

Interactive Effects of Perception and Production Training on L2 Phonetic Acquisition

Ruri Ueda

A thesis submitted for the degree of Doctor of Philosophy of
The Australian National University



**Australian
National
University**

October 2024

© Copyright by Ruri Ueda, 2024
All Rights Reserved

Declaration

I hereby declare that this thesis is my own work and, to the best of my knowledge, does not contain any materials previously published or written by another person, to a substantial extent, has been accepted for the award of any other degree of diploma, except where otherwise indicated.

The ethical aspects of this research have been approved by the ANU Human Research Ethics Committee (Protocol 2020/523)

October 2024

A handwritten signature in black ink, consisting of two characters, '瑠璃' (Ruri), written in a cursive style.

Ruri Ueda

上田瑠璃

Acknowledgements

I would like to express my deepest gratitude to my primary supervisor, Dr Solène Inceoglu, for her kind encouragement, insightful feedback, and invaluable guidance throughout my PhD journey. Her expertise and encouragement have been instrumental in shaping this work. As a researcher and educator in the field of Second Language Acquisition, particularly in L2 phonetic learning, she has been an exceptional role model and has generously shared her knowledge and insights with me.

My heartfelt thanks also go to the rest of my supervisory panel. I am deeply grateful to Dr Ksenia Gnevsheva, whose course, consultations, and feedback equipped me with essential research skills, especially in quantitative analysis methods. Her advice has been invaluable in helping me navigate the complexities of my research topic. I am also indebted to Dr Ken-ichi Hashimoto for the immense support he gave me during my data collection with Japanese participants, particularly under the challenging circumstances of remote research during the COVID-19 pandemic. Without his assistance and dedication, this thesis would not have been possible. Additionally, I would like to thank Dr Toshiyuki Nakamura for his support, career guidance, and encouragement during my PhD journey.

I am grateful to all the participants—both L1 Japanese and L1 English speakers—who took part in this research. Your time, effort, and commitment to completing the tasks were invaluable to this research, and I greatly appreciate your contributions.

I also wish to extend my thanks to everyone who has encouraged and supported me along this journey. Your belief in me has been a constant source of motivation. Although I cannot name everyone, I would like to give special thanks to my undergraduate supervisor, Professor Hiroyuki Imai. His support has been unwavering, even after my graduation, and every conversation with him has enriched both my career and life, contributing significantly to my personal and intellectual growth. My sincere thanks also go to the late Professor Haruyo Yoshida, whose guidance and encouragement were instrumental in my decision to commence a PhD. I am deeply thankful for her support.

I must also extend my thanks to my colleagues and fellow teachers with whom I have had the privilege of working. Your mentorship and shared experiences have helped shape my academic path, and I am deeply grateful for your guidance and wisdom. I am also grateful to fellow HDR students. The mutual encouragement we shared made the PhD journey not only manageable but also rewarding.

To all of my students who have studied English or Japanese with me, thank you. Sharing your language learning journeys has been a privilege and a joy. Teaching you has not only helped me grow as an educator but also inspired me as a researcher. I hope that your language learning continues to bring you joy, as it has for me.

A special thanks goes to my friends, whose constant support, encouragement, and understanding have carried me through this journey. The moments of laughter, deep conversations, and simple companionship have made every high and low more bearable and enjoyable. To all my friends—both near and far—thank you for being there.

Finally, I would like to express my profound gratitude to my family. To my sister, whose journey in second/foreign language education and research sparked my curiosity and opened the door to my own academic pursuits, thank you always for taking on a big sister role and, more importantly, for being my best friend. And to my parents, who provide endless love and support, I am forever grateful. Your belief in me has been a pillar of strength throughout this journey, and I am incredibly fortunate to have you two as my parents. ありがとう！

Abstract

Although previous research has found that perception and production training improve L2 speech perception (e.g., Bradlow et al., 1997) and production (e.g., Hattori, 2009), respectively, the interactive effects of these two types of training are still an open question. Furthermore, studies have shown mixed findings on the effects of perception-and-production-mixed training (e.g., Baese-Berk & Samuel, 2016; Lu et al., 2015; Thorin et al., 2018).

In addition to the effects of perception and production training, this thesis also considers the role of phonological short-term memory (PSTM) in phonetic acquisition, which has recently been suggested as a predictor of phonetic learning (e.g., Lengeris & Nicolaidis, 2014; MacKay et al., 2001). This thesis investigates the effects of production during perceptual training on the L2 phonetic acquisition of the English /l/–/r/ contrast with L1 Japanese speakers and examines the relationship between PSTM and L2 phonetic learning. Based on the findings, the relationship between perception and production is explored and the effective methods for phonetic acquisition are discussed.

Experiments 1 ($n = 46$) and 2 ($n = 61$) examined how oral repetition during perceptual training affects perceptual learning using discrimination and identification tasks, respectively. The results revealed that perception-only training led to the most significant improvements in perceptual learning. Disruption occurred especially when participants repeated the stimulus after answering a perceptual question, but not when repeating immediately after hearing it.

Experiment 3 ($n = 60$) investigated whether and how the formation of phonetic categories through exposure to target phonemes facilitates phonetic learning. The results showed that training involving production practice improved production accuracy more than perception-only training, especially when training began with perception-only training and shifted to production-only training. These findings suggest that while production practice is important for pronunciation learning, an initial focus on perception followed by production practice leads to better learning outcomes.

The PSTM results revealed no significant relationship between PSTM capacity and either perception or production accuracy before training. Regarding the degree of learning after phonetic training, the findings suggest that the relationship between PSTM capacity and perceptual learning may vary based on initial perceptual ability, with greater PSTM potentially having a positive influence as learning progresses.

Table of Contents

Declaration	i
Acknowledgements	ii
Abstract.....	iv
Table of Contents	v
List of Tables	viii
List of Figures.....	x
Chapter 1 Introduction.....	1
Chapter 2 Review of the Literature	8
2.1 Speech acquisition of the first language	8
2.2 Speech acquisition of the second language.....	11
2.2.1 Speech Learning Model.....	12
2.2.2 Native Language Magnet Model	13
2.2.3 Perceptual Assimilation Model.....	14
2.2.4 Second Language Linguistic Perception Model	15
2.3 Perceptual training studies	15
2.3.1 Factors affecting the effects of perceptual training	16
2.3.1.1 Feedback in perceptual training	16
2.3.1.2 High Variability Phonetic/Perceptual Training.....	18
2.3.1.3 Discrimination and identification training	22
2.3.2 Effects of perceptual training on production	25
2.4 Production training studies	30
2.4.1 Effects of production training on production.....	30
2.4.2 Effects of production training on perception	43
2.5 Perception and production-mixed training.....	46
2.5.1 Production effect on vocabulary learning.....	47
2.5.2 Effects on perceptual learning	54
2.5.3 Effects on production learning.....	64
2.6 Phonological Short-Term Memory (PSTM).....	66
2.6.1 Relationship between PSTM and perception.....	68
2.6.2 Relationship between PSTM and production	73
2.7 English /l/ and /r/ sounds for L1 Japanese speakers	75
2.8 Current study.....	77
Chapter 3 Experiment 1: The effects of production during discrimination training and the relationship between L2 perception and PSTM	79
3.1 Introduction.....	79
3.2 Method	80
3.2.1 Participants	80
3.2.1.1 Japanese learners of English.....	80
3.2.1.2 L1 English speakers for stimulus recording	81
3.2.2 Materials	81
3.2.2.1 English stimuli for training sessions and tests.....	81
3.2.2.2 Stimuli for the PSTM tasks	82
3.2.2.3 Language background questionnaire	82
3.2.3 Procedure	83
3.2.3.1 Phonetic training.....	84

3.2.3.2	PSTM task	85
3.2.3.3	Discrimination task	85
3.2.3.4	Identification task	85
3.2.4	Statistical analysis	86
3.3	Results	87
3.3.1	Discrimination task	87
3.3.1.1	Descriptive statistics	87
3.3.1.2	Inferential Statistics	93
3.3.2	Identification task	97
3.3.2.1	Descriptive statistics	97
3.3.2.2	Inferential statistics	103
3.3.3	Summary of the results on perceptual learning	106
3.3.4	Relationship between PSTM and perceptual learning	107
3.3.4.1	PSTM task scores	107
3.3.4.2	PSTM and discrimination task score	108
3.3.4.3	PSTM and identification task score	110
3.4	Discussion	113
3.5	Summary of Experiment 1	121
Chapter 4	Experiment 2: The effects of production during identification training and the relationship between L2 perception and PSTM	122
4.1	Introduction	122
4.2	Method	123
4.2.1	Participants	123
4.2.2	Materials	124
4.2.2.1	English stimuli for training sessions and tests	124
4.2.2.2	Stimuli for the PSTM tasks	124
4.2.2.3	Language background questionnaire	124
4.2.3	Procedure	124
4.2.3.1	Phonetic training	125
4.2.3.2	PSTM task	126
4.2.3.3	Discrimination task	126
4.2.3.4	Identification task	126
4.2.4	Statistical analysis	126
4.3	Results	126
4.3.1	Discrimination task	126
4.3.1.1	Descriptive statistics	126
4.3.1.2	Inferential statistics	132
4.3.2	Identification task	137
4.3.2.1	Descriptive statistics	137
4.3.2.2	Inferential statistics	143
4.3.3	Summary of results on perceptual learning	144
4.3.4	Relationship between PSTM and perceptual learning	145
4.3.4.1	PSTM task scores	145
4.3.4.2	PSTM and discrimination task score	146
4.3.4.3	PSTM and identification task score	150
4.3.5	Summary of results on PSTM	152
4.4	Discussion	153
4.5	Summary of Experiment 2	157
Chapter 5	Experiment 3: The effects of perception and production training on phonetic learning and its relationship with PSTM	158
5.1	Introduction	158
5.2	Method	159
5.2.1	Participants	159
5.2.1.1	Japanese learners of English	159
5.2.1.2	L1 English speakers for an L1 rating task	160
5.2.2	Materials	160
5.2.2.1	English stimuli for training sessions and tests	160
5.2.2.2	Stimuli for phonological short-term memory tasks	161

5.2.2.3	Language background questionnaire	161
5.2.3	Procedure	161
5.2.3.1	Phonetic training.....	162
5.2.3.2	PSTM task	163
5.2.3.3	Picture description task	163
5.2.3.4	Sentence reading task	163
5.2.3.5	Word-list reading task	163
5.2.3.6	Discrimination task.....	163
5.2.3.7	Identification task	164
5.2.3.8	Exit questionnaire.....	164
5.2.4	Rating tasks.....	164
5.2.5	Statistical analysis.....	165
5.3	Results.....	166
5.3.1	Discrimination test.....	166
5.3.1.1	Descriptive Statistics	166
5.3.1.2	Inferential Statistics	171
5.3.2	Identification task	175
5.3.2.1	Descriptive Statistics	175
5.3.2.2	Inferential Statistics	180
5.3.3	Summary of results on perceptual learning	182
5.3.4	Identification rating task	183
5.3.4.1	Descriptive statistics.....	183
5.3.4.2	Inferential statistics.....	188
5.3.5	Goodness rating task.....	192
5.3.5.1	Descriptive statistics.....	192
5.3.5.2	Inferential statistics.....	197
5.3.6	Summary of results on production learning.....	200
5.3.7	Relationship between PSTM and phonetic learning.....	200
5.3.7.1	PSTM task score.....	201
5.3.7.2	PSTM and discrimination scores.....	202
5.3.7.3	PSTM and identification scores	203
5.3.7.4	PSTM and identification rating scores	207
5.3.7.5	PSTM and goodness rating scores.....	210
5.3.8	Summary of results on PSTM.....	211
5.3.9	Learners' Perceptions of Phonetic Training	212
5.3.9.1	PERC-ONLY training	214
5.3.9.2	MIX training.....	217
5.3.9.3	PROD-ONLY training	220
5.3.9.4	PERC-PROD training.....	222
5.4	Discussion.....	228
5.5	Summary of Experiment 3	239
Chapter 6	General Discussion	241
6.1	Production effects in perceptual training.....	241
6.2	Relationship between speech perception and production learning.....	243
6.3	Relationship between PSTM and phonetic learning.....	246
Chapter 7	Conclusion	249
7.1	Pedagogical implications	249
7.2	Limitations and further research	250
References		252
Appendices		270

