0.018	
		Female Aud	-0.20	0.07	-2.93	0.005	
		Female AV	-0.16	0.07	-2.28	0.028	
		Male Child	-0.14	0.10	-1.45	0.162	
		Male Adult	-0.21	0.07	-2.85	0.008	
		Male Aud	-0.14	0.08	-1.70	0.104	
		Male AV	-0.21	0.08	-2.55	0.018	
		/u/	All	-1.73	0.13	-12.97	0.000
			Female	-1.81	0.16	-11.34	0.000
			Male	-1.60	0.24	-6.56	0.000
			Adult	-1.33	0.19	-7.10	0.000
			Aud	-1.60	0.20	-7.99	0.000
			AV	-1.86	0.18	-10.47	0.000
			Child	-2.17	0.16	-13.39	0.000

		Female Child	-2.16	0.22	-9.70	0.000
		Female Adult	-1.47	0.21	-6.98	0.000
		Female Aud	-1.74	0.24	-7.19	0.000
		Female AV	-1.88	0.21	-8.85	0.000
		Male Child	-2.21	0.20	-10.91	0.000
		Male Adult	-1.12	0.35	-3.22	0.003
		Male Aud	-1.34	0.36	-3.74	0.001
		Male AV	1.83	0.33	-5.54	0.000
Euclidean Distance (F1+F2)	/i/	All	-0.17	0.04	-4.78	0.000
		Female	-0.18	0.05	-3.94	0.000
		Male	-0.15	0.06	-2.66	0.011
		Adult	-0.16	0.04	-3.62	0.001
		Aud	-0.18	0.05	-3.40	0.001
		AV	-0.16	0.05	-3.31	0.002
		Child	-0.18	0.06	-3.16	0.002
		Female Child	-0.20	0.07	-2.78	0.008
		Female Adult	-0.16	0.06	-2.77	0.008
		Female Aud	-0.20	0.07	-3.03	0.004
		Female AV	-0.16	0.06	-2.48	0.017
		Male Child	-0.14	0.10	-1.46	0.159
		Male Adult	-0.16	0.07	-2.26	0.032
		Male Aud	-0.13	0.09	-1.54	0.139
	Male AV	-0.17	0.08	-2.16	0.041	
	/u/	All	-1.64	0.14	-11.74	0.000
		Female	-1.72	0.17	-10.09	0.000
		Male	-1.51	0.25	-6.10	0.000
		Adult	-1.17	0.20	-5.95	0.000
		Aud	-1.46	0.20	-7.17	0.000
		AV	-1.82	0.19	-9.57	0.000
		Child	-2.16	0.16	-13.55	0.000
		Female Child	-2.18	0.21	-10.24	0.000
		Female Adult	-1.25	0.23	-5.42	0.000
		Female Aud	-1.58	0.24	-6.47	0.000
		Female AV	-1.85	0.24	-7.77	0.000
		Male Child	-2.12	0.23	-9.42	0.000
		Male Adult	-1.02	0.36	-2.87	0.008
Male Aud		-1.22	0.37	-3.31	0.003	
Male AV	-1.77	0.33	-5.39	0.000		
F1	/i/	All	-0.05	0.04	-1.12	0.264
		Female	-0.13	0.05	-2.47	0.016
		Male	0.10	0.05	1.94	0.058
		Adult	-0.01	0.06	-0.17	0.865
		Aud	-0.08	0.06	-1.32	0.193
		AV	-0.01	0.05	-0.17	0.869
		Child	-0.08	0.05	-1.61	0.112
		Female Child	-0.14	0.07	-2.07	0.045
		Female Adult	-0.12	0.08	-1.45	0.153
		Female Aud	0.19	0.08	-2.42	0.020
		Female AV	-0.07	0.07	-1.00	0.322
		Male Child	0.03	0.07	0.43	0.673
		Male Adult	0.16	0.08	2.11	0.045

		Male Aud	0.11	0.09	1.25	0.224	
		Male AV	0.10	0.07	1.49	0.150	
	/u/	All	-0.08	0.05	-1.67	0.098	
		Female	-0.17	0.06	-2.80	0.006	
		Male	0.09	0.06	1.42	0.161	
		Adult	-0.02	0.06	-0.25	0.804	
		Aud	-0.11	0.07	-1.58	0.118	
		AV	-0.05	0.07	-0.76	0.450	
		Child	-0.15	0.07	-2.12	0.038	
		Female Child	-0.17	0.10	-1.72	0.092	
		Female Adult	-0.17	0.08	-2.31	0.026	
		Female Aud	-0.22	0.09	-2.47	0.017	
		Female AV	-0.13	0.09	-1.46	0.152	
		Male Child	-0.10	0.07	-1.50	0.149	
		Male Adult	0.24	0.08	3.07	0.005	
		Male Aud	0.10	0.09	1.14	0.267	
		Male AV	0.08	0.10	0.87	0.394	
F2		/i/	All	-0.15	0.04	-4.08	0.000
	Female		-0.17	0.05	-3.65	0.000	
	Male		-0.13	0.07	-1.95	0.058	
	Adult		-0.14	0.05	-2.74	0.008	
	Aud		-0.12	0.06	-2.19	0.032	
	AV		-0.18	0.05	-3.66	0.001	
	Child		-0.17	0.06	-3.00	0.004	
	Female Child		-0.19	0.07	-2.75	0.009	
	Female Adult		-0.14	0.06	-2.36	0.023	
	Female Aud		-0.15	0.07	-2.37	0.022	
	Female AV		-0.18	0.06	-2.75	0.009	
	Male Child		-0.14	0.11	-1.28	0.215	
	Male Adult		-0.13	0.09	-1.41	0.170	
	Male Aud		-0.07	0.11	-0.61	0.551	
	Male AV		-0.19	0.08	-2.36	0.027	
			/u/	All	-2.09	0.15	-13.56
		Female		-2.20	0.19	-11.38	0.000
		Male		-1.88	0.25	-7.40	0.000
		Adult		-1.74	0.23	-7.46	0.000
		Aud		-1.91	0.23	-8.41	0.000
		AV		-2.27	0.21	-10.89	0.000
		Child		-2.47	0.18	-13.74	0.000
		Female Child		-2.49	0.24	-10.30	0.000
		Female Adult		-1.92	0.30	-6.48	0.000
		Female Aud		-2.12	0.29	-7.25	0.000
		Female AV		-2.29	0.26	-8.82	0.000
		Male Child		-2.44	0.25	-9.83	0.000
		Male Adult		-1.44	0.37	-3.88	0.001
		Male Aud		-1.50	0.34	-4.46	0.000
		Male AV	-2.23	0.36	-6.15	0.000	
F3	/i/	All	-0.03	0.03	-1.12	0.265	
		Female	-0.03	0.04	-0.91	0.367	
		Male	-0.03	0.04	-0.66	0.515	
		Adult	-0.01	0.04	-0.36	0.717	

	Aud	0.00	0.04	-0.12	0.906
	AV	-0.06	0.04	-1.49	0.142
	Child	-0.05	0.04	-1.26	0.212
	Female Child	-0.05	0.05	-1.00	0.324
	Female Adult	-0.01	0.05	-0.28	0.785
	Female Aud	-0.03	0.06	-0.54	0.594
	Female AV	-0.04	0.05	-0.76	0.451
	Male Child	-0.04	0.05	-0.82	0.423
	Male Adult	-0.01	0.06	-0.24	0.816
	Male Aud	0.05	0.04	1.35	0.192
	Male AV	-0.09	0.07	-1.43	0.167
/u/	All	0.02	0.05	0.39	0.699
	Female	-0.04	0.04	-0.95	0.345
	Male	0.13	0.12	1.07	0.290
	Adult	0.04	0.08	0.53	0.599
	Aud	-0.01	0.07	-0.12	0.905
	AV	0.05	0.07	0.63	0.528
	Child	0.00	0.06	-0.07	0.947
	Female Child	-0.01	0.07	-0.21	0.834
	Female Adult	-0.07	0.05	-1.25	0.217
	Female Aud	-0.06	0.07	-0.91	0.368
	Female AV	-0.02	0.05	-0.37	0.714
	Male Child	0.02	0.14	0.11	0.912
	Male Adult	0.22	0.19	1.18	0.249
	Male Aud	0.09	0.16	0.56	0.581
	Male AV	0.17	0.18	0.90	0.379

Pre-exposure/post-exposure

		Estimate	Std. Error	t-value	p-value	
Duration	English-like	All	-12.66	2.33	-5.44	0.000
		Female	-13.45	2.94	-4.58	0.000
		Male	-11.13	3.81	-2.93	0.004
		Adult	-1.72	1.12	-1.54	0.124
		Aud	-4.83	3.13	-1.54	0.123
		AV	-20.71	3.45	-6.00	0.000
		Child	-24.47	3.99	-6.13	0.000
		Female Child	-23.70	4.88	-4.86	0.000
		Female Adult	-3.13	1.49	-2.10	0.036
		Female Aud	-4.83	3.94	-1.23	0.220
		Female AV	-22.86	4.41	-5.19	0.000
		Male Child	-26.16	7.00	-3.74	0.000
		Male Adult	0.68	1.62	0.42	0.675
		Male Aud	-4.82	5.17	-0.93	0.351
		Male AV	-16.90	5.58	-3.03	0.003