List of Tables

Table 3.1 Experiment 1: Stimulus Types of Different Word and Speaker Familiarity	82
Table 3.2 Experiment 1: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-Tests of the Discrimination Task.....	88
Table 3.3 Experiment 1: Model Output of a Logistic Mixed Effects Model for the Discrimination Task	94
Table 3.4 Experiment 1: Summary of Pairwise Comparisons between Time and Group for the Discrimination Task	95
Table 3.5 Experiment 1: Summary of Pairwise Comparisons between Time and Stimuli for the Discrimination Task	96
Table 3.6 Experiment 1: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-Tests of the Identification Task	98
Table 3.7 Experiment 1: Model Output of a Logistic Mixed Effects Model for the Identification Task	104
Table 3.8 Experiment 1: Summary of Pairwise Comparisons between Time and Stimuli for the Identification Task.....	106
Table 3.9 Experiment 1: Mean Scores, Standard Deviations, and Minimum and Maximum Scores of the PSTM Task.....	108
Table 4.1 Experiment 2: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-Tests of the Discrimination Task.....	127
Table 4.2 Experiment 2: Model Output of a Logistic Mixed Effects Model for the Discrimination Task	134
Table 4.3 Experiment 2: Summary of Pairwise Comparisons between Group and Stimuli for the Discrimination Task	135
Table 4.4 Experiment 2: Summary of Pairwise Comparisons between Time and Stimuli for the Discrimination Task	136
Table 4.5 Experiment 2: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-Tests of the Identification Task	137
Table 4.6 Experiment 2: Model Output of a Logistic Mixed Effects Model for the Discrimination Task	143
Table 4.7 Experiment 2: Summary of Pairwise Comparisons between Time and Group for the Discrimination Task	144
Table 4.8 Experiment 2: Mean Scores, Standard Deviations, and Minimum and Maximum Scores of the PSTM Task.....	145
Table 4.9 Experiment 2: Model Output of a Linear Regression Model with the PSTM score for the Discrimination Task	148

Table 4.10 Experiment 2: Model Output of a Linear Regression Model with the PSTM score for the Identification Task.....	151
Table 5.1 Experiment 3: Stimulus Types of Different Word and Speaker Familiarity	161
Table 5.2 Experiment 3: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-tests of the Discrimination Task	166
Table 5.3 Experiment 3: Summary of Pairwise Comparisons across Groups and Stimuli at Pre-test for the Discrimination Task	173
Table 5.4 Experiment 3: Summary of Pre-test and Post-test Pairwise Comparisons across Groups and Stimuli for the Discrimination Task	174
Table 5.5 Experiment 3: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-tests of the Identification Task	175
Table 5.6 Experiment 3: Summary of Pre-test and Post-test Pairwise Comparisons across Groups and Stimuli for the Identification Task	182
Table 5.7 Experiment 3: Summary of Pairwise Comparison Results (<i>p</i> values) for Perceptual Learning	183
Table 5.8 Experiment 3: Mean Identification Rating Scores.....	184
Table 5.9 Experiment 3: Summary of Pairwise Comparisons across Groups and Tasks at Pre-test for the Identification Rating Score	190
Table 5.10 Experiment 3: Summary of Pre-test and Post-test Pairwise Comparisons across Groups and Speaking Tasks for the Identification Rating Score	191
Table 5.11 Experiment 3: Summary of Pre-test and Post-test Pairwise Comparisons across Groups and Tasks for the Identification Rating Score	191
Table 5.12 Experiment 3: Mean Goodness Rating Scores and the Standard Deviations	193
Table 5.13 Experiment 3: Summary of Pairwise Comparisons of Tasks within Groups at Pre-test for the Goodness Rating Score.....	199
Table 5.14 Experiment 3: Summary of Pre-test and Post-test Pairwise Comparisons across Groups and Speaking Tasks for the Goodness Rating Score.....	199
Table 5.15 Experiment 3: Summary of Pairwise Comparison Results (<i>p</i> values) for Production Learning.....	200
Table 5.16 Experiment 3: Mean Scores, Standard Deviations, and Minimum and Maximum Scores of the PSTM Task	201
Table 5.17 Experiment 3: Model Output of a Linear Regression Model with the PSTM score for the Identification Task.....	205
Table 5.18 Experiment 3: Model Output of a Linear Regression Model with the PSTM score for the Identification Rating Score with All Tasks	209
Table 5.19 Experiment 3: Summary of the results of the Exit Questionnaire	213

List of Figures

Figure 3.1 Experiment 1: Timeline of Data Collection	84
Figure 3.2 Experiment 1: Discrimination Task Results with All Stimuli	89
Figure 3.3 Experiment 1: Discrimination Task Results with the TWTS Stimuli.....	90
Figure 3.4 Experiment 1: Discrimination Task with the TWUS Stimuli	91
Figure 3.5 Experiment 1: Discrimination Task Results with the UWTS Stimuli	92
Figure 3.6 Experiment 1: Discrimination Task Results with the UWUS Stimuli.....	93
Figure 3.7 Experiment 1: Identification Task Results with All Stimuli.....	99
Figure 3.8 Experiment 1: Identification Task Results with the TWTS Stimuli	100
Figure 3.9 Experiment 1: Identification Task Results with the TWUS Stimuli.....	101
Figure 3.10 Experiment 1: Identification Task Results with the UWTS Stimuli.....	102
Figure 3.11 Experiment 1: Identification Task Results with the UWUS Stimuli	103
Figure 3.12 Experiment 1: PSTM Score per Training Group	108
Figure 3.13 Experiment 1: Correlation between the PSTM Score and the Discrimination Task Pre- Test Score	109
Figure 3.14 Experiment 1: Correlation between the PSTM Score and the Gain Score of the Discrimination Task	110
Figure 3.15 Experiment 1: Correlation between the PSTM Score and the Identification Task Pre-Test Score	111
Figure 3.16 Experiment 1: Correlation between the PSTM Score and the Gain Score of the Identification Task.....	112
Figure 4.1 Experiment 2: Timeline of Data Collection	125
Figure 4.2 Experiment 2: Discrimination Task Results with All Stimuli	128
Figure 4.3 Experiment 2: Discrimination Task Results with the TWTS Stimuli.....	129
Figure 4.4 Experiment 2: Discrimination Task Results with the TWUS Stimuli	130
Figure 4.5 Experiment 2: Discrimination Task Results with UWTS Stimuli	131
Figure 4.6 Experiment 2: Discrimination Task Results with the UWUS Stimuli.....	132
Figure 4.7 Experiment 2: Identification Task Results with All Stimuli.....	138
Figure 4.8 Experiment 2: Identification Task Results with the TWTS stimuli.....	139

Figure 4.9 Experiment 2: Identification Task Results with the TWUS Stimuli.....	140
Figure 4.10 Experiment 2: Identification Task Results with the UWTS Stimuli.....	141
Figure 4.11 Experiment 2: Identification Task Results with the UWUS Stimuli	142
Figure 4.12 Experiment 2: PSTM Scores per Training Group.....	146
Figure 4.13 Experiment 2: Correlation between the PSTM Score and the Discrimination Task Pre-Test Score	147
Figure 4.14 Experiment 2: Interaction between the PSTM and Pre-test Scores on the Gain Score in the Discrimination Task across Groups	149
Figure 4.15 Experiment 2: Correlation between the PSTM Score and the Identification Task Pre-Test Score	150
Figure 4.16 Experiment 2: Interaction between the PSTM and Gain Scores in the Identification Task across Groups	152
Figure 5.1 Experiment 3: Timeline of Data Collection	162
Figure 5.2 Experiment 3: Discrimination Task Results with All Stimuli	167
Figure 5.3 Experiment 3: Discrimination Task Results with the TWTS Stimuli.....	168
Figure 5.4 Experiment 3: Discrimination Task Results with the TWUS Stimuli	169
Figure 5.5 Experiment 3: Discrimination Task Results with the UWTS Stimuli	170
Figure 5.6 Experiment 3: Discrimination Task Results with the UWUS Stimuli.....	171
Figure 5.7 Experiment 3: Identification Task Results with All Stimuli.....	176
Figure 5.8 Experiment 3: Identification Task Results with the TWTS stimuli	177
Figure 5.9 Experiment 3: Identification Task Results with the TWUS Stimuli.....	178
Figure 5.10 Experiment 3: Identification Task Results with the UWTS Stimuli.....	179
Figure 5.11 Experiment 3: Identification Task Results with the UWUS Stimuli	180
Figure 5.12 Experiment 3: Identification Rating Task for All Tasks	185
Figure 5.13 Experiment 3: Identification Rating Task for the Word-List Reading Task.....	186
Figure 5.14 Experiment 3: Identification Rating Task for the Sentence Reading Task.....	187
Figure 5.15 Experiment 3: Identification Rating Task for the Picture Description Task.....	188
Figure 5.16 Experiment 3: Identification Rating Task for All Tasks	194
Figure 5.17 Experiment 3: Goodness Rating Task for the Word-List Reading Task	195
Figure 5.18 Experiment 3: Goodness Rating Task for the Sentence Reading Task.....	196

Figure 5.19 Experiment 3: Goodness Rating Task for the Picture Description Task	197
Figure 5.20 Experiment 3: PSTM Scores per Training Group.....	201
Figure 5.21 Experiment 3: Correlation between the PSTM Score and the Discrimination Task Pre-Test Score	202
Figure 5.22 Experiment 3: Correlation between the PSTM Score and the Gain Score of the Discrimination Task	203
Figure 5.23 Experiment 3: Correlation between the PSTM Score and the Pre-Test Score of the Identification Task.....	204
Figure 5.24 Experiment 3: Interaction between the PSTM and Pre-test Scores on the Gain Score in the Identification Task across Groups	206
Figure 5.25 Experiment 3: Correlation between the PSTM Score and the Pre-Test Score of the Identification Rating Scores on Production.....	208
Figure 5.26 Experiment 3: Correlation between the PSTM Score and Gain Score of the Identification Rating Scores on Production	209
Figure 5.27 Experiment 3: Correlation between the PSTM Score and the Pre-test Score of the Goodness Rating Scores on Production	210
Figure 5.28 Experiment 3: Correlation between the PSTM Score and the Gain Score of the Goodness Rating Scores on Production	211
Figure 5.29 Experiment 3: Participants' Desires to Improve Their English Skills and Pronunciation	214
Figure 5.30 Experiment 3: Participants' Perceptions towards PERC-ONLY Training	214
Figure 5.31 Experiment 3: Participants' Perceptions towards MIX Training.....	217
Figure 5.32 Experiment 3: Participants' Perceptions towards PROD-ONLY Training.....	220
Figure 5.33 Experiment 3: Participants' Perceptions towards PERC-PROD Training.....	223