	Foreign	All	-14.57	1.95	-7.47	0.000
		Female	-13.26	2.48	-5.36	0.000
		Male	-17.08	3.16	-5.41	0.000
		Adult	-0.90	0.93	-0.96	0.335
		Aud	-7.43	2.70	-2.76	0.006
		AV	-21.88	2.83	-7.74	0.000
		Child	-29.32	3.35	-8.75	0.000
		Female Child	-26.07	4.15	-6.28	0.000
		Female Adult	-0.41	1.25	-0.33	0.742
		Female Aud	-7.09	3.47	-2.04	0.042
		Female AV	-19.98	3.54	-5.65	0.000
		Male Child	-36.45	5.66	-6.44	0.000
		Male Adult	-1.74	1.33	-1.30	0.193
		Male Aud	-8.15	4.09	-1.99	0.047
		Male AV	-25.23	4.73	-5.34	0.000
f0	English-like	All	-0.84	0.50	-1.69	0.091
		Female	0.63	0.57	1.11	0.269
		Male	-3.68	0.90	-4.11	0.000
		Adult	-0.58	0.73	-0.79	0.429
		Aud	1.37	0.69	1.98	0.048
		AV	-3.12	0.72	-4.31	0.000
		Child	-1.13	0.65	-1.74	0.082
		Female Child	-0.60	0.69	-0.86	0.390
		Female Adult	1.86	0.88	2.12	0.035
		Female Aud	1.39	0.82	1.69	0.092
		Female AV	-0.21	0.78	-0.27	0.790
		Male Child	-2.31	1.22	-1.89	0.059
		Male Adult	-4.76	1.06	-4.48	0.000
		Male Aud	1.32	1.17	1.13	0.259
		Male AV	-8.26	1.32	-6.24	0.000
	Foreign	All	-0.53	0.41	-1.30	0.195
		Female	0.34	0.48	0.71	0.478
		Male	-2.22	0.68	-3.26	0.001
		Adult	0.47	0.63	0.75	0.451
		Aud	2.09	0.59	3.57	0.000
		AV	-3.22	0.58	-5.59	0.000
		Child	-1.62	0.52	-3.12	0.002
		Female Child	-1.91	0.59	-3.25	0.001
		Female Adult	2.60	0.74	3.52	0.000
		Female Aud	1.99	0.69	2.86	0.004
		Female AV	-1.45	0.66	-2.20	0.028
		Male Child	-1.00	0.84	-1.19	0.236
Male Adult	-3.19	0.94	-3.40	0.001		
Male Aud	2.31	0.96	2.41	0.016		
Male AV	-6.36	0.95	-6.68	0.000		
Euclidean Distance (F1+F2+F3)	English-like	All	-0.06	0.03	-2.14	0.032
		Female	-0.06	0.03	-1.79	0.073
		Male	-0.07	0.06	-1.22	0.224
		Adult	-0.01	0.04	-0.35	0.726
		Aud	-0.05	0.04	-1.06	0.291
		AV	-0.08	0.04	-2.03	0.042

		Child	-0.12	0.04	-2.94	0.003		
		Female Child	-0.11	0.04	-2.53	0.012		
		Female Adult	-0.01	0.05	-0.22	0.829		
		Female Aud	-0.08	0.05	-1.59	0.112		
		Female AV	-0.04	0.04	-0.90	0.370		
		Male Child	-0.13	0.08	-1.59	0.112		
		Male Adult	-0.02	0.08	-0.28	0.777		
		Male Aud	0.02	0.09	0.20	0.845		
		Male AV	-0.15	0.08	-2.00	0.046		
		Foreign	All	-0.03	0.01	-1.91	0.057	
	Female		-0.04	0.02	-2.16	0.031		
	Male		-0.01	0.03	-0.31	0.755		
	Adult		0.00	0.02	0.06	0.950		
	Aud		0.01	0.02	0.40	0.693		
	AV		-0.07	0.02	-3.36			
	Child		-0.06	0.02	-2.87	0.004		
	Female Child		-0.07	0.02	-2.81	0.005		
	Female Adult		-0.01	0.02	-0.40	0.692		
	Female Aud		-0.01	0.03	-0.47	0.635		
	Female AV		-0.07	0.02	-2.94	0.003		
	Male Child		-0.04	0.04	-1.08	0.282		
	Male Adult		0.02	0.03	0.60	0.546		
	Male Aud		0.05	0.04	1.51	0.132		
	Male AV		-0.06	0.03	-1.84	0.066		
	Euclidean Distance (F1+F2)		English-like	All	-0.09	0.03	-2.82	0.005
				Female	-0.09	0.04	-2.61	0.009
		Male		-0.07	0.06	-1.29	0.199	
Adult		-0.03		0.04	-0.80	0.422		
Aud		-0.06		0.05	-1.34	0.182		
AV		-0.11		0.04	-2.76	0.006		
Child		-0.14		0.04	-3.42	0.001		
Female Child		-0.14		0.05	-3.02	0.003		
Female Adult		-0.04		0.05	-0.87	0.387		
Female Aud		-0.11		0.05	-2.21	0.027		
Female AV		-0.07		0.05	-1.43	0.155		
Male Child		-0.14		0.08	-1.71	0.088		
Male Adult		-0.02		0.07	-0.24	0.813		
Male Aud		0.05		0.09	0.61	0.540		
Male AV		-0.19	0.07	-2.57	0.010			
Foreign		All	-0.04	0.02	-2.81	0.005		
		Female	-0.05	0.02	-2.70	0.007		
		Male	-0.03	0.02	-1.03	0.303		
		Adult	0.00	0.02	-0.16	0.870		
		Aud	-0.02	0.02	-0.81	0.416		
		AV	-0.07	0.02	-3.39	0.001		
		Child	-0.09	0.02	-3.74	0.000		
		Female Child	-0.10	0.03	-3.80	0.000		
		Female Adult	0.00	0.02	-0.02	0.986		
		Female Aud	-0.03	0.03	-1.09	0.275		
		Female AV	-0.07	0.02	-3.01	0.003		
		Male Child	-0.05	0.04	-1.13	0.258		

		Male Adult	-0.01	0.03	-0.28	0.779
		Male Aud	0.01	0.04	0.28	0.782
		Male AV	-0.06	0.03	-1.72	0.086
F1	English-like	All	0.07	0.01	5.10	0.000
		Female	0.12	0.02	6.31	0.000
		Male	-0.01	0.02	-0.22	0.823
		Adult	0.06	0.02	3.57	0.000
		Aud	0.11	0.02	5.08	0.000
		AV	0.04	0.02	2.00	0.046
		Child	0.08	0.02	4.41	0.000
		Female Child	0.12	0.02	4.92	0.000
		Female Adult	0.12	0.02	4.92	0.000
		Female Aud	0.15	0.03	5.43	0.000
		Female AV	0.08	0.02	3.35	0.001
		Male Child	0.02	0.03	0.57	0.570
		Male Adult	-0.02	0.03	-0.91	0.362
		Male Aud	0.03	0.03	0.83	0.407
		Male AV	-0.04	0.03	-1.10	0.274
	Foreign	All	0.17	0.01	12.22	0.000
		Female	0.21	0.02	11.37	0.000
		Male	0.09	0.02	4.72	0.000
		Adult	0.15	0.02	8.27	0.000
		Aud	0.18	0.02	8.79	0.000
		AV	0.16	0.02	8.53	0.000
		Child	0.19	0.02	9.77	0.000
		Female Child	0.23	0.02	9.27	0.000
		Female Adult	0.19	0.03	7.35	0.000
		Female Aud	0.23	0.03	8.61	0.000
		Female AV	0.19	0.03	7.42	0.000
		Male Child	0.10	0.03	3.44	0.001
Male Adult	0.09	0.02	3.81	0.000		
Male Aud	0.07	0.03	2.49	0.013		
Male AV	0.11	0.03	4.20	0.000		
F2	English-like	All	0.10	0.04	2.51	0.012
		Female	0.06	0.05	1.29	0.199
		Male	0.17	0.07	2.48	0.013
		Adult	0.03	0.06	0.49	0.622
		Aud	0.08	0.06	1.27	0.206
		AV	0.12	0.05	2.39	0.017
		Child	0.18	0.05	3.51	0.001
		Female Child	0.14	0.06	2.39	0.017
		Female Adult	-0.01	0.07	-0.20	0.844
		Female Aud	0.06	0.07	0.83	0.405
		Female AV	0.06	0.06	1.01	0.311
		Male Child	0.26	0.10	2.65	0.008
		Male Adult	0.10	0.09	1.11	0.269
		Male Aud	0.11	0.11	1.02	0.309
		Male AV	0.23	0.09	2.58	0.010
	Foreign	All	0.01	0.02	0.28	0.782
		Female	0.03	0.03	1.04	0.301
		Male	-0.04	0.04	-1.08	0.281

		Adult	0.00	0.03	-0.09	0.932	
		Aud	0.03	0.03	0.73	0.468	
		AV	-0.01	0.03	-0.49	0.626	
		Child	0.02	0.04	0.43	0.666	
		Female Child	0.06	0.04	1.49	0.138	
		Female Adult	0.00	0.03	-0.14	0.893	
		Female Aud	0.10	0.04	2.20	0.028	
		Female AV	-0.04	0.03	-1.25	0.212	
		Male Child	-0.09	0.06	-1.41	0.158	
		Male Adult	0.00	0.03	0.06	0.954	
		Male Aud	-0.12	0.06	-2.23	0.026	
		Male AV	0.04	0.04	0.89	0.374	
F3	English-like	All	-0.01	0.02	-0.72	0.470	
		Female	-0.02	0.02	-1.13	0.258	
		Male	0.00	0.04	0.10	0.919	
		Adult	-0.03	0.03	-1.02	0.308	
		Aud	-0.01	0.02	-0.27	0.787	
		AV	-0.02	0.02	-0.76	0.447	
		Child	0.00	0.02	0.10	0.922	
		Female Child	-0.02	0.02	-0.87	0.383	
		Female Adult	-0.02	0.03	-0.83	0.404	
		Female Aud	0.00	0.03	-0.12	0.901	
		Female AV	-0.04	0.02	-1.60	0.110	
		Male Child	0.05	0.05	1.01	0.316	
		Male Adult	-0.03	0.05	-0.62	0.533	
		Male Aud	-0.01	0.05	-0.27	0.787	
		Male AV	0.02	0.05	0.38	0.701	
			Foreign	All	0.01	0.01	0.61
		Female		0.02	0.01	1.72	0.085
		Male		-0.02	0.02	-0.94	0.347
		Adult		0.00	0.02	-0.12	0.902
		Aud		-0.01	0.02	-0.86	0.389
		AV		0.03	0.02	1.77	0.078
		Child		0.02	0.02	1.06	0.288
		Female Child		0.03	0.02	1.43	0.154
		Female Adult		0.02	0.02	1.14	0.256
		Female Aud		0.01	0.02	0.81	0.416
		Female AV		0.03	0.02	1.63	0.103
		Male Child		0.00	0.03	-0.06	0.952
		Male Adult	-0.04	0.03	-1.17	0.241	
	Male Aud	-0.07	0.03	-2.23	0.026		
	Male AV	0.03	0.03	0.86	0.392		

Children:

Children: pre-test/post-exposure

			Estimate	Std. Error	t-value	p-value
Duration	/i/	All	-1.84	16.17	-0.11	0.910
		Female	0.08	19.04	0.00	0.997
		Male	-5.23	31.10	-0.17	0.869
		Child	148.15	26.72	5.54	0.002
		Adult	-30.40	9.90	-3.07	0.004
		Aud	30.10	25.73	1.17	0.254
		AV	-36.43	14.03	-2.60	0.016
		Female Child	152.02	51.22	2.97	0.097
		Female Adult	-21.63	12.63	-1.71	0.099
		Female Aud	35.32	28.82	1.23	0.241
		Female AV	-35.17	19.21	-1.83	0.089
		Male Child	144.29	40.39	3.57	0.070
		Male Adult	-47.95	14.54	-3.30	0.006
		Male Aud	21.75	53.00	0.41	0.692
	Male AV	-38.96	20.95	-1.86	0.112	
	/u/	All	25.63	21.87	1.17	0.247
		Female	9.60	23.02	0.42	0.680
		Male	54.11	45.39	1.19	0.251
		Child	277.69	43.90	6.33	0.000
		Adult	-12.43	19.20	-0.65	0.522
		Aud	12.38	23.72	0.52	0.606
		AV	44.36	9.64	4.60	0.000
		Female Child	159.53	46.37	3.44	0.075
		Female Adult	-11.82	19.82	-0.60	0.556
		Female Aud	45.40	29.26	1.55	0.143
		Female AV	-26.20	32.39	-0.81	0.432
Male Child		277.72	77.21	3.60	0.069	
Male Adult	-9.77	13.69	-0.71	0.489		
Male Aud	89.46	81.74	1.09	0.306		
Male AV	9.94	7.43	1.34	0.230		
f0	/i/	All	10.66	7.28	1.47	0.149
		Female	12.22	10.01	1.22	0.232
		Male	7.89	10.27	0.77	0.454
		Child	23.69	12.80	1.85	0.114
		Adult	8.18	8.31	0.98	0.331
		Aud	9.97	12.58	0.79	0.436
		AV	11.41	7.31	1.56	0.133
		Female Child	36.01	20.64	1.74	0.223
		Female Adult	8.82	10.97	0.80	0.429
		Female Aud	17.42	17.48	1.00	0.336
		Female AV	7.02	10.79	0.65	0.526

		Male Child	11.37	15.91	0.72	0.549	
		Male Adult	6.89	12.97	0.53	0.605	
		Male Aud	-1.96	17.93	-0.11	0.916	
		Male AV	20.19	2.65	7.61	0.000	
	/u/	All	14.72	6.02	2.45	0.018	
		Female	13.18	7.95	1.66	0.108	
		Male	17.46	9.47	1.84	0.084	
		Child	23.56	16.16	1.46	0.166	
		Adult	18.01	10.22	1.76	0.088	
		Aud	15.30	7.76	1.97	0.060	
		AV	-6.60	12.56	-0.53	0.605	
		Female Child	39.22	27.25	1.44	0.287	
		Female Adult	9.46	8.16	1.16	0.257	
		Female Aud	19.20	14.95	1.28	0.220	
		Female AV	7.15	6.06	1.18	0.257	
		Male Child	12.12	16.44	0.74	0.538	
		Male Adult	18.98	11.37	1.67	0.121	
		Male Aud	7.88	16.29	0.48	0.642	
		Male AV	29.43	3.09	9.51	0.000	
Euclidean Distance (F1+F2+F3)	/i/	All	-0.01	0.13	-0.08	0.938	
		Female	0.15	0.17	0.87	0.389	
		Male	-0.27	0.18	-1.49	0.157	
		Child	-0.70	0.13	-5.50	0.003	
		Adult	0.10	0.14	0.72	0.474	
		Aud	-0.15	0.19	-0.76	0.457	
		AV	0.14	0.17	0.78	0.443	
		Female Child	-0.35	0.50	-0.69	0.615	
		Female Adult	0.21	0.18	1.19	0.247	
		Female Aud	0.24	0.23	1.03	0.322	
		Female AV	0.08	0.25	0.30	0.768	
		Male Child	-0.80	0.12	-6.75	0.021	
		Male Adult	-0.12	0.20	-0.61	0.551	
		Male Aud	-0.69	0.08	-8.48	0.000	
		Male AV	0.25	0.16	1.63	0.155	
		/u/	All	-0.93	0.15	-6.01	0.000
			Female	-0.95	0.21	-4.53	0.000
			Male	-0.90	0.23	-3.89	0.001
			Child	-1.41	0.42	-3.37	0.004
			Adult	-0.91	0.18	-4.98	0.000
			Aud	-0.91	0.25	-3.68	0.001
			AV	-0.85	0.18	-4.81	0.000
			Female Child	-1.80	0.93	-1.93	0.194
			Female Adult	-0.83	0.20	-4.18	0.000
			Female Aud	-1.28	0.30	-4.27	0.001
			Female AV	-0.62	0.26	-2.39	0.031
			Male Child	-0.80	0.59	-1.35	0.310
			Male Adult	-0.93	0.27	-3.41	0.005
			Male Aud	-0.72	0.33	-2.16	0.063
			Male AV	-1.12	0.32	-3.46	0.014
Euclidean Distance (F1+F2)	/i/	All	0.17	0.15	1.09	0.280	
		Female	0.35	0.19	1.82	0.079	