Chapter 1 Introduction

Today, many people communicate in their second language (L2) on a daily basis due to globalisation. However, even after migrating to an English-speaking country, many L2 speakers continue to have heavy foreign accents (Derwing et al., 2008). Research has suggested that speech with strong accents leads to underestimation of the speakers' abilities and credibility. Tsalikis et al. (1991) found that Greek-accented English sales talks were perceived as less credible, competent, and friendly than those with a standard American English accent. Likewise, Lev-Ari and Keysar (2010) showed that accented speech negatively impacted the credibility of information. They found that listeners tended to rate trivia statements as less truthful when these are spoken with a heavy accent as compared to no foreign accent or mild accent, even when listeners were aware that accented speech could impact their judgment. These studies have suggested that attaining comprehensible and intelligible pronunciation is important for L2 speakers to be perceived as competent and credible as they are. Moreover, learning proper pronunciation plays a key role in overall language acquisition, influencing not just how speakers are perceived but also how they internalise and produce language. Research on language acquisition has highlighted the important relationship between pronunciation abilities and language development (Ferguson & Farwell, 1975; Kaushanskaya & Yoo, 2011; Schwartz & Leonard, 1982; Zamuner & Thiessen, 2018).

For example, research on first language (L1) acquisition has identified a relationship between children's production abilities and their lexical development (Ferguson & Farwell, 1975; Schwartz & Leonard, 1982; Zamuner & Thiessen, 2018). Zamuner and Thiessen (2018) found that children tend to imitate new words when learning their L1, and that the likelihood of spontaneous imitation is influenced by the familiarity of the consonants in the word. Specifically, their study revealed that children were more likely to imitate words composed of consonants that they had previously produced. The researchers argued that this phenomenon can be explained by the two accounts. The first is Ferguson and Farwell's (1975) lexical avoidance effect, which suggests that children are reluctant to produce words with unfamiliar sounds. The second account is that children are simply unable to produce words containing sounds that they are not yet capable of articulating (Schwartz & Leonard, 1982).

Furthermore, familiarity with sounds also plays a significant role in L1 lexical development. Schwartz and Leonard (1982) demonstrated that children learned new words

more effectively when those words contained familiar sounds, as opposed to unfamiliar ones. This finding aligns with research on adult learners. Kaushanskaya and Yoo (2011) discovered that adult learners could memorise more words when these did not contain foreign sounds, highlighting the crucial role of familiarity with sounds in word retention. This suggests that both children and adults rely on their phonological knowledge to learn language, emphasising the importance of mastering L2 pronunciation for successful language acquisition.

Although research has shown the importance of L2 phonetic acquisition (Levis, 2005; Munro & Derwing, 2006a), many language teachers neglect pronunciation instructions, focusing instead on other skills such as reading, grammar, and vocabulary (Burri & Baker, 2021; Foote et al., 2013). Macdonald (2002) conducted in-depth interviews with eight ESL (English as a Second Language) teachers in Australia to investigate their perspectives on pronunciation teaching and whether they taught pronunciation in their classes. The study found that teachers did not prioritise teaching and assessing pronunciation because the curriculum did not require pronunciation teaching; instead, it focused on other skills. Similarly, in the UK, Burgess and Spencer (2000) surveyed English teachers using a questionnaire and reported that most teachers integrated pronunciation teaching into other class activities, which means they did not dedicate specific time to it. In the United States, Darcy et al. (2012) conducted a questionnaire survey with ESL teachers and found that 71% reported no explicit pronunciation instruction in their classes. In Canada, 40 hours of classroom observation in three ESL teachers' classrooms revealed that while pronunciation teaching occurred, much less time was allocated to it compared to grammar and vocabulary teaching (Foote et al., 2016). Likewise, a study by Henderson et al. (2012) exploring pronunciation teaching practices in seven European countries through a questionnaire survey, found that while the majority of teachers reported spending up to a quarter of their class time on pronunciation, some teachers believed even less time should be spent on pronunciation instruction.

Regarding the features of pronunciation that are taught, a Likert-scale questionnaire survey conducted by Burns (2006) showed that teachers focused more on segmental features than on suprasegmental features. Teacher training in pronunciation instruction also appears to be lacking. Henderson et al. (2012) found that most teachers had little or no training in pronunciation instruction, which affects their confidence in teaching it. This lack of confidence is problematic, as demonstrated by Bai and Yuan's (2019) qualitative study in

Hong Kong, which included teachers' written reflections and focused group interviews. This study revealed that teachers often avoided pronunciation instruction due to a lack of confidence.

In the context of English education in Japan, Teshima (2011) reported a lack of pronunciation teaching in English classrooms. Based on classroom observations and interviews with teachers and students, Teshima found that very little pronunciation instruction was conducted in junior high (ages 13–15) and high schools (ages 16–18). More specifically, almost no instruction on segmental features, such as vowels and consonants, was provided. However, at the suprasegmental level, word stress and sentence intonation were taught in classrooms because these aspects frequently appear in Japanese entrance exams.

An empirical study conducted by Otsuka and Ueda (2012) supported Teshima's findings. They carried out a pronunciation check sheet quiz and a questionnaire survey with first-year university students and a separate questionnaire for English teachers in junior (years 7 to 9) and senior high (years 10 to 12) schools. The results indicated that the majority of first-year students (about 70%) felt incompetent in reading phonetic symbols. On the contrary, many of these students correctly answered questions on word stress on the quiz. This was consistent with the teachers' responses to the questionnaire that over 80% of junior high school teachers and all high school teachers reported teaching word stress. The accuracy rate of intonation varied depending on the sentence type; for example, most students answered correctly to yes/no questions, whilst the accuracy rates in alternative questions and enumerations were less than 50%.

Some researchers have argued that teachers in Japan avoid teaching pronunciation because they had not been trained to teach it during their teacher education programmes. Kochiyama et al. (2013) conducted an extensive review of syllabi from various teacher education programs focused on English teaching in Japan. In their analysis, they examined the syllabi of courses offered within these programs to assess the inclusion of pronunciation instruction. Through the examination, they found that three-fifths of the programmes did not offer courses covering pronunciation instruction. Even in programmes that included phonetics courses, the time spent on how to teach pronunciation was limited. The authors argued that this lack of opportunities to study how to teach pronunciation forms a vicious circle (i.e., students in teacher education programmes not studying how to teach pronunciation → a lack of confidence in teaching pronunciation → not teaching pronunciation in schools → students

whose English is heavily accented entering a teacher education programme → ...). Similarly, Uchida and Sugimoto (2018) conducted a questionnaire survey in 2015 to explore which pronunciation features were taught in schools and how they were taught. A hundred English teachers in public junior high schools in Tokyo completed the survey which showed that while both segmental and suprasegmental features were taught, instruction primarily relied on listening and repeating. More complex suprasegmental features were often avoided, likely because teaching them requires knowledge and skills.

Another reason why Japanese teachers of English tend to avoid pronunciation instruction may come from the belief that their own pronunciation is not as accurate as that of L1 speakers (Inoi, 2010). However, as Levis et al. (2016) pointed out, the important thing is not whether teachers are L1 speakers, but whether they possess the knowledge to teach pronunciation effectively. Teacher training courses can cultivate teachers' self-confidence in pronunciation. For example, Inoi (2017) found that many elementary school teachers who attended a pronunciation training course gained confidence in their pronunciation through the training. This highlights the essential role of teacher training in enabling teachers to conduct pronunciation instruction with more confidence.

Another possible reason why pronunciation is not widely taught in Japan is due to the washback effects from the entrance exam system. Although there are various methods to enter schools and universities, written exams are the most common, and English is one of the subjects that is the most unavoidable. Some schools and universities have introduced English speaking tests in their entrance exams, but due to the administrative difficulties, such as time constraints and the burden on the interviewers, only a small number of schools and universities conduct English speaking tests.

In terms of pronunciation, some English entrance exams include questions that assess pronunciation knowledge. For example, the National Center Test for University Admissions, which used to be a standardised test administered as part of the university entrance process, included questions on stress positions and syllable counting. However, these questions were usually limited to the word level and were designed to test knowledge rather than actual pronunciation ability. The test was replaced by the Common Test for University Admissions in 2021, which no longer includes pronunciation questions. The absence of speaking tests in entrance exams drives teachers to focus more on other areas, especially reading, grammar, and vocabulary, because these skills are more likely to be tested.

Despite the importance of pronunciation, it has become clear that it is not sufficiently taught in many language classes. However, even with dedicated effort, L2 pronunciation remains difficult to master for many learners, especially adults. Lenneberg (1967) hypothesised that after a certain age—the *critical* period—children lose their ability to learn a new language in the same way they learn their L1. This critical period has been hypothesised to end around age 12 (Lenneberg, 1967). While the Critical Period Hypothesis originally focussed on language acquisition ending after puberty, more recent research suggests that the ability to achieve nativelike pronunciation decreases after a sensitive window, even if language learning continues throughout life (Hakuta et al., 2003). For instance, Flege (1999a) argued that learners who passed puberty often develop a noticeable foreign accent because their well-established L1 phonetic categories interfere with L2 pronunciation. Thus, while learners can still acquire an L2 after the critical period (Flege et al., 1995), the likelihood of achieving a nativelike pronunciation becomes considerably lower as age increases, supporting the view that the sensitive period plays a role in L2 phonetic development (Ortega, 2009). Indeed, many L2 learners struggle to attain intelligible pronunciation. For example, a longitudinal study showed that some L2 speakers' speech comprehensibility does not improve much even after ten years of living in an English-speaking country (Munro et al., 2024). These findings suggest that, despite the lifelong ability to learn L2 sounds, L2 learners might need explicit training in speech both perception and production, especially for challenging sounds.

One major issue is that although many researchers have investigated the effects of L2 phonetic training, rigorous methods for acquiring accurate speech perception and production have not been fully established yet. For example, it is well-known that the English /l/–/r/ contrast is difficult for L1 Japanese speakers. Although many studies have investigated this specific case using intensive perceptual training (e.g., Bradlow et al., 1997) and production training (e.g., Hattori, 2009) and showed significant improvement in each modality, these training designs have not been sufficiently effective in developing robust phonetic categories.

To develop effective speech learning training methods, it is crucial to understand the relationship between speech perception and production. Research has shown that integrating both skills can enhance learning outcomes. For example, research on L1 vocabulary learning has shown that reading or repeating nonwords aloud enhanced memorisation of new words (MacLeod et al., 2010), suggesting that adding a production element to perceptual training could improve learning outcomes. This finding aligns with the broader idea that production

supports processing new linguistic items. However, the findings of L2 phonetic acquisition studies have shown contradicted findings. Studies have found that repeating stimuli during perceptual training did not yield additional learning benefits (Lu et al., 2015; Thorin et al., 2018) or even disrupted perceptual learning (Baese-Berk, 2010, 2019; Baese-Berk & Samuel, 2016). Possible reasons for this disruption are listening to one's own non-target-like pronunciation, the cognitive cost of switching between two tasks (i.e., perception and production), and the absence of feedback on production. The exact source of this disruption or the reason for the null-effect is still uncertain.