		Male	-0.17	0.21	-0.81	0.431
		Child	0.06	0.61	0.10	0.926
		Adult	0.19	0.15	1.25	0.219
		Aud	0.08	0.25	0.33	0.745
		AV	0.26	0.17	1.50	0.147
		Female Child	0.73	0.84	0.87	0.478
		Female Adult	0.30	0.18	1.63	0.114
		Female Aud	0.55	0.30	1.82	0.090
		Female AV	0.16	0.25	0.65	0.526
		Male Child	-0.61	0.23	-2.68	0.115
		Male Adult	-0.04	0.25	-0.18	0.860
		Male Aud	-0.66	0.12	-5.68	0.001
		Male AV	0.45	0.12	3.75	0.010
	/u/	All	-1.14	0.20	-5.81	0.000
		Female	-1.15	0.26	-4.36	0.000
		Male	-1.11	0.29	-3.78	0.002
		Child	-1.71	0.55	-3.11	0.007
		Adult	-1.08	0.23	-4.81	0.000
		Aud	-1.30	0.32	-4.09	0.000
		AV	-1.04	0.34	-3.07	0.006
		Female Child	-2.13	1.29	-1.66	0.239
		Female Adult	-1.01	0.25	-4.08	0.000
		Female Aud	-1.56	0.38	-4.12	0.001
		Female AV	-0.74	0.33	-2.27	0.040
		Male Child	-0.66	0.46	-1.44	0.288
		Male Adult	-1.24	0.35	-3.49	0.005
		Male Aud	-0.70	0.33	-2.11	0.068
		Male AV	-1.62	0.42	-3.86	0.008
F1	/i/	All	0.13	0.12	1.04	0.304
		Female	0.28	0.16	1.73	0.094
		Male	-0.14	0.16	-0.91	0.377
		Child	0.17	0.66	0.26	0.801
		Adult	0.12	0.10	1.24	0.224
		Aud	0.14	0.21	0.67	0.511
		AV	0.11	0.12	0.92	0.367
		Female Child	1.06	0.84	1.27	0.333
		Female Adult	0.17	0.13	1.34	0.191
		Female Aud	0.46	0.27	1.72	0.107
		Female AV	0.10	0.18	0.57	0.577
		Male Child	-0.71	0.08	-9.24	0.012
		Male Adult	0.02	0.15	0.14	0.895
		Male Aud	-0.36	0.24	-1.51	0.169
		Male AV	0.13	0.09	1.43	0.203
	/u/	All	-0.34	0.08	-4.26	0.000
		Female	-0.23	0.10	-2.40	0.023
		Male	-0.52	0.12	-4.40	0.000
		Child	-0.69	0.02	-29.18	0.000
		Adult	-0.33	0.13	-2.58	0.015
		Aud	-0.27	0.14	-1.96	0.061
		AV	-0.33	0.20	-1.68	0.108
		Female Child	-0.70	0.02	-42.99	0.001

		Female Adult	-0.17	0.10	-1.68	0.104
		Female Aud	-0.21	0.16	-1.32	0.208
		Female AV	-0.25	0.12	-2.11	0.053
		Male Child	-0.68	0.05	-12.71	0.006
		Male Adult	-0.48	0.15	-3.19	0.008
		Male Aud	-0.60	0.11	-5.57	0.001
		Male AV	-0.43	0.24	-1.74	0.132
F2	<i>/i/</i>	All	0.13	0.26	0.52	0.607
		Female	0.29	0.38	0.76	0.456
		Male	-0.14	0.22	-0.65	0.527
		Child	-1.20	0.86	-1.39	0.215
		Adult	0.39	0.23	1.66	0.104
		Aud	-0.07	0.45	-0.15	0.881
		AV	0.35	0.16	2.21	0.038
		Female Child	-2.19	1.54	-1.42	0.291
		Female Adult	0.64	0.30	2.13	0.043
		Female Aud	0.27	0.70	0.39	0.702
		Female AV	0.31	0.24	1.30	0.215
		Male Child	-0.20	0.25	-0.82	0.501
		Male Adult	-0.13	0.24	-0.52	0.612
		Male Aud	-0.61	0.22	-2.85	0.022
	Male AV	0.44	0.12	3.66	0.011	
	<i>/u/</i>	All	-1.10	0.21	-5.35	0.000
		Female	-1.14	0.27	-4.18	0.000
		Male	-1.03	0.32	-3.19	0.006
		Child	-1.66	0.57	-2.93	0.010
		Adult	-1.05	0.22	-4.71	0.000
		Aud	-1.31	0.31	-4.17	0.000
		AV	-0.93	0.38	-2.46	0.023
		Female Child	-2.04	1.40	-1.46	0.283
		Female Adult	-1.01	0.26	-3.93	0.001
		Female Aud	-1.59	0.36	-4.36	0.001
		Female AV	-0.68	0.33	-2.05	0.060
		Male Child	-0.54	0.44	-1.23	0.345
		Male Adult	-1.18	0.39	-2.98	0.012
Male Aud		-0.59	0.33	-1.80	0.110	
Male AV	-1.59	0.52	-3.08	0.022		
F3	<i>/i/</i>	All	-0.13	0.06	-1.98	0.054
		Female	-0.07	0.09	-0.81	0.426
		Male	-0.22	0.08	-2.66	0.017
		Child	-0.43	0.07	-6.25	0.002
		Adult	-0.08	0.07	-1.17	0.251
		Aud	-0.12	0.09	-1.33	0.197
		AV	-0.13	0.09	-1.40	0.174
		Female Child	-0.23	0.22	-1.05	0.483
		Female Adult	-0.05	0.09	-0.57	0.575
		Female Aud	-0.02	0.12	-0.19	0.851
		Female AV	-0.11	0.12	-0.89	0.391
		Male Child	-0.51	0.02	-22.31	0.002
		Male Adult	-0.13	0.08	-1.70	0.114
		Male Aud	-0.26	0.11	-2.44	0.041

	Male AV	-0.16	0.14	-1.18	0.282
/u/	All	-0.11	0.06	-1.79	0.080
	Female	-0.10	0.06	-1.53	0.138
	Male	-0.14	0.14	-1.00	0.331
	Child	-0.26	0.16	-1.58	0.135
	Adult	-0.11	0.08	-1.41	0.170
	Aud	-0.07	0.06	-1.24	0.227
	AV	-0.11	0.15	-0.74	0.471
	Female Child	-0.32	0.04	-7.25	0.019
	Female Adult	-0.07	0.07	-0.96	0.345
	Female Aud	-0.14	0.11	-1.33	0.205
	Female AV	-0.05	0.07	-0.71	0.490
	Male Child	-0.48	0.31	-1.56	0.260
	Male Adult	-0.04	0.13	-0.32	0.753
	Male Aud	-0.31	0.20	-1.55	0.161
Male AV	0.07	0.16	0.44	0.677	

Children: pre-exposure/post-exposure

			Estimate	Std. Error	t-value	p-value
Duration	English-like	All	8.17	2.72	3.01	0.003
		Female	6.53	3.03	2.15	0.032
		Male	11.03	5.29	2.08	0.037
		Adult	42.08	11.90	3.54	0.000
		Aud	1.67	1.63	1.02	0.306
		AV	15.66	4.86	3.22	0.001
		Child	0.11	1.76	0.06	0.950
		Female Child	25.45	15.62	1.63	0.104
		Female Adult	3.71	1.78	2.09	0.037
		Female Aud	11.39	5.52	2.06	0.039
		Female AV	1.68	2.27	0.74	0.458
		Male Child	59.30	18.65	3.18	0.002
		Male Adult	-2.27	3.29	-0.69	0.491
		Male Aud	22.32	9.11	2.45	0.015
		Male AV	-2.93	2.70	-1.09	0.277
	Foreign	All	7.49	3.47	2.16	0.031
		Female	7.47	4.11	1.82	0.069
		Male	7.52	6.31	1.19	0.234
		Adult	4.93	7.05	0.70	0.485
		Aud	11.89	6.46	1.84	0.066
		AV	4.67	5.08	0.92	0.358
		Child	16.32	5.16	3.16	0.002
		Female Child	40.23	21.71	1.85	0.066
		Female Adult	2.56	2.29	1.12	0.264
		Female Aud	14.87	7.57	1.96	0.050

		Female AV	0.13	2.91	0.04	0.965	
		Male Child	39.04	21.89	1.78	0.076	
		Male Adult	-1.14	3.89	-0.29	0.770	
		Male Aud	15.94	10.83	1.47	0.142	
		Male AV	-2.72	3.48	-0.78	0.434	
f0	English-like	All	4.51	0.92	4.88	0.000	
		Female	3.02	1.21	2.50	0.013	
		Male	7.10	1.40	5.08	0.000	
		Adult	2.20	2.48	0.89	0.376	
		Aud	4.96	0.98	5.06	0.000	
		AV	7.49	1.48	5.07	0.000	
		Child	1.30	1.06	1.22	0.222	
		Female Child	8.90	3.31	2.69	0.008	
		Female Adult	2.14	1.28	1.67	0.095	
		Female Aud	7.53	1.99	3.78	0.000	
		Female AV	-1.50	1.35	-1.11	0.266	
		Male Child	-4.70	3.51	-1.34	0.182	
		Male Adult	10.35	1.45	7.16	0.000	
		Male Aud	7.44	2.10	3.54	0.000	
		Male AV	6.67	1.69	3.95	0.000	
		Foreign	All	3.95	1.20	3.28	0.001
			Female	1.99	1.61	1.24	0.217
			Male	7.32	2.53	4.28	0.000
			Adult	4.93	2.11	1.95	0.052
			Aud	0.41	1.79	0.19	0.847
			AV	3.88	1.78	2.17	0.030
			Child	2.82	4.11	1.59	0.112
			Female Child	6.51	1.73	1.58	0.115
			Female Adult	1.31	2.67	0.76	0.448
			Female Aud	6.37	1.78	2.39	0.017
			Female AV	-2.39	4.08	-1.35	0.179
			Male Child	-5.34	1.80	-1.31	0.192
			Male Adult	10.79	2.57	6.00	0.000
			Male Aud	5.74	2.10	2.24	0.026
			Male AV	9.23	0.03	4.40	0.000
Euclidean Distance (F1+F2+F3)	English-like	All	-0.10	0.04	-3.06	0.002	
		Female	-0.10	0.05	-2.65	0.008	
		Male	-0.08	0.09	-1.64	0.102	
		Adult	-0.37	0.03	-3.89	0.000	
		Aud	-0.05	0.05	-1.65	0.098	
		AV	-0.13	0.04	-2.66	0.008	
		Child	-0.06	0.16	-1.48	0.140	
		Female Child	-0.59	0.04	-3.74	0.000	
		Female Adult	-0.06	0.06	-1.53	0.127	
		Female Aud	-0.14	0.05	-2.27	0.024	
		Female AV	-0.07	0.12	-1.33	0.185	
		Male Child	-0.24	0.05	-1.99	0.047	
		Male Adult	-0.04	0.08	-0.73	0.469	
		Male Aud	-0.11	0.06	-1.46	0.144	
		Male AV	-0.04	0.04	-0.69	0.494	
		Foreign	All	-0.14	0.05	-3.71	0.000

		Female	-0.12	0.06	-2.61	0.009	
		Male	-0.17	0.07	-2.70	0.007	
		Adult	-0.12	0.06	-1.71	0.087	
		Aud	-0.22	0.05	-3.43	0.001	
		AV	-0.14	0.06	-2.62	0.009	
		Child	-0.16	0.16	-2.82	0.005	
		Female Child	-0.23	0.05	-1.41	0.162	
		Female Adult	-0.11	0.07	-2.32	0.021	
		Female Aud	-0.15	0.06	-2.10	0.036	
		Female AV	-0.09	0.17	-1.45	0.148	
		Male Child	-0.47	0.06	-2.80	0.006	
		Male Adult	-0.08	0.10	-1.29	0.199	
		Male Aud	-0.19	0.08	-1.97	0.050	
		Male AV	-0.14	0.03	-1.85	0.065	
Euclidean Distance (F1+F2)	English-like	All	-0.09	0.04	-2.70	0.007	
		Female	-0.12	0.06	-2.80	0.005	
		Male	-0.05	0.10	-0.84	0.403	
		Adult	-0.25	0.03	-2.52	0.012	
		Aud	-0.06	0.05	-1.84	0.066	
		AV	-0.08	0.04	-1.54	0.125	
		Child	-0.10	0.15	-2.35	0.019	
		Female Child	-0.32	0.04	-2.20	0.029	
		Female Adult	-0.09	0.06	-2.14	0.033	
		Female Aud	-0.12	0.06	-1.88	0.060	
		Female AV	-0.11	0.14	-2.05	0.041	
		Male Child	-0.18	0.06	-1.35	0.179	
		Male Adult	-0.01	0.08	-0.23	0.815	
		Male Aud	-0.01	0.07	-0.16	0.876	
	Male AV	-0.09	0.04	-1.22	0.221		
		Foreign	All	-0.13	0.05	-3.09	0.002
			Female	-0.13	0.07	-2.68	0.008
			Male	-0.11	0.08	-1.64	0.101
			Adult	-0.11	0.07	-1.38	0.168
			Aud	-0.23	0.06	-3.15	0.002
			AV	-0.10	0.06	-1.62	0.106
			Child	-0.17	0.16	-2.88	0.004
			Female Child	-0.03	0.05	-0.21	0.835
			Female Adult	-0.15	0.08	-2.89	0.004
			Female Aud	-0.14	0.07	-1.77	0.077
			Female AV	-0.13	0.19	-2.04	0.042
	Male Child		-0.37	0.07	-1.93	0.055	
	Male Adult	-0.05	0.11	-0.69	0.489		
	Male Aud	-0.07	0.09	-0.61	0.542		
	Male AV	-0.18	0.03	-2.06	0.040		
F1	English-like	All	0.05	0.04	1.62	0.106	
		Female	0.02	0.05	0.43	0.671	
		Male	0.11	0.09	2.03	0.043	
		Adult	-0.03	0.03	-0.30	0.763	
		Aud	0.06	0.05	1.97	0.049	
		AV	-0.02	0.04	-0.42	0.675	
		Child	0.12	0.11	3.13	0.002	

		Female Child	-0.29	0.04	-2.53	0.012	
		Female Adult	0.06	0.06	1.56	0.120	
		Female Aud	-0.04	0.05	-0.69	0.491	
		Female AV	0.07	0.12	1.42	0.156	
		Male Child	0.25	0.06	2.08	0.039	
		Male Adult	0.07	0.08	1.19	0.235	
		Male Aud	0.01	0.07	0.16	0.874	
		Male AV	0.23	0.04	3.43	0.001	
	Foreign	All	0.12	0.05	2.96	0.003	
		Female	0.13	0.07	2.50	0.013	
		Male	0.12	0.08	1.60	0.110	
		Adult	0.15	0.07	1.83	0.068	
		Aud	0.18	0.06	2.42	0.016	
		AV	0.09	0.06	1.43	0.152	
		Child	0.17	0.15	2.79	0.005	
		Female Child	-0.12	0.05	-0.83	0.406	
		Female Adult	0.16	0.08	3.04	0.002	
		Female Aud	0.10	0.07	1.34	0.182	
		Female AV	0.15	0.19	2.22	0.027	
		Male Child	0.23	0.08	1.25	0.214	
		Male Adult	0.09	0.11	1.11	0.269	
		Male Aud	-0.06	0.10	-0.58	0.561	
		Male AV	0.34	0.04	3.56	0.000	
F2		English-like	All	0.01	0.05	0.34	0.733
	Female		0.04	0.06	0.83	0.407	
	Male		-0.03	0.12	-0.50	0.619	
	Adult		0.29	0.04	2.55	0.011	
	Aud		-0.03	0.06	-0.91	0.363	
	AV		0.04	0.05	0.67	0.506	
	Child		-0.01	0.15	-0.27	0.784	
	Female Child		0.50	0.05	3.26	0.001	
	Female Adult		-0.02	0.07	-0.50	0.618	
	Female Aud		0.04	0.06	0.60	0.549	
	Female AV		0.03	0.08	0.55	0.579	
	Male Child		0.04	0.06	0.27	0.786	
	Male Adult		-0.05	0.09	-0.87	0.383	
	Male Aud		0.02	0.07	0.18	0.854	
	Male AV		-0.10	0.10	-1.38	0.169	
			Foreign	All	0.08	0.05	1.95
		Female		0.07	0.06	1.49	0.136
		Male		0.08	0.07	1.28	0.200
		Adult		-0.01	0.07	-0.18	0.857
		Aud		0.22	0.06	3.17	0.002
		AV		0.07	0.06	1.19	0.233
		Child		0.10	0.14	1.82	0.070
		Female Child		0.24	0.05	1.70	0.092
		Female Adult		0.05	0.07	1.03	0.301
		Female Aud		0.10	0.07	1.37	0.171
		Female AV		0.05	0.08	0.76	0.445
		Male Child		0.32	0.06	1.88	0.062
		Male Adult		0.02	0.10	0.28	0.783

		Male Aud	0.19	0.07	1.96	0.051
		Male AV	-0.05	0.11	-0.67	0.507
F3	English-like	All	0.02	0.02	1.14	0.254
		Female	0.00	0.03	-0.18	0.854
		Male	0.07	0.06	1.88	0.060
		Adult	0.13	0.02	2.22	0.027
		Aud	0.00	0.03	0.17	0.862
		AV	0.08	0.03	2.61	0.009
		Child	-0.04	0.09	-1.32	0.187
		Female Child	-0.02	0.03	-0.19	0.850
		Female Adult	-0.01	0.04	-0.25	0.806
		Female Aud	0.02	0.03	0.61	0.543
		Female AV	-0.03	0.04	-1.01	0.311
		Male Child	0.22	0.04	3.08	0.002
		Male Adult	0.02	0.05	0.49	0.626
		Male Aud	0.15	0.05	3.14	0.002
		Male AV	-0.05	0.06	-0.98	0.329
		Foreign	All	0.03	0.03	1.19
	Female		-0.02	0.04	-0.52	0.606
	Male		0.10	0.04	2.28	0.023
	Adult		0.06	0.04	1.39	0.166
	Aud		0.03	0.04	0.64	0.522
	AV		0.01	0.04	0.39	0.699
	Child		0.04	0.11	1.17	0.242
	Female Child		-0.22	0.03	-2.05	0.043
	Female Adult		0.00	0.05	0.09	0.930
	Female Aud		0.00	0.04	0.01	0.995
	Female AV		-0.03	0.05	-0.73	0.464
	Male Child		0.31	0.05	3.24	0.001
	Male Adult	0.04	0.06	0.78	0.437	
Male Aud	0.19	0.05	2.96	0.003		
Male AV	-0.01	0.08	-0.24	0.814		

References

- Adamson, L., & Frick, J. (2003). The Still-Face: A history of a shared experimental paradigm. *Infancy*, 4, 451-473.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV). Washington, DC: APA.
- Babel, M. & Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, 1-18.
- Babel, M. (2009). *Phonetic and social selectivity in speech accommodation*. Doctoral Dissertation, University of California Berkeley.
- Babel, M. (2010). Dialect convergence and divergence in New Zealand English. *Language in Society*, 39 (4), 437-456.
- Bahrick, L. & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, 36(2), 190-201.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): evidence from Asperger Syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31, 5-17.
- Bebko, J., Weiss, J., Demark, J., & Gomez, P. (2006). Discrimination of temporal synchrony intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry*, 47, 88-98.
- Bernstein, L., Auer, E., Jr., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, 44, 5–18.