This thesis aims to investigate the interactive effects of perception and production training. Understanding how these two modalities are intertwined is key to refining our understanding of L2 phonetic acquisition. By examining the learning processes involved in perception and production, this research contributes to the theoretical framework that underpins pronunciation training. Moreover, the insights gained from this investigation will provide pedagogical implications for developing more effective learning strategies for L2 learners.

In addition to exploring the relationship between the two speech modalities, this thesis also considers individual differences in speech acquisition. One factor that has gained attention in recent studies is phonological short-term memory (PSTM) (Ghaffarvand Mokari & Werner, 2019; Inceoglu, 2019; Moorman, 2017). This thesis sheds lights on the nature of L2 speech perception and production, and the relationship between L2 speech learning and PSTM by comparing the effects of various training types on the English /l/-/r/ contrast on perceptual and production learning with Japanese learners of English. Based on these findings, effective learning methods and strategies for both modalities are discussed.

This dissertation is organized as follows. Chapter 2 reviews the relevant literature on the acquisition of speech perception and production. It first examines theories of L1 and L2 speech acquisition, followed by studies on training in perception, production, and combination of the two modalities. Then, the relationship between speech acquisition and PSTM is discussed. Furthermore, the challenges that Japanese learners face in acquiring the English /l/ and /r/ sounds are explored. Finally, the purposes and structure of this research and each experiment are presented.

Chapter 3 and Chapter 4 present the methods, results, and discussions for Experiments 1 and 2, respectively. Experiment 1 focused on the effect of production during discrimination

training on perceptual learning, while Experiment 2 explored the production effect using identification training. Chapter 5 follows a similar structure but covers Experiment 3, which further investigated the relationship between perceptual learning and production learning by comparing the perception-only, production-only, perception-production-mixed, and perception-to-production conditions. Chapter 6 discusses the interactive effects of perception and production training and the relationship between phonetic learning and PSTM based on the findings from the three experiments. In Chapter 7, the key findings are summarised, and pedagogical implications and future research directions are suggested.

Chapter 2 Review of the Literature

In this chapter, the literature on speech acquisition of the first language is reviewed to understand the general ideas of how humans acquire speech. Then, the scope of the review is narrowed down to second language speech acquisition by examining four well-known speech models. Next, phonetic training studies of perceptual, production, and the combination of the two modalities, which are the main focuses of this study, are reviewed to understand their effectiveness and limitations, as well as their gap in the literature. Then, how speech acquisition relates to PSTM is demonstrated, and the reasons why English /l/–/r/ is a challenging contrast for L1 Japanese speakers. Lastly, the purpose of this study is presented.

2.1 Speech acquisition of the first language

This thesis addresses the acquisition of the L2 speech and its potential application to L2 classrooms. L2 acquisition theories have been developed in relation to L1 acquisition. Due to the close nature between L1 and L2 acquisition, it is important to first grasp a bigger picture of how humans acquire speech by reviewing the literature on speech acquisition of the L1 for a better understanding of L2 speech acquisition.

Speech acquisition starts in the womb, even before a baby is born. Although it is still uncertain when exactly a foetus starts hearing the external speech in utero, a number of studies have shown that they can listen and respond to ambient sounds. A study conducted by Hepper and Shahidullah (1994) showed that 96% of foetuses start responding to low-frequent external sounds (i.e., 500-Hz, 250-Hz and 100-Hz tones) from the beginning of week 19 to 27. As for high-frequent sounds, their responses to 1000-Hz and 3000-Hz tones start to be observed from week 29 to 33 and from week 31 to 35, respectively. Griffiths et al. (1994) studied a pregnant sheep, which has similar transmissibility of sound in the womb to that of humans, to understand whether the external speech can be heard in utero; they also tried to test the intelligibility of the external sounds by implanting a miniature hydrophone into their uterus. The results showed that speech was relatively intelligible, especially when a given sound was voiced, although the place or manner of articulation was less determined. Also, a series of studies showed that foetuses can learn the vowels by listening to external sounds (Moon et al., 2013), while consonant sounds are unclear in the womb because a foetus is surrounded by the amniotic fluid in the womb and the voice they hear is low-pass filtered. Foetuses can also hear and learn the suprasegmental features of speech, such as pitch

differences. Research conducted by Mampe et al. (2009) showed that newly born babies (aged 2–5 days) cried with the pitch patterns of their mothers.

After birth, infants show an excellent discriminative ability for phonetic contrasts. A series of studies have found that infants can discriminate both phonemic contrasts, which are relevant to their mother tongue, and irrelevant contrasts. Aslin et al. (1981) reported that infants aged 6 to 12 months and surrounded by English can discriminate irrelevant voice onset time contrasts used in Thai and Spanish languages. Similar findings have been reported for the phonetic contrasts of Hindi and Thompson, one of the Salishan languages spoken in British Columbia, Canada (Werker & Tees, 1984), and French and Czech (Trehub, 1976). Japanese infants aged 6–8 months can also discriminate the English /l/–/r/ contrast, which is notoriously difficult for Japanese adult learners of English (Kuhl et al., 2006; Tsushima et al., 1994). Regarding vowels, four-month-old infants in an English-speaking environment can discriminate German vowel contrasts (Polka & Werker, 1994).

Infants, however, start learning which phonemes are relevant and irrelevant in their surrounding language(s) to develop the ability to identify which perceived sound should be categorised as which phoneme, a process known as categorical perception (Kuhl, 1994, 2004; Liberman et al., 1967; Liberman et al., 1957; Werker & Tees, 1984). As they learn categorical perception, their discriminative ability comes to be attuned to relevant phonetic contrasts by being exposed to the surrounding language(s) and learning the relevant phonemes (Cutler, 2012). Infants develop their phonetic system for categorical perception, which is compatible with the surrounding language(s), by the end of the first year of their life. Research shows that English-learning infants at 11 to 13 months cannot discriminate the Hindi contrasts, although they can do so when aged six to eight months (Werker & Lalonde, 1988). It has also been reported that infants who can discriminate foreign contrasts lose their discriminative ability at the age of 10 months (Werker & Tees, 1984). Similarly, Japanese infants lose their discriminative ability for the English /l/–/r/ contrast at 10–12 months (Kuhl et al., 2006; Tsushima et al., 1994). Regarding vowels, six-month-old infants show a lower score than four-month-old infants, suggesting that the shift of language specificity begins earlier for vowels than consonants (Polka & Werker, 1994). Based on these findings, language specificity for a phonetic system by infants is likely to appear after four months for vowels and eight months for consonants.

Infants go through these processes and develop their phonetic system based on speech input by family and caregivers. On the whole, babies spend much time listening to their

family and/or caregivers. A study showed that a baby in the Netherlands heard two and a half hours of speech daily on average at six to nine months old (van de Weijer, 1999). When speaking to babies, people speak slowly with longer and more frequent pauses. This infant-directed speech is said to assist and lead the acquisition of an L1 by infants. Liu et al. (2003) found that infants who have their mother speak clearly to them (i.e., an expanded vowel space) performed better in a phonetic discrimination task. The impact of infant-directed speech can also be observed in babies' babblings. A study by Werker et al. (2007) showed that infants reflected emphasised-speech input in their babblings. In their study, English- and Japanese-speaking mothers taught their babies minimal pairs (e.g., 'Gidda' ギダ and 'Geeda' ギーダ). The results showed that two groups of mothers produced the minimal pairs differently; the English-speaking mothers emphasised the spectral differences between the vowels, whilst the Japanese-speaking mothers emphasised the durational differences. Their babies' babblings also changed accordingly. These findings are consistent with Goldstein and Schwade's (2008) claim that infants' production reflects the speech input they receive.

Speech perception and production are closely linked because they share auditory and motor information. Kuhl (2000) explained that infants store perceptual representations by being exposed to ambient sounds, and those representations assist in motor speech development. Accordingly, researchers have reported that children become able to perceive phonemes accurately before being able to produce them. A well-known phenomenon that supports this claim is 'fis phenomenon,' documented by Berko and Brown (1960). During an observation conducted by one of the researchers, a boy called his toy 'fis.' Then, the researcher asked him whether the toy was 'fis.' However, he denied it repeatedly and only said 'yes' when the researcher said 'fish.' From this phenomenon, they argued that the boy could identify the phonetic difference between /s/ and /ʃ/ prior to becoming able to produce /ʃ/ as adults do. Thus, it suggests that children develop their phonetic categories faster in their mind than they become able to articulate them.

After the babbling period, infants generally start producing real words by the time they reach age one (Newman, 2017). Regarding the onset of speech production, the existence of the acquisition order of English consonant sounds has been confirmed by research. According to Setter (2019), the acquisition periods vary depending on each sound between ages one and a half to eight. Some phonemes such as /p, n, h, m, w, b/ can be acquired mostly by three-year-olds, whilst /r, dʒ/ are likely to take eight years to acquire. Jakobson (1968, 1971) who has extensively researched child language acquisition, claims that the easier phonemes are, the

faster children acquire them; they start imitating easily articulated sounds. He explained that children generally begin with wide vowels, especially the /a/ sound, with labial consonants. Next, they acquire nasal and dental sounds. Then, they are likely to acquire stop consonants followed by fricatives. Compared to the front consonants, children take more time to acquire back consonants, such as /k/ and /g/. The difficulty in back sounds has been indicated for vowels as well. Studies of newborn babies showed that they barely produce back vowels, such as /u/ and /ɑ/ (Irwin & Curry, 1941). An explanation for the difficulty in producing back vowels is the difference of tongue–jar interaction (Kent, 1981). Kent explained that seven of the front vowels are produced with similar tongue positions. In contrast, producing back vowels require shifting the tongue relative to the jar. In addition, back vowels include rounded vowels such as /u, u, o, o/, which require lip movements. Thus, infants need to learn complex tongue–jar–lip relationships among the phonemes, leading them to acquire those back sounds with more difficulty.

In sum, the humans’ speech learning trajectory starts in the womb, beginning with learning vowels and suprasegmental features, with their phonetic acquisition developing further after their birth. They are born with an ability to distinguish both relevant and irrelevant phonetic contrasts in the surrounding language(s). Their discriminative ability starts to be restricted to their mother tongue after four months for vowels and eight months for consonants when they start learning categorical perception. By the age of 10 months, they can discriminate phonetic contrasts according to the phonetic system of their mother tongue. When they reach age one, they start producing sounds from easy-articulating phonemes; front sounds are generally easier than back sounds. Regarding English consonants, children generally take eight years to articulate all phonemes like adult L1 speakers. The studies on L1 phonetic acquisition reviewed so far suggest that speech perception precedes speech production.

2.2 Speech acquisition of the second language

There is a consensus that L2 speakers have difficulty in perceiving and producing the target language, and the learning process is not as simple as the one with the L1. Previous studies have shown that even after spending many years of learning an L2, many learners continue struggling with acquiring intelligible pronunciation of the target language (Munro et al., 2024). To account for the difficulty, researchers have developed several speech models: speech learning model (Flege, 1995a), native language magnet model (Kuhl, 1994; Kuhl et

al., 2008; Kuhl & Iverson, 1995), perceptual assimilation model (Best, 1995; Best & Tyler, 2007) and second language perception model (Escudero, 2005; van Leussen & Escudero, 2015). These speech models predict the degree of difficulty in the acquisition of speech perception and production for L2 phonetic contrasts in terms of (dis)similarity between learners' L1 and L2 phonetic systems.

2.2.1 Speech Learning Model

One of the most well-known and influential models for L2 phonetic acquisition is the speech learning model (SLM) developed by Flege (1995a). This model mainly focuses on the acquisition difficulties of L2 phonemes based on the perceived differences between those of L1 and L2 by experienced learners. The model predicts the following broad assumptions: L1 and L2 phonetic categories share the same long-term memory space, the ability to form new phonetic categories remains throughout the life, and learners' production reflects their mental representations of L2 phonemes (Flege, 2003).

First, Flege (1995a) explains that mental representation is composed of phonetic categories which possess acoustic and articulatory information. The L1 and L2 phonetic categories share the same phonological space in the long-term memory, which makes them influence each other. L2 speakers perceive speech sounds based on the perceived phonetic differences among L1 and L2 phones as 'identical,' 'new,' or 'similar.' L2 phonemes, which also exist in learners' L1, are processed smoothly since there is no need for category formation. When L2 learners perceive a sound, which is significantly far from each of their L1 category, they recognise it as a 'new' sound. Because category formation is triggered auditorily if they continue listening to the sounds repeatedly, the category of the sound is likely to form. The most problematic sounds are ones which have a 'similar' sound in learners' L1. To establish a phonetic category of these 'similar' sounds, learners need to detect the phonetic differences. Once they detect the differences by listening to the sounds repeatedly, the category will be gradually constructed.

The next important assumption about the SLM is that speech learning ability remains over one's life span. This assumption challenges the Critical Period Hypothesis (Lenneberg, 1967). The hypothesis posits that people biologically lose their ability to acquire a language around age 12. After this critical age, it is impossible to attain the level of L1 speakers in L2. Flege agrees that early learners are better at phonetic category formation of new phonemes than late learners. However, he explains that this is not only due to a biological constraint but

also because of the differences of input. Adult learners receive less input, both quantitatively and qualitatively, than young learners (Flege, 1995a). The quantity and quality of input affect L2 category formation. Thus, the SLM hypothesises that sufficient authentic input plays a role in learning L2 phonemes, and an innate ability for language acquisition remains throughout one's life.

Lastly, the SLM hypothesises about the relationship between speech perception and production skills. The model predicts that perception skill precedes production skill; in other words, being able to perceive an L2 sound does not always mean being able to produce that sound accurately. Since learners' perceptual accuracy reflects their production accuracy, difficulties in L2 perception can lead to L2 production errors (Flege, 2003).

2.2.2 Native Language Magnet Model

The native language magnet (NLM) model predicts difficulties in distinguishing non-native phonetic contrasts by explaining how language specificity develops during L1 speech acquisition (Kuhl, 1994; Kuhl & Iverson, 1995). The model was expanded later to include neural commitment and the effects of L1 exposure on the brain (NLM-e, Kuhl et al., 2008).

This model predicts that infants who are younger than one year old can discriminate any phonemes. During the first year, infants tune to their L1 phonological system by developing prototypes of L1 phonemes (phonetic categories) and creating a sound map in their brain. The sound map helps them in hearing the L1 phonemes more clearly. On the contrary, it obstructs distinguishing non-native phonemes, which fall into the same prototypes due to the 'perceptual magnet' effect. The L1 prototypes function as perceptual magnets, attracting unfamiliar sounds to them when listeners perceive unfamiliar sounds. Because the perceptual magnet effect is greater for sounds that are close to one of the L1 prototypes, L2 phonemes with similar L1 counterparts are more difficult to perceive correctly.

Similar to the SLM, the NLM model also makes a conflicting claim against the CPH (Lenneberg, 1967). The model predicts that speech perception by adults and infants are different because of their previous language input. The difference is not attributed to brain plasticity, as claimed in the CPH. Thus, adult learners can also learn non-native sounds. As for the production ability, similar to the SLM, the NLM model predicts that speech perception precedes production (Kuhl et al., 2008).

2.2.3 Perceptual Assimilation Model

The perceptual assimilation model (PAM), developed by Best (1994, 1995), predicts the perception of non-native phonetic contrasts by naïve listeners. The core of the model is based on a direct realist approach (Best, 1995) that posits that listeners perceive information on articulatory gestures of speech sounds (i.e., organs' movement, articulation place, and articulation manner). The model was later expanded to L2 learning, PAM-L2 (Best & Tyler, 2007), using the same principles as the original PAM. The models predict the perception and acquisition difficulties depending on the gestural similarity or difference of the phonetic contrasts.

According to the models, six assimilation patterns exist whereby non-native phonetic contrasts can be assimilated into native sounds:

- Two-category (TC) assimilation: Listeners listen to non-native phonetic contrasts, which have similar individual phonemes in L1. Perception has no problem for these contrasts.
- Single-category (SC) assimilation: Both phonemes of a non-native contrast can be perceived as a single phoneme in L1. This assimilation pattern is arduous for L2 listeners to both perceive and learn the sounds. The English /l/–/r/ for L1 Japanese listeners are categorised into this pattern.
- Category-goodness (CG) assimilation: Two non-native phonemes can be assimilated into a phoneme in L1 just like SC. However, one of the non-native phonemes is more similar to the native phoneme than the other. Due to the difference in the similarity, listeners can perceive the difference between the phonemes and relatively learn the sounds easily.
- Uncategorised-uncategorised (UU) contrasts: Neither of the non-native phonemes falls into any L1 categories. Depending on the similarity between L1 and non-native phonemes, perception can be either good or poor.
- Uncategorised-categorised (UC) contrasts: Only one of the non-native phonetic contrasts is assimilated into an L1 phoneme. Perception of the contrasts is good.
- Non-assimilable (NA): Non-native phonemes cannot be perceived as speech sounds, such as click sounds for people whose L1 does not have clicks. Despite the absence of the acoustic features in the L1 phonological system, perception of the non-native sounds is excellent.

The PAM-L2 posits that it is possible to learn and improve L2 perception by exposure and experience of target sounds. When learning L2 phonemes, L1 phonetic categories (pre-existing categories) evolve to incorporate the sounds of the learners' phonetic system rather than construct a new category for L2, as the SLM predicts.

2.2.4 Second Language Linguistic Perception Model

The second language linguistic perception model (L2LP), developed by Escudero (Escudero, 2005; van Leussen & Escudero, 2015), is a relatively newer model that predicts the development of L2 speech perception. According to the model, there are separate perception grammars for L1 and L2, and L1 and L2 do not share the same phonological space. To form the separate grammars, L2 learners copy the L1 perception grammar when perceiving the L2 sounds at the initial stage of their learning (full copying hypothesis). The duplicated perception grammar gradually changes to a more optimal one for L2 sounds perception by sufficient L2 input and/or experience. Thus, the model implies the necessity of L2 input and experience for accurate L2 phonetic perception.

Regarding the degree of difficulty in each phonetic contrast, the model predicts three scenarios: new, similar, and subset. L2 phonetic contrasts, which do not exist in L1 (new scenario), are the most difficult because the contrasts classified in the new scenario need to adjust the L1 category by making a new category or dividing the existing category. A similar scenario includes patterns when L2 contrasts are acoustically similar to two separate L1 sounds. This scenario is less difficult than the new scenario because only the boundaries of the categories need to be adjusted for L2 sounds. The subset scenario consists of a multiple-category assimilation pattern. In other words, an L2 sound can be perceived as two L1 sounds. This scenario is considered as being less problematic.

The models reviewed above (i.e., SLM, PAM, NLM, and L2LP) argue that the difficulty in L2 perception depends on the similarities between the L1 and L2 acoustic features or articulatory gestures. Another noteworthy and encouraging point of the models for L2 learners, instructors, and researchers is that the L2 perception can be improved by sound input and linguistic experience of L2, such as through perception training.

2.3 Perceptual training studies

Speech acquisition models reviewed in the previous section confirm the importance of exposure to L2 sounds for improving learners' perception and production. To test and explore

the effects of L2 speech input on L2 perception and production development, numerous studies have used lab-based perceptual training paradigms (e.g., Bradlow et al., 1997; Sakai & Moorman, 2017; Thomson, 2012b). Perceptual training aims at providing participants with speech input of target L2 sounds. The length of training sessions varies among studies from one session (Wang & Munro, 1999) to 45 sessions over two weeks (Bradlow et al., 1997). Previous studies have found that perceptual training improves L2 learners' perceptual ability of target phonemes because the speech input can influence learners' phonetic categories (Logan et al., 1991; Nishi & Kewley-Port, 2007). Interestingly, learning effects can also be observed on the production of the trained phonemes (e.g., Bradlow et al., 1997; Thomson, 2011). Because of its efficacy and benefits along with practicality, perceptual training has drawn L2 instructors' attention to introduce the method in their class as homework (e.g., Hamada & Tsushima, 2001; Iino & Thomson, 2018) and a classroom activity (Lee et al., 2020; Ueda & Hashimoto, 2019). In addition, apps and websites that provide perceptual training to L2 learners. One of the most well-known perceptual training platforms is English Accent Coach developed by Thomson (2012a). This platform is accessible from laptops and smartphones free of charge, and learners can practice identifying English vowels and consonants by listening to target sounds spoken by multiple speakers, following the High Variability Phonetic/Perceptual Training (HVPT) method (see details for HVPT in Chapter 2.3.1.2). More recently, Thomson (2024) launched an app version called English Accent Coach Asteroids, which incorporates a gamification element.

2.3.1 Factors affecting the effects of perceptual training

Although perceptual training has been found to be effective for phonetic learning in general, previous research has suggested that several factors may determine the effectiveness of perceptual training, such as feedback, high variability in stimuli, and training types.

2.3.1.1 Feedback in perceptual training

Feedback plays a significant role in L2 acquisition. Ellis (2005) stated that feedback facilitates the rapid learning of L2, and studies have confirmed the positive effects of feedback in L2 acquisition (e.g., Henshaw, 2011; Saito & Lyster, 2012). Feedback in perceptual training can be divided into two broad types: summative feedback and immediate feedback. In terms of perceptual training, summative feedback can be provided at the end of a training session as in the form of final scores. In contrast, immediate feedback is given in each

trial, such as displaying ‘Correct’ on the screen with a high-pitched chime for a correct answer.

Feedback has been found to be effective in L2 learning. Felker et al. (2021) examined perceptual learning of a novel L2 accent in interactive dialogue settings. The findings indicated that explicit corrective feedback was more effective than generic feedback, and both types of feedback improved participants’ ability to process the accented words.

In L2 phonetic acquisition, Lee and Lyster (2016) explored the impact of corrective feedback on L2 speech perception with L1 Korean speakers. Participants learned the English phonetic contrast /i/–/ɪ/ in either an instruction-only condition or an instruction+corrective feedback condition. Results showed that the instruction+corrective feedback group outperformed the instruction-only group in both immediate and delayed post-tests.