- Bernstein Ratner N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics*, 14, 303–309.
- Biersack, S., Kempe, V., & Knapton, L. (2005). Fine-tuning speech registers: a comparison of the prosodic features of child-directed and foreigner-directed speech. In: *Proceedings of the 9th European Conference on Speech Communication and Technology*, Lisbon.
- Binnie, C. Montgomery, A., & Jackson, P. (1974). Auditory and visual contribution to the perception of consonants. *Journal of Speech and Hearing Research*, 17, 619-630.
- Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355-387.
- Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [computer program]. Version t.3.53, retrieved 9 July 2013 from <http://www.praat.org/>
- Boucher, J., Lewis, V., & Collis, G. (1998). Familiar face and voice matching and recognition in children with autism. *Journal of Child Psychology and Psychiatry*, 39, 171–181.
- Bourhis, R. & Giles, H. (1977). The language of intergroup distinctiveness. In *Language, ethnicity, and intergroup relations*, ed. Howard Giles, 119–136.
- Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 445-463.
- Brugger, A., Larivière, L., Mumme, D., and Bushnell, E. (2007). Doing the right thing: Infants' selection of actions to imitate from observed event sequences. *Child Development*, 78, 806-824.
- Burnham, D., & Dodd, B. (2004). Auditory visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, 44, 209220.

- Burnham D., Lau S., Tam H., & Schoknecht C. (2001). Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers, and by non-tonal language speakers. In *Proceedings of the International Conference on Auditory-Visual Speech Processing*, eds. Massaro D., Light J., and Geraci K., 155–160.
- Busby, P. & Plant, G. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *Journal of the Acoustical Society of America*, 97(4), 2603-2606.
- Calliope (1989) *La parole et son traitement automatique*, Collection Technique et Scientifique des Télécommunications, Ed. Masson.
- Cavé, C., Guaitella, I., Bertrand, R., Santi, S., Harlay, F., & Espesser, R. (1996). About the relationship between eyebrow movements and F0 variation. *ICSLP 96*, 4, 2175–2179.
- Chartrand, T. & Bargh, J. (1996). The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76, 893–910.
- Clopper, C. & Pierrehumbert, J. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America*, 124, 1682–1688.
- Clopper, C. & Pisoni, D. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, 32, 111–140.
- Clopper, C., Pisoni, D., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*, 118, 1661–1676.
- Cooper, R., & Aslin, R. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, 61, 1584–1595.
- de Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, 3, 69–86.

- Desjardins, R. & Werker, J. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, 45(4), 187-203.
- Desjardins, R., Rogers, J., & Werker, J. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology*, 66(1), 85-110.
- Dias, J. & Rosenblum, L. (2011). Visual influences on interactive speech alignment. *Perception*, 40, 1457-1466.
- Eaton, C. & Bernstein Ratner, N. (to appear). Rate and phonological variation in preschool children: effects of modeling and directed influence. *Journal of Speech, Language, and Hearing Research*.
- Fant, G. (1959). Acoustic analysis and synthesis of speech with application to Swedish. *Ericsson Technics*, 15, 3-108.
- Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.
- Fant, G. (1983). Feature analysis of Swedish vowels –a revisit. *STL-QPSR* 2-3, 1-19.
- Ferguson, C. (1964). Baby talk in six languages. *American Anthropologist*, 66, 103–114.
- Ferguson, C. (1977). Baby talk as a simplified register. In *Talking to Children*, C.E. Snow and C.A. Ferguson (eds.). Cambridge: Cambridge University.
- Fernald, A. & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20(1), 104-113.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3), 477-501.
- Fernald, A. (1982). *Acoustic determinants of infant preference for motherese*. Doctoral

- Dissertation, University of Oregon.
- Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. In L. Feagans, C. Garvey, R. Golinkoff (Eds.), *The origins and growth of communication*, Ablex, Norwood, NJ.
- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8, 181-195.
- Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, 64, 657–674.
- Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development*, 67, 3192–3218.
- Fisher, C. (1968). Confusions among visually perceived consonants. *Journal of Speech & Hearing Research*, 11(4), 796-804.
- Flege, J. (1987). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, 8(2), 162-177.
- Francis A., Kaganovich N., & Driscoll-Huber C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *Journal of the Acoustical Society of America*, 124, 1234–1251.
- Freed, B. (1981). Foreigner talk, baby talk, native talk. *International Journal of Society Language*, 28, 19-39.
- Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11, 149–75.
- Garnica, O. (1977). Some prosodic and paralinguistic features of speech to young children. In C.E. Snow, C.A. Ferguson (Eds.), *Talking to children: Language input and acquisition*, Cambridge University Press, Cambridge, England.

- Garrod, S. & Doherty, G. (1994). Conversation, coordination, and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition*, 53, 181-215.
- Gendrot, C., & Adda-Decker, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. *Variations*, 2(22.5), 2-4.
- Gergerly, G., Bekkering, H., & Kiraly, I., (2002). Developmental psychology: rational imitation in preverbal infants. *Nature*, 415, 755.
- Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- Green, K. & Gerdeman, A. (1995). Cross-Modal discrepancies in coarticulation and the integration of speech information: the McGurk effect with mismatched vowels. *Journal of Experiment Psychology: Human Perception and Performance*, 21(6), 1409-1426.
- Green, K., & Miller, J. (1985). On the role of visual rate information in phonetic perception. *Perception & Psychophysics*, 38, 269–276.
- Green, K., Kuhl, P., & Meltzoff, A. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception and Psychophysics*, 50, 524-536.
- Green, J., Nip, I., Wilson, E., Mefferd, A., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*, 53, 1529-1542.
- Gregory, S., Dagan, K., & Webster, S. (1997). Evaluating the relation between vocal accommodation in conversational partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior*, 21, 23–43.

- Gregory, S., Green, B., Carrothers, R., & Dagan, K. (2001). Verifying the primacy of voice fundamental frequency in social status accommodation. *Language and Communication*, 21, 37–60.
- Gregory, S. & Hoyt, B. (1982). Conversation partner mutual adaptation as demonstrated by fourier series analysis. *Journal of Psycholinguistic Research*, 11, 35–46.
- Gregory, S. & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, 70, 1231–1240.
- Gregory, S., Webster, S., & Huang, G. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language and Communication*, 13, 195–217.
- Grieser, D. & Kuhl, P. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, 24, 14–20.
- Hanna, E. & Meltzoff, A. N. (1993). Peer imitation by toddlers in laboratory, home, and day-care contexts: Implications for social learning and memory. *Developmental Psychology*, 29, 701-710.
- Hardison, D. (2003). Acquisition of second-language speech: effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, 24, 495-522.
- Harrell, F. (2001). Regression modeling strategies. Oxford: Springer- Verlag.
- Hasek, C., Singh, S., & Murry, T. (1980). Acoustic attributes of preadolescent voices. *Journal of the Acoustical Society of America*, 68(5), 1262-1265.
- Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perception training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3), 360-378.