The importance of feedback was further supported by a study that examined an adaptation effect in L2 phonetic learning. Fujita et al. (2019) investigated whether listeners could learn L2 sounds implicitly in a perceptual training paradigm without feedback. The findings showed that L2 learners’ perceptual performance did not improve after 15 to 20 minutes of training, though their response times decreased.

Similarly, studies have shown that explicit training is more effective than implicit learning (Reber, 1967). Lacabex and Gallardo-del-Puerto (2020) examined the impact of explicit phonetic training and implicit exposure to target sounds on students’ awareness of the English schwa. Results revealed a significant improvement in both explicit and implicit learning groups, with explicit instruction groups showing greater gains.

However, Vlahou et al. (2012) found contradictory findings. They compared explicit training with feedback, implicit training with feedback, and implicit training without feedback. Surprisingly, participants in all conditions improved, but the implicit training without feedback improved the most. Other studies have also shown that perceptual learning in an implicit learning paradigm where learners learn the target phonemes by listening to them repeatedly without feedback lead to perceptual learning (Baese-Berk, 2010; Hayes-Harb, 2007; Maye & Gerken, 2000, 2001; Toro et al., 2005).

Taken together, while some studies have demonstrated perceptual learning without explicit feedback, previous research has generally suggested that feedback plays an important role in perceptual learning, and its importance has been reflected in numerous L2 phonetic studies incorporating immediate and/or summative feedback in perceptual training (e.g., Bradlow et al., 1997; Inceoglu, 2016; Shinohara & Iverson, 2018; Thomson, 2012b).

2.3.1.2 High Variability Phonetic/Perceptual Training

Another factor that should be considered for perceptual training is the stimuli, especially the high variability of stimuli (i.e., various types of tokens spoken by multiple speakers). Previous research has found that training with highly variable stimuli, so called HVPT, leads to learning generalisation to untrained stimuli. The efficacy of HVPT research was first explored by Logan et al. (1991) who replicated an earlier perceptual training study conducted by Strange and Dittmann (1984). In Strange and Dittmann (1984), L1 Japanese speakers underwent AX (same or different) perceptual training on the English /l/-/r/ contrast with immediate feedback over three weeks. The training stimuli were recorded by only one L1 American English speaker. After 14 to 18 sessions of training, participants increased their discrimination accuracy, and the learning led to an improvement in categorical perception. However, their perception of the target phonemes was not robust for untrained natural speech.

Logan et al. (1991) replicated their study with highly variable stimuli for perceptual training. A total of 68 minimal pairs were recorded by four L1 English speakers. The researchers claimed that using highly variable stimuli could help learners be exposed to the full range of the exemplars of to-be-learned phonemes. After completing 15 training sessions, each consisting of 272 trials, participants improved their identification accuracy for both trained stimuli, untrained words, and untrained speakers, suggesting that perceptual learning through HVPT generalised to untrained stimuli. Logan et al. explained that exposure to the various examples of to-be-learned sounds helped learners construct phonetic representations that encompass a broader range of speech.

Lively et al. (1993) expanded on Logan et al.'s (1991) study by conducting two experiments that investigated the role of variability in training Japanese listeners to identify the English /r/-/l/ contrast. In Experiment 1, similar to Logan et al., participants were trained using multiple talkers and showed moderate but significant improvement in accuracy and response latency. Additionally, participants were able to generalise their learning to new words produced by both trained and untrained talkers. In Experiment 2, Lively et al. introduced a condition where participants were trained with tokens from a single talker. While participants improved during training, they failed to generalise to new talkers, emphasizing the importance of variability in stimuli for robust category formation. The results suggest that variability plays a crucial role in facilitating perceptual learning and generalising it to untrained stimuli.

Following these studies, many researchers have investigated the effects of HVPT and reported similar findings on English consonants (Bradlow et al., 1999; Bradlow et al., 1997; Huensch & Tremblay, 2015; Iino, 2018, 2019; Iino & Thomson, 2018; Iverson et al., 2005; Shinohara, 2014; Shinohara & Iverson, 2018, 2021), English vowels (Carlet & Cebrian, 2019; Iverson et al., 2012; Lengeris, 2018; Rato, 2014; Thomson, 2011; Wong, 2014), French vowels (Inceoglu, 2022), Japanese geminate-singleton (Sadakata & McQueen, 2013), Thai tones (Wright, 2021), and Mandarin tones (Perrachione et al., 2011).

Sadakata and McQueen (2013) investigated the effects of HPVT by conducting a 5-day perceptual training on the Japanese geminate and the singleton fricative /ss/–/s/ contrast with L1 Dutch speakers. Participants underwent HVPT or low high variability phonetic training (LVPT). The identification training consisted of 160 trials per day. Participants who received HVPT with stimuli recorded by five speakers improved more than those who underwent LVPT with stimuli recorded by a single speaker. The HVPT group also showed learning generalisation to new voices, and untrained vowel and consonants. However, the superiority of HVPT appeared only in an identification task but not in a discrimination task. These findings suggest that variability in stimuli facilitates constructing phonetic categories instead of improving pre-categorical processing.

Wong (2014) also compared the effects of HVPT and LVPT on the English /e/–/æ/ contrast for L1 Cantonese speakers. The results showed that although both groups improved their perception performance, the HVPT group outperformed the LVPT group. Wong claimed that exposure to highly variable stimuli prevents learners from relying on talker-specific features, leading to the attainment of robust L2 phonetic representations. This effect can be observed as the generalisation of perceptual learning to new stimuli, as reported by Logan et al. (1991).

Iverson et al. (2005) also found that L1 Japanese speakers successfully improved their identification accuracy of English /l/ and /r/ sounds after 10 sessions of 1.5-hour HVPT with 300 trials over 2 to 3 weeks. The perceptual learning generalised to new talkers and new words as well. However, the learning did not generalise to different positions of the target phonemes in a word; participants improved their identification accuracy for the trained position (e.g., word-initial position), but not for the other positions (i.e., medials and clusters).

Iino and colleagues conducted several studies examining the impact of HVPT on the perceptual accuracy of Japanese learners of English. In Iino and Thomson (2018), non-English majoring students participated in HVPT for ten weeks using English Accent Coach

(Thomson, 2012a), leading to significant improvement in their ability to perceive the English /l/ and /r/ contrast. Iino (2018) extended this research to identify which English phonemes were most difficult for Japanese learners, finding that /θ/, /z/, /s/, /l/, and /r/ were particularly challenging. After HVPT, perceptual accuracy improved for most consonants and vowels, though some sounds, such as /s/, /l/, and /v/, showed little improvement. Iino (2019) further tested the effects of HVPT on the perception of English fricatives, including /θ/, /f/, and /v/. Participants showed significant gains in perceptual accuracy following training.

The positive effect of HVPT was also observed on English vowels. Iverson et al. (2012) conducted HVPT on 14 English vowels in a 3- or 4-alternative forced-choice (AFC) identification task format with L1 French speakers. Stimuli were recorded by eight L1 British English speakers, and participants received eight training sessions of 225 trials over one to two weeks. Results revealed that participants significantly improved their identification accuracy for most vowels, excluding /ɪ/, /aɪ/, and /ʌ/.

Thomson (2011) investigated the effects of HVPT to improve the perception of English vowels. In the study, L1 Mandarin speakers completed eight HPVT training sessions, where they identified 200 vowel tokens produced by 20 L1 English speakers. The study found that learners significantly improved their identification of English vowels both in the training context and in one novel context, although the improvement did not generalise to a third phonetic context. The improvements were retained for at least one month after training, but no additional gains were observed. Additionally, learners reported increased confidence in their vowel identification after training.

Rato (2014) examined the effects of HVPT on the perception of English vowels by Portuguese learners of English. Participants underwent five training sessions using stimuli from six L1 American English speakers, focusing on three contrasts: /i/–/ɪ/, /ɛ/–/æ/, and /u/–/ʊ/. The training included both identification and discrimination tasks. The results showed significant improvement in participants' perceptual accuracy for these vowel contrasts, with the effects persisting in a delayed post-test two months later. Moreover, the learning generalised to new words and speakers, for two of the vowel contrasts.

Wright (2021) found that high variability in tokens within training trials also played a significant role. Wright conducted a series of experiments on HPVT of four Thai tone categories with L1 English speakers. Participants were divided into two groups: in the High Variability within trials group, they were exposed to five different tokens of the same tone in

each trial, while in the Identical Token group, they heard the same token repeatedly within a trial. In each trial, participants were required to choose one of four boxes, with the location of the "X" corresponding to the tone they heard. Results showed that the High Variability group achieved higher accuracy and faster reaction times compared to the Identical Token group, and the learning generalised to new stimuli and talkers.

The positive effects of high variability in stimuli are supported by an adaptation study as well. Lee and Baese-Berk (2021) investigated how L2 English listeners enhance their ability to comprehend unfamiliar speech—process known as adaptation—and whether talker variability affects this adaptation. In the study, L2 listeners were exposed to speech from L1 English speakers in two conditions: a high variability condition, where participants heard speech from multiple talkers, and a low variability condition, where participants heard speech from a single talker. During the task, participants listened to sentences and then identified whether a word presented on the screen matched the final word of the sentence they had just heard. Results revealed that participants showed rapid adaptation to English speech, with response times decreasing significantly across the first three blocks. The high variability condition initially caused slower response times but led to better cross-talker generalisation and faster within-block adaptation to novel speakers. In contrast, participants in the low variability condition adapted more quickly to a single talker but did not generalise to new speakers. Overall, exposure to high talker variability improved listeners' ability to generalise to unfamiliar talkers, despite the initial difficulty.

Although many studies have confirmed that HVPT is effective on L2 perceptual learning, some studies have suggested that variability in stimuli might slow perceptual learning. Perrachione et al. (2011) investigated how individual differences in perceptual aptitude influence the effectiveness of HVPT for learning untrained phonetic contrasts, specifically lexical tones. L1 English speakers were divided into high-aptitude and low-aptitude groups based on their initial pitch perceptual abilities. The study found that high-aptitude participants benefited from HVPT, showing better learning and generalisation to new stimuli. In contrast, low-aptitude participants struggled with HVPT, particularly when faced with trial-by-trial variability, which made it difficult for them to focus on relevant phonetic features. To address this, the second experiment reduced trial-by-trial variability while maintaining overall variability. Results showed that the reduced variability improved learning outcomes for both groups, especially for low-aptitude participants. The findings suggest that

HVPT is more effective when tailored to learners' perceptual abilities, with lower variability or gradual exposure benefiting those with weaker perceptual skills.

The reason why high variability in stimuli might slow perceptual learning may be attributed to increased cognitive load when processing different speakers. Magnuson et al. (2021) investigated how listeners manage cognitive load when processing speech across talker variability. Participants were exposed to speech from family members, newly familiar talkers, and unfamiliar speakers. The cognitive load, or 'talker-change cost,' was measured by tracking reaction times and processing efficiency in speech recognition tasks. Results showed that, even with familiar voices, a constant cognitive load remained, suggesting that adjusting to each new talker's acoustic characteristics is effortful.

In conclusion, the evidence from various studies underscores the effectiveness of HVPT in improving perceptual accuracy and facilitating generalisation to new stimuli and voice. However, research also highlights that variability in stimuli can slow perceptual learning, particularly for learners with weaker perceptual abilities.

2.3.1.3 Discrimination and identification training

Considerable discussion has focused on the differences between two types of training, discrimination training and identification training. Discrimination training targets the ability to hear the differences between two sounds and includes tasks such as AX and ABX. In an AX task, learners listen to two sounds ('A' and 'X') in a row and determine whether those sounds are the same or different, in other words whether the 'X' was the same as the 'A'. In an ABX task, learners listen to three sounds and determine whether the last sound 'X' was the same as the first one 'A' or the second one 'B.' Of the two types of discrimination task, AX is easier than ABX due to the differences in cognitive load (Daidone, 2017). Because cognitive demand can increase according to a memory load, ABX, where learners need to store three stimuli in their short-term memory for each trial, can be more demanding than AX. Stimuli used in a discrimination task can be either within-categorical or categorical, meaning that participants judge differences between stimuli that are phonetically distinct but fall within the same category or between two sounds that belong to different phonetic categories.

Identification tasks focus on whether learners can identify which phoneme they have heard in a forced-choice paradigm. During the task, learners hear a word and then choose the corresponding word or the symbol representing the sound they heard from the visually

presented options. In contrast to discrimination tasks, stimuli used in identification tasks are generally composed of minimal pairs of target phonemes.

Some researchers claim that identification training is more effective than discrimination training (Jamieson & Morosan, 1986; Logan et al., 1991; Strange, 1992; Strange & Dittmann, 1984). Strange and Dittmann (1984) conducted AX discrimination training on the English /l/-/r/ contrast with L1 Japanese speakers and found that Japanese speakers improved both identification and discrimination task scores only for trained stimuli, but not for untrained stimuli. Strange (1992) also showed that categorical discrimination tasks improved perceptual accuracy only for trained stimuli. In contrast, Logan et al. (1991) later reported generalisation effects to untrained stimuli following identification training on the same phonetic contrast with the same L1 population. They argued that while discrimination tasks increase sensory resolution, they do not contribute to the formation of robust L2 phonetic categories. Moreover, Jamieson and Morosan (1986) explained that identification tasks can help learners tap into their phonetic representations by having learners try to label stimuli, whilst discrimination tasks require learners to only compare the acoustic differences.

More recent studies also suggested the superiority of identification training (Carlet & Cebrian, 2019; Grenon et al., 2018; Law et al., 2019). Law et al. (2019) compared the effects of discrimination training and identification training on the English /z/-/dz/ with L1 Japanese speakers. Participants underwent one of the two training types with a total of 512 tokens. The results revealed that neither training type significantly improved accuracy in the identification task, but identification training produced slightly better outcomes compared to discrimination training. Grenon et al. (2018) also showed that after discrimination training, participants performed significantly better at using vowel quality to distinguish the /i/-/ɪ/ contrast. However, Grenon et al. also found that 25% of them mislabelled the phonemes in the identification task. Therefore, they hypothesised that discrimination training can be useful with the instructions of phoneme-grapheme correspondences.

Carlet and Cebrian (2019) compared the effects of two types of perceptual training, identification and discrimination training, on the perception and production of English vowels by L2 Spanish/Catalan speakers. Participants were assigned to one of three groups: identification training, discrimination training, or a control group with no training. Both treatment groups showed improvements in vowel perception and outperformed the control group. However, the identification training group demonstrated significantly larger gains in perception compared to the discrimination training group. These findings suggest a potential

superiority of identification training for L2 perceptual learning. The study suggests that both training methods can enhance L2 learning, but identification training may provide additional benefits, especially for tasks involving vowel identification. However, participants' perceptual accuracy was measured only in the identification task. Therefore, the superiority of identification training may have been attributed to a task familiarity effect.

In contrast, other researchers argue that both training tasks have similar effects. A study conducted by Flege (1995b) showed that both identification and discrimination training groups improved their L2 perception of the target phonemes. Moreover, the learning generalised to untrained stimuli and was retained for two months after the training sessions, suggesting the effectiveness of both training. Likewise, Wayland and Li (2008) compared the effects of two training types, AX training and 2AFC training, on two Thai tones with L1 English and L1 Chinese participants. After two training sessions of 80 trials, both groups demonstrated similar outcomes. Likewise, Wee et al. (2019) compared the effects of the two types of training on the L2 English /i/–/ɪ/. They found that participants in both training groups significantly improved their identification accuracy by ignoring vowel length and focusing on spectral cues, similar to L1 speakers of English.

A study with a different methodological design also supported that discrimination training was as effective as identification training. Shinohara and Iverson (2018) combined identification training and discrimination training and compared the effects of these training types directly. L1 Japanese learners received five 30-min sessions of both training types (ten sessions in total) on the English /l/–/r/ contrast. Each session of identification training was composed of 300 2AFC trials. As for discrimination training, each session consisted of 200 trials conducted in a 3AFC task. Results showed that both training types showed significant improvement in both identification and discrimination of the target phonetic contrast. Therefore, Shinohara and Iverson claimed that both training types are equally effective for perceptual improvement.

Nozawa (2015) identified differential effects of training on vowels versus consonants. In the study, L1 Japanese speakers participated and received training on either English vowels or coda nasals in either an identification or discrimination training condition. After two sessions of training, the results of pre- and post-test identification tasks showed similar gains for coda nasals in both the identification and discrimination training groups. However, for vowels, the identification training group outperformed the discrimination training group,

indicating that identification training was more effective for vowel learning than discrimination training.

It is also possible that discrimination and identification abilities are independent. Iverson et al. (2012) found that L1 French speakers improved discrimination accuracy for only three pairs out of 17 after identification training on English vowels, although they improved identification accuracy for 11 out of 14 vowels. Therefore, the researchers claimed that identification training does not necessarily affect discrimination accuracy.

In summary, studies have shown mixed findings on whether discrimination training is as effective as identification training, and more studies are needed to further investigate the effects of the two types of training.

2.3.2 Effects of perceptual training on production

The relationship between L2 speech perception and production remains complex and has yet to be fully untangled. However, much of the literature suggests that L2 learners' speech perception tends to be better than or at least equal to their speech production abilities.

One prominent theory explaining this relationship is Flege's SLM (1995a). According to SLM, accurate speech perception is a prerequisite for accurate speech production. This model suggests that improvement in L2 perceptual abilities lead to corresponding improvement in speech production.

A study conducted by Kissling (2014) supported the hypothesis that perceptual ability is a significant predictor of improvement in pronunciation after explicit phonetics instruction. The study found that learners who could better perceive subtle differences between L2 and L1 sounds were more likely to produce the target sounds accurately after pronunciation instruction, especially when accompanied by explicit phonetic training.

Another influential theory is the Motor Theory of Speech Perception, proposed by Liberman and Mattingly (1985). This theory posits a direct link between perception and production, as both rely on the information of articulatory gestures. According to this theory, the same neural mechanisms are involved in both perceiving and producing speech, suggesting that these two processes develop simultaneously. Consequently, the Motor Theory predicts that improvement in one domain should be closely mirrored by improvement in the other.

Supporting these theoretical claims, correlational studies have shown a strong relationship between perception and production accuracy in L2 learners (Flege, 1999b; Flege

et al., 1997). These studies consistently demonstrate that higher accuracy in L2 perception correlates with better accuracy in L2 production, reinforcing the notion of a close, interdependent relationship between the two.

In contrast, some researchers claim that speech production learning can precede perceptual learning (Goto, 1971; Hattori, 2009; Sheldon & Strange, 1982; Yamada et al., 1994). Goto's (1971) study revealed a significant disconnection between perception and production by exploring the difficulties that L1 Japanese speakers face when perceiving and producing the English /l/ and /r/ sounds. Although some Japanese participants could produce the /l/ and /r/ sounds with relative accuracy, they still struggled to perceive the distinction between them. This suggests that accurate production of sounds does not necessarily correspond to accurate perception.

A study conducted by Sheldon and Strange (1982) also showed that perception performance is better than production performance. Their study investigated the relationship between perceptual accuracy and intelligibility of the English /l-/r/ contrast with L1 Japanese speakers. The results showed that, at least for some participants, production performances were better than perception performances. However, the sample size was small for this kind of study (only six participants), and a bigger sample size is necessary to enhance the generalisability of the findings for the relationship between perception and production.

Hattori (2009) found that production can improve without perceptual learning after production training. In the study, L1 Japanese learners of English received pronunciation instructions on the English /l-/r/ contrast. As a result, they significantly improved their production accuracy of the target contrast but did not show a significant improvement in their perceptual accuracy. These findings support the theory that production learning can precede perception learning, and that learning in each modality, at least for production learning, does not always mirror the other modality.

Baese-Berk (2019) found that the relationship between the two modalities can be independent and can even be negative. The study showed that while perceptual training often leads to improvement in both perception and production, production training does not consistently improve perceptual abilities. In fact, production training can disrupt perceptual learning, indicating a potential negative relationship between the two modalities. In addition, the study revealed individual differences in how learners responded to training. For some learners, production training may hinder perceptual development, while others may show progress in both areas.

In fact, studies have reported no correlation between the two modalities and individual differences in terms of the degree of learning (Bradlow et al., 1997; de Jong et al., 2009; Huensch & Tremblay, 2015; Sakai & Moorman, 2017; Thorin et al., 2018). For example, Bradlow et al. (1997) found that while perceptual training led to improvement in both perception and production after perceptual training, there was no significant correlation between the degree of learning in the two modalities. This indicates that improvement in perception did not necessarily predict improvement in production, suggesting that perception and production can develop independently with individual differences. Thorin et al. (2018) also found that perceptual learning transferred to production learning after perceptual training, but the correlation between the two modalities was not strong. These findings suggest no direct link between perception and production learning.

Nagle and Baese-Berk (2021) proposed that the relationship between perception and production varies based on the learners' developmental stage, with a weaker connection in the early stages of learning. Therefore, longitudinal studies should be conducted for better understanding of the relationship between perception and production.

Indeed, some studies suggest that the relationship between perception and production abilities change over the course of learning. Yamada et al. (1994) found a positive correlation between the two modalities in their analysis of L1 Japanese speakers' learning of English /l/, /r/, and /w/ sounds in both Japan and the USA in an immersion context. However, when the participants in the USA were excluded, the regression line flattened, indicating a weaker correlation. Similarly, Jia et al. (2006) found a weaker relationship between perception and production among beginners compared to more advanced learners. These findings suggest that, in the early stage of learning, the relationship between perception and production is not strong, and some learners may produce sounds more accurately than they can perceive them. However, as learning progresses, the relationship strengthens.

Although there are mixed findings regarding the link between L2 speech perception and production, many studies have shown that perceptual training improves not only L2 learners' perceptual accuracy but also production accuracy (Sakai & Moorman, 2017). Studies have reported a transfer from perceptual learning to production learning on English consonants (Bradlow et al., 1999; Bradlow et al., 1997; Huensch & Tremblay, 2015; Lopez-Soto & Kewley-Port, 2009; Shinohara & Iverson, 2018), English vowels (Iverson et al., 2012;

Lambacher et al., 2005; Lengeris, 2018; Thomson, 2011; Wong, 2014), Japanese geminates (Motohashi-Saigo & Hardison, 2009), and French nasal vowels (Inceoglu, 2016).