- Hietanen, J., Manninen, P., Sams, M., & Surakka, V. (2001). Does audiovisual speech perception use information about facial configuration? *European Journal of Cognitive Psychology*, 13, 395-407.
- Hockley, N. & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *Journal of the Acoustical Society of America*, 96, 3309.
- Iarocci, G. & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *Journal of Autism and Developmental Disorders*, 36, 77–90
- Irwin, J., Whalen, D., & Fowler, C. (2006). A sex difference in visual influence on heard speech. *Perception and Psychophysics*, 68, 582-592.
- Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47–B57.
- Jacobson, J., Boersma, D., Fields, R., & Olson, K. (1983). Paralinguistic features of adult speech to infants and small children. *Child Development*, 54, 436–442.
- Jaeger, T. & Snider, N. (2013). Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime’s prediction error given both prior and recent experience. *Cognition*, 127, 57-83.
- Jaeger, T. (2011). Corpus-based research on language production: Information density and reducible subject relatives. In E. M. Benders & J. E. Arnold (Eds.), *Language from a cognitive perspective: Grammar, usage, and processing studies in honor of Tom Wasow* (161–197). Stanford: CSLI Publications.
- Jerger, S., Damian, M., Spence, M., Tye-Murray, N., & Abdi, H. (2009). Developmental shifts in children's sensitivity to visual speech: A new multimodal picture word task. *Journal of*

- Experimental Child Psychology*, 102, 40–59.
- Jiang, J. (2003). Relating optical speech to speech acoustic and visual speech perception. Ph.D. dissertation, Department of Electrical Engineering, UCLA.
- Johnson, F., Hicks, L., Goldberg, T., & Myslobodsky, M. (1988). Sex differences in lipreading. *Bulletin of the Psychonomic Society*, 26, 106-108.
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. *Journal of Phonetics*, 43, 485-499.
- Kawahara, H., de Cheveigne, A., & Patterson, R. (1998). An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: revised TEMPO in the STRAIGHT-suite. *ICSLP-1998*, paper 0659.
- Keane, B., Rosenthal, O., Chun, N., & Shams, L. (2010). Audiovisual integration in high functioning adults with autism. *Research in Autism Spectrum Disorders*, 4, 276-289.
- Kemler Nelson, D., Hirsh-Pasek, K., Jusczyk, P., & Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, 16, 55–68.
- Kim, S. & Lee, E. (2001). A comparative study of Korean and French vowel systems. *Speech Sciences*, 8, 53-66
- Kim, M. (2011). Phonetic convergence after perceptual exposure to native and nonnative speech: Preliminary findings based on fine-grained acoustic-phonetic measurement. *Proceedings of The 17th International Congress of Phonetic Sciences*.
- Kim, M., Horton, W., & Bradlow, A. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Journal of Laboratory Phonology*, 2, 125-156.
- Kugiumutzakis, G. (1999) Genesis and development of early infant mimesis to facial and vocal

- models, J. Nadel, G. Butterworth, Eds., *Imitation in Infancy*, Cambridge University Press, pp. 36–59.
- Kuhl, P. & Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science*, 218, 1138-1141.
- Kuhl, P. & Metzloff, A. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100, 2425–2438.
- Kuhl, P., Andruski, J., Chistovich, I., Chistovich, L., Kozhevnikova, E., & Ryskina, V. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277, 684-686.
- Kushnerenko E., Teinonen T., Volein A., & Csibra G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences*, 105(11), 442–445.
- Langdell, T. (1978). Recognition of faces: an approach to the study of autism. *Journal of Child Psychology and Psychiatry*. 19, 255–268.
- Lansing, C. & McConkie, G. (1999). Attention to facial regions in segmental and prosodic visual speech perception tasks. *Journal of Speech, Language, and Hearing Research*, 42, 526–539.
- Lenneberg, E. (1967). *Biological Foundations of Language*. New York: John Wiley and Sons.
- Lindell, A., Notice, K., & Withers, K. (2009). Reduced language processing asymmetry in non-autistic individual with high levels of autism traits. *Laterality: Asymmetries of Body, Brain, and Cognition*, 14(5), 457-472.
- Linden, S. & Vroomen, J. (2008). Audiovisual speech recalibration in children. *Journal of Child Language*, 35, 809-822.

- Liu, H., Kuhl, P., & Tsao, F. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, 6, F1–F10.
- Massaro, D. & Cohen, M. (1990). Perception of synthesized audible and visible speech. *Psychological Science*, 1, 55-63.
- Massaro, D., Thompson, L., & Laron, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, 41, 93–113.
- Massaro, D. (1984). Children's perception of visual and auditory speech. *Child Development*, 55, 1777-1788.
- Massaro, D. (1987). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale, NJ: Erlbaum.
- Massaro, D. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. MIT Press, Cambridge, MA
- Massaro, D., Tsuzaki, M., Cohen, M., Gesi, A., & Heridia, R. (1993). Bimodal speech perception: an examination across languages. *Journal of Phonetics*, 21, 445-478.
- McGurk, H., & MacDonald, J. (1976) Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Meltzoff, A. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31, 838-850.
- Meltzoff, A. (2007). "Like me": A foundation for social cognition. *Developmental Science*, 10, 12~134.
- Meltzoff, A., Kuhl, P., Movellan, J., & Sejnowski, T. (2009). Foundations for a new science of learning. *Science*, 325, 284-288.
- Meltzoff, A. & Moore, M. (1977). Imitation of facial and manual gestures by human neonates.

- Science*, 198, 75-78.
- Meltzoff, A. & Moore, M. (1983). Newborn infants imitate adult facial gestures. *Child Development*, 54, 702-709.
- Meltzoff, A. & Moore, M. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, 25, 954-962.
- Meltzoff, A. & Moore, M. (1994). Imitation, memory, and the representation of persons. *Infant Behavior and Development*, 17, 83-99.
- Meltzoff, A. & Moore, M. (1997). Explaining facial imitation: a theoretical model. *Early Development and Parenting*, 6, 179-192.
- Meltzoff, A., & Williamson, R. (2010). The importance of imitation for theories of social-cognitive development. In G. Bremner & T. Wachs (Eds.), *Handbook of infant development* (2nd ed., 345-364). Oxford: Wiley-Blackwell.
- Meltzoff, A. & Williamson, R. (2013). Imitation: social, cognitive, and theoretical perspectives. In P. R. Zelazo (Ed.). *Oxford handbook of developmental psychology* (vol. 1, pp. 651-682). NY: Oxford University Press.
- Mielke, J., Nielsen, K., & Magloughlin, L. (2013). Phonetic imitation by individuals with autism spectrum disorders: investigating the role of procedural and declarative memory. *Proceedings of Meetings on Acoustics*, 19.
- Miller, R., Sanchez, K., & Rosenblum, L. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics*, 72. 1614–1625.
- Mitterer, H. (2013). Phonetic convergence is not a consequence of stimulus-response compatibility. Talk presented at the ASA ICA Conference, Montreal, Quebec.
- Morton, E. (1977). On the occurrence and significance of motivation structural rules in some

- bird and mammal sounds. *American Naturalist*, 111, 855–869.
- Munhall, K., Jones, J., Callan, D., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility. *Psychological Science*, 15, 133–137.
- Musacchia, G., Sams, M., Nicol, T., & Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Experimental Brain Research*, 168(1-2), 1-10.
- Nagy, E., & Molnar, P. (1994). Homo imitans or homo provocans? *International journal of Psychophysiology*, 18, 128.
- Nagy, E., & Molnar, P. (2004). Homo imitans or homo provocans? Human imprinting model of neonatal imitation. *Infant Behavior & Development*, 27, 54-63.
- Namy, L., Nygaard, L., & Sauerteig, D. (2002). Gender differences in vocal accommodation: the role of perception. *Journal of Language and Social Psychology*, 21, 422–432.
- Natale, M. (1975a). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32, 790–804.
- Natale, M. (1975b). Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills*, 40, 827–830.
- Navarra, J. & Soto-Faraco, S. (2007). Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychological Research*, 71, 4-12.
- Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, 42, 555-565.
- Nielsen, K. (2004). Segmental differences in the visual contribution to speech intelligibility. Masters Thesis. University of California, Los Angeles.

- Nielsen, K. (2008). The specificity of allophonic variability and its implications for accounts of speech perception. Doctoral Dissertation, University of California, Los Angeles.
- Nielsen, K. (2011a). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39, 132-142.
- Nielsen, K. (2011b). Phonetic imitation by school-age children. Poster presented at the 162nd Meeting of the Acoustical Society of America. San Diego, CA.
- Nye, P. & Fowler, C. (2003). Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. *Journal of Phonetics*, 31(1), 63-79.
- Ota, M. & Stewart, M. (2007). Autism and lexical context effects on speech perception. *Proceedings of ICPHS XVI, 2021-2024*.
- Owens, E. & Blazek, B. (1985). Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech and Hearing Research*, 28, 381-393.
- Pardo, J. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119, 2382–2393.
- Pardo, J., Jay, I., & Krauss, R. (2010). Conversational role influences speech imitation. *Attention, Perception & Psychophysics*, 72(8), 2254-2264.
- Patterson, M., & Werker, J. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, 22, 237-247.
- Patterson, M., & Werker, J. (2002). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6, 191-196.
- Patterson, M. & Werker, J. (2003). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, 81, 93-115.
- Penfield, W. & Roberts L. (1959). *Speech and Brain Mechanisms*. Princeton, N.J.: Princeton

University Press.