One of the of the earliest studies that investigated the transfer from perceptual learning to production learning was conducted by Bradlow et al. (1997). They investigated the effects of perceptual training on the English /l/-/r/ with L1 Japanese speaker. Stimuli were recorded by five speakers, and the training consisted of 45 sessions. Results showed that the treatment group significantly outperformed the control group at post-test, and the learning generalised to untrained words recorded by a trained speaker and an untrained speaker. The researchers also confirmed that the learning effects persisted at the delayed post-test which was conducted three months after training.

Motohashi-Saigo and Hardison (2009) conducted perceptual training on Japanese geminates for L1 English speakers. After ten training sessions, significant improvement in production performance was observed with a high effect size ($d = 1.24$), which was the largest effect size among perception training studies analysed in Sakai and Moorman's (2017) meta-analysis.

Lee et al. (2020) investigated the effects of pronunciation training under perception or production conditions in a classroom setting with L1 Japanese university students. The training consisted of two sessions of 30-minute pronunciation treatment focusing on either segmental or suprasegmental features, incorporating explicit instructions, teacher-led activities and pair work. After the training, while all conditions resulted in significant improvement in pronunciation, the perception condition led to greater gains in both segmental and suprasegmental features compared to the production condition.

In Lengeris (2018), the study investigated the effectiveness of perceptual training on the production and perception of English vowels by L1 Greek speakers. The training was conducted in an identification format and was composed of five sessions. Results indicated that perceptual training significantly improved the participants' production of English vowels, both in sentence reading and spontaneous speech, though improvement was greater in the former. Wong (2014) also revealed production improvement in English vowels after perceptual training. Participants were L1 Cantonese speakers who received 10 sessions of 10-minute training over 5 days. After the training, they significantly improved both perception and production.

Some studies, however, reported that transfer effects from perception to production by perception training had not been significant. Aliaga-Garcia and Mora (2009) conducted six two-hour perceptual training on English vowels for L1 Catalan/Spanish speakers. Results showed improvement in the production of voiceless stops but not vowels. Hwang and Lee (2015) also conducted perceptual training with English vowels and consonants for 11 to 12-year-old Korean speakers and found no significant effects of perceptual training on production.

Iino and Thomson (2018) examined the effects of HVPT on the perception and production of the English /l/ and /r/ sounds. Non-English majoring students at a Japanese university participated in training over ten weeks. While perceptual accuracy improved significantly, production improvement was minimal, suggesting that articulation practice may be necessary for production gains in EFL contexts. A follow-up study (Iino, 2019) further supported the findings with different target sounds, English fricatives /θ/, /f/, and /v/. While perception improved, production gains were minimal, reinforcing the need for additional articulation practice.

Iverson et al. (2012) found that L1 French speakers significantly improved their identification accuracy for 11 out of 14 English vowels but improved their production accuracy for only three vowels. The researchers analysed the production data and found that errors were due to reading mistakes because participants were unfamiliar with the reading of certain words. During training, participants learned to associate each spelling with its target-like pronunciation, which might have led to production improvement. Therefore, the researchers claimed that the improvement in perception and that of production were caused by fundamentally different factors.

Research has suggested that the effects of perceptual training may vary depending on the target phonemes. Sakai and Moorman (2017) showed that perception training improves learners' production of obstruents better than sonorants or vowels. They explained that the degree of differences among phonemes is attributed to the place of articulation and the ease in perceiving phonetic gestures from listening to the sounds. The finding aligns with that of a study conducted by Ueda and Hashimoto (2019). In the study, L1 Japanese speakers investigated the effects of classroom-based perceptual training on the English /b/-/v/, /l/-/r/, and /s/-/θ/ contrasts. Results showed that although participants improved the production of /b/-/v/ and /s/-/θ/ contrasts (obstruents), they did not improve the /l/-/r/ contrast (sonorants).

In summary, while the relationship between perception and production remains open to question, perceptual training appears to be effective for both. However, more research is needed to understand the extent of its effectiveness for speech perception and production.

2.4 Production training studies

Previous research has shown that training focusing on L2 production is effective in achieving more accurate pronunciation. However, there are several variables that determine the effectiveness of production training, such as the type of training and the nature of feedback. In addition, similar to perceptual training, some studies have demonstrated transfer effects from production learning to perceptual learning; specifically, production training can enhance perceptual accuracy even without direct perceptual training. The following section will explore these variables and transfer effects in greater detail.

2.4.1 Effects of production training on production

Previous studies have shown that pronunciation practice is effective in improving learners' pronunciation accuracy. According to a meta-analysis conducted by Thomson and Derwing (2014), 82% of the 75 studies reviewed showed significant improvement in the pronunciation of target sounds following pronunciation instruction. Similarly, a meta-analysis conducted by Saito (2012) reported that all treatment groups experiences significant improvements in production, except for the studies where the treatment lasted only for 15 to 30 minutes and participants exhibited a ceiling effect at pre-test. Additionally, a meta-analysis conducted by Lee et al. (2015) confirmed that pronunciation instruction significantly impacts L2 production. The analysis highlights that learner age influences the effect of treatment, and that laboratory-based pronunciation instruction tends to be more effective than classroom-based one. However, it was noted that using technology for instruction was less effective than instruction provided exclusively by instructors. The meta-analysis also revealed that more significant effects were observed when productions were elicited in controlled measures, such as word-list reading, compared to an open-ended one. Importantly, the analysis underscored that instruction incorporating feedback brought about better learning outcomes.

The last point aligns with Saito and Lyster's (2012) findings on the importance of feedback in production learning. In their study, L1 Japanese speakers participated in English lessons for 4 hours in Canada. The first group received form-focus instruction (FFI) on the English /r/ sound accompanied by corrective feedback (CF). The second group received FFI

without CF, and the third group, serving as a control, received similar instruction but targeted English vowels instead. The FFI targeting English /r/ involved 38 minimally-paired words within lessons on English argument skills. Corrective feedback was provided through recasts when students mispronounced words. The results revealed that the group receiving FFI with CF produced significantly more accurate /r/ sounds in both controlled and spontaneous speech, as measured by F3 values and L1 listener ratings. In contrast, the groups without CF did not show significant improvement, clearly demonstrating the critical role of feedback in enhancing pronunciation accuracy.

Other studies also support the importance of feedback in production training (Dlaska & Krekeler, 2013; Flege, 1988; Kartushina et al., 2015; Neri et al., 2006; Tejedor-García et al., 2017; Wang & Young, 2015). For example, Flege (1988) conducted a production training study on the English vowel contrasts /i/–/ɪ/ and /æ/–/ɑ/ with one L1 Spanish speaker using an optoelectronic glossometer, which visually displays speaker's tongue position. After three-day training, the participant became able to produce only the /i/–/ɪ/ contrast more distinctively. Flege suggested that the participant's lack of improvement on the /æ/–/ɑ/ contrast was likely due to the subtle differences in tongue position for these vowels not being clearly visible on the glossometer images, which highlights the crucial role of clear feedback in production learning.

Kartushina et al. (2015) took a different approach by focusing on acoustic properties, specifically F1 and F2 values, to investigate the effectiveness of immediate feedback on learners' production. L1 French speakers practiced producing four Danish vowels either with or without immediate feedback on their F1 and F2 values of their productions. The feedback was provided using a real-time acoustic analysis tool that displays the F1 and F2 values on the screen. After five training sessions, the group that received immediate feedback during training became able to produce the target vowels more similarly to L1 speakers' pronunciation.

Although receiving feedback is important to achieve accurate pronunciation as discussed above, many learners lack opportunities to receive feedback from L1 speakers and language teachers. However, recent advancements in speech processing technologies have made it possible for learners to receive feedback on their speech and perception through various tools. Pronunciation training using such technologies is known as Computer Assisted Pronunciation Training (CAPT) (Bliss et al., 2018). Studies have demonstrated that CAPT

and other Computer Assisted Language Learning (CALL) systems are effective in improving pronunciation (Martin, 2020a, 2020b; Neri et al., 2008; Wang & Young, 2015).

Martin (2020b) explored the effectiveness of CAPT in a distance language course. Participants were L2 learners enrolled in an online German course. The CAPT focused on both segmental and suprasegmental features, with participants receiving immediate feedback on their perceptual accuracy. Regarding production, participants practiced it by imitating recorded L1 speakers' models. The results showed that participants who underwent CAPT improved their perception and production skills more than those who did not receive such training. In another study, Martin (2020a) included a teacher-led pronunciation component and further examined the effectiveness of CAPT. The study found no significant difference in learning outcomes between the two groups that received pronunciation training either online or from a teacher. A questionnaire survey revealed that participants in both treatment groups were satisfied with the training they received.

Neri et al. (2008) also found that CAPT was as effective as in-class pronunciation training. Eleven-year-old Italian children practised English word-level pronunciation either in a CAPT condition or a teacher-led condition for four weeks. The CAPT system, named PARLING, focused on improving pronunciation through interactive exercises that allowed learners to hear L1 speakers' models and practice their pronunciation. It provided immediate feedback, helping learners improve their pronunciation with each attempt. Results indicated that both groups equally improved their pronunciation.

Some studies have focused on the effects of specific speech technologies, such as waveforms, spectrograms, and pitch displays on pronunciation learning (Akahane-Yamada et al., 1997; Hardison, 2004; Herd et al., 2013; Hirata, 2004; Olson, 2014; Thorin, 2020). Akahane-Yamada et al. (1997) conducted a case study to investigate whether using spectrographic images could help L1 Japanese speakers learn the English /l/ and /r/ sounds. In the study, a male Japanese speaker first received an instruction on how to pronounce the target sounds. Then, he practiced pronunciation by reproducing words containing the target phonemes, using the spectrographic images of his pronunciation compared to an L1 speakers' model as feedback. The practice lasted five hours over three days. His pre-test and post-test speech productions were judged in a 2AFC task and a paired-comparison task by an L1 American English speaker. In the paired-comparison test, the judge listened to the pre-test and post-test versions of each word and rated which version was better or whether both were

equal. Results revealed a significant increase in both identification scores (from 44.1% to 97.1%) and paired-comparison scores (from 16.2% in the pre-test to 71.3% in the post-test).

Hattori (2009) also explored the effectiveness of using spectrographic images to improve the pronunciations of the English /l/ and /r/ sounds with a group of L1 Japanese speakers. Participants first received instructions on the target sounds, which included watching a video of an L1 speaker producing the sounds. Then, they completed 10 sessions of 30-40 minute one-on-one training. The results indicated that participants' pronunciation became more intelligible as assessed by L1 speakers' ratings.

Herd et al. (2013) conducted production training using Praat for visual images, such as waveform and spectrograms of model speech recorded by L1 speakers. L1 American English learners of Spanish at the intermediate level learned three Spanish intervocalic sounds, /d, ɾ, r/. In each trial, participants first inspected an L1 speaker's production of a word, produced the same word, and then compared the waveform and spectrogram of their own speech with those of an L1 speaker. Participants underwent six training sessions, and each lasted 35 to 45 minutes, except for the first session, which took over an hour due to a tutorial on identifying the target phonemes using visual images. Participants