- Peterson, G. & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 119, 2382-2393.
- Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169–226.
- Pickering, M. & Ferreira, V. (2008). Structural priming: a critical review. *Psychological Bulletin*, 134, 427-459.
- Pinker, S. (1994). *The language instinct*. New York: Morrow.
- R Development Core Team. 2008. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>, ISBN 3-900051-07-0.
- Reisberg, D., McLean, J. and Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In R. Campbell and B. Dodd (Eds.), *Hearing by Eye: The Psychology of Lip-Reading* (pp. 97-114). Hillsdale, N.J.: Erlbaum
- Remick, H. (1976). Maternal speech to children during language acquisition. In W. von Raffler-Engel & Y. Lebrun (Eds.), *Baby talk and infant speech* (pp. 223–233). Amsterdam: Swets & Zeitlinger.
- Repacholi, B., Meltzoff, A., & Olsen, B. (2008). Infants' understanding of the link between visual perception and emotion: "If she can't see me doing it, she won't get angry". *Developmental Psychology*, 44, 561-574.
- Ronquest, R., Levi, S., & Pisoni, D. (2010). Language identification from visual-only speech. *Attention, Perception, and Psychophysics*, 72, 1601-1613.
- Rosen, S. & Corcoran, T. (1982). A video-recorded test of lipreading for British English. *British*

- Journal of Audiology*, 16, 245-254.
- Rosenblum, L. (2005). The primacy of multimodal speech perception. In D. Pisoni & R. Remez (Eds.). *Handbook of Speech Perception*. Blackwell: Malden, MA. pp. 51-78.
- Rosenblum, L. (2008). Speech Perception as a Multimodal Phenomenon. *Current Directions in Psychological Science*, 17, 405-409.
- Rosenblum, L., Miller, R. & Sanchez, K. (2007). Lipread me now, hear me better later: Crossmodal transfer of talker familiarity effects. *Psychological Science*, 18, 392-396.
- Rosenblum, L. & Saldana, H. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 318-331.
- Rosenblum, L., Schmuckler, M., & Johnson, J. (1997). The McGurk effect in infants. *Perception and Psychophysics*, 59, 347357.
- Ryalls, B. & Pisoni, D. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, 33, 441-452.
- Sachs, J. (1977). The adaptive significance of linguistic input to prelinguistic infants. In C.E. Snow & C.A. Ferguson (Eds.), *Talking to children: Language input and acquisition*. Cambridge, England: Cambridge University Press.
- Sanchez, K., Miller, R., & Rosenblum, L. (2010). Visual influences on alignment to voice onset time. *Journal of Speech, Language, and Hearing Research*, 53, 262-272.
- Scarborough, R., Keating, P., Mattys, S. L., Cho, T. & Alwan, A. (2009). Optical phonetics and visual perception of lexical and phrasal stress in English. *Language and Speech*, 52, 135-175.
- Schaal, D. (1999). Is imitation learning the route to humanoid robots? *Trends in Cognitive*

- Sciences*, 3(6), 233-242.
- Schachner, A., & Hannon, E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental Psychology*, 47, 19-25.
- Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143–165.
- Schorr, E., Fox, N., van Wassenhove, V., Knudsen, E. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proceedings of the National Academy of Sciences*, 102(51), 18748- 18750.
- Sekiyama, K. & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America*, 90, 1797-1805.
- Sekiyama, K. & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, 21, 427-444.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, 408, 788.
- Shockley, K., Sabadini, L. & Fowler, C. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66, 422–429.
- Shue, Y., Keating, P. & Vicenik, C. (2009). VOICESAUCE: A program for voice analysis. *Journal of the Acoustical Society of America*, 126, 2221.
- Singh, L., Morgan, J., & Best, C. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy*, 3, 365–394.
- Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of Infant-Directed Speech on Early Word Recognition. *Infancy*, 14, 654–666.

- Sjolander, K. (2004). Snack sound toolkit, KTH. Stockholm, Sweden.
<http://www.speech.kth.se/snack>.
- Smith, E. & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, 48(8), 813-821.
- Snow, C. (1977). Mothers' speech research: from input to interaction. In C.E. Snow, C.A. Ferguson (Eds.), *Talking to children: Language input and acquisition*, Cambridge University Press, Cambridge, England.
- Soto-Faraco, S., Navarra, J., Weikum, W., Vouloumanos, A., Sebastian-Galles, N., Werker, J. (2007). Discriminating languages by speechreading. *Perception and Psychophysics*, 69, 218-231.
- Stern, D., Spieker, S., Barnet, R., & MacKain, D. (1983). The prosody of maternal speech: infant age and context related changes. *Journal of Child Language*, 10, 1-15.
- Stevens, K. & House, A. (1963). Perturbation of vowel articulations by consonant context: An acoustical study. *Journal of Speech and Hearing Research*, 6, 111– 128.
- Stewart, M. & Ota, M. (2008). Lexical effects on speech perception in individuals with “autistic” traits. *Cognition*, 109(1), 157-162.
- Strange, W., Weber, A., Levy, E., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: phonetic context effects. *Journal of the Acoustical Society of America*, 122(2), 1111-1129.
- Street, R. Jr. Cappella, J. (1989). Social and linguistic factors influencing adaptation in children's speech. *Journal of Psycholinguistic Research*, 18, 497–519.
- Sumby, W. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.

- Summerfield, A. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London*, 335, 71-78.
- Summerfield, A. & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quarterly Journal of Experimental Psychology*, A36, 517-4.
- Sundara, M., Ward, N., Conboy, B., Kuhl, P. (submitted). Production benefits of short term exposure to a second language in infancy.
- Swanson, L., Leonard, L., & Gandour, J. (1992). Vowel duration in mothers' speech to young children. *Journal of Speech and Hearing Research*, 35, 617-625.
- Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F.R. Volkmar, R. Paul, K. A.J. & D.J. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders*, Vol. 1 (3rd edn., pp. 335–364). Hoboken: Wiley
- Thiessen, E., Hill, E., & Saffran, J. (2005). Infant directed speech facilitates word segmentation. *Infancy*, 7(1), 49-67.
- Thompson, L., Driscoll, D., & Markson, L. (1998). Memory for visual-spoken language in children and adults. *Journal of Nonverbal Behavior*, 22(3), 167-187.
- Trainor, L. & Desjardins, R. (2002). Pitch characteristics of infant-directed speech affects infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, 9(2), 335-340.
- Traunmüller, H. & Öhrström, N. (2007) Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics*, 35(2), 244-258.
- Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, 88, 97–100.
- Trudgill, P. (1981). Linguistic accommodation: Sociolinguistic observations on a

- sociopsychological theory. In *Papers from the Parasession on Language and Behavior*, ed. Roberta Hendrick, Carrie Masek, and Mary Frances Miller, 218–237. Chicago: Chicago Linguistics Society.
- Uther, M., Knoll, M., & Burnham, D. (2007). Do you speak E-NG-L-I-S-H? A comparison of foreigner- and infant-directed speech. *Speech Communication*, 17, 2-7.
- Vorperian, H. & Kent, R., (2007). Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. *Journal of Speech Language and Hearing Research*, 50(6), 1510-1545.
- Walden, B., Prosek, R., Montgomery, A., Scherr, C., & Johnes, C. (1977). Effect of training on the visual recognition of consonants. *Journal of Speech and Hearing Research*, 20, 130-145.
- Ward, N. & Sundara, M. (2010). Development of native language preference in the visual modality. Poster presented at *2nd Pan-American/Iberian Meeting on Acoustics*, Cancun, MX.
- Ward, N. & Sundara, M. (2012). The development and basis of visual language preference in infancy. Poster presented at *International Conference on Infant Studies*, Minneapolis, MN.
- Ward, N., Sundara, M., Conboy, B., & Kuhl, P. (2009). Consequences of short-term language exposure in infancy on babbling. Poster presented at *158th meeting of the Acoustical Society of America*, San Antonio, TX.
- Watson, C., Qui, W., Chamberlain, M., & Li, X. (1996). Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition. *Journal of the Acoustical Society of America*, 100, 1153–1162.

- Weikum, W., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastian-Galles, N., & Werker, J. (2007). Visual language discrimination in infancy. *Science*, 316, 1159.
- Welch, R. & Warren, D. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88, 638-667.
- Werker, J., & McLeod, P. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, 43(2), 230-246.
- Werker, J., McGurk, H., & Frost, P. (1992). La langue et les lèvres: Cross-language influences on bimodal speech perception. *Canadian Journal of Psychology*, 46(4), 551-568.
- Werker, J., Pegg, J., & McLeod, P. (1994). A cross-language comparison of infant preference for infant-directed communication. *Infant Behavior and Development*, 17(3), 323-333.
- Williams, I., Whiten, A., & Singh, T. (2004) A Systematic Review of Action Imitation in Autistic Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 34, 285–99.
- Williamson, R., Meltzoff, A., & Markman, E. (2008). Prior experiences and perceived efficacy influence 3-year-olds' imitation. *Developmental Psychology*, 44, 275-285.
- Woodhouse, L., Hickson, L., & Dodd, B. (2009). Review of visual speech perception by hearing and hearing - impaired people: Clinical implications. *International Journal of Language and Communication Disorders*, 44(3), 253-270.
- Yehia, H., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, 30, 555–568.
- Yu, A. (2010). Perceptual compensation is correlated with individuals “autistic” traits: implications for models of sound change. *PLoS ONE*, 5(8): e11950.

Yu, A., Grove, J., Martinovic, M., & Sondregger, M. (2011). Effects of working memory capacity and “autistic” traits on phonotactic effects in speech perception. *Proceedings of ICPhS XVII*, 2236-2239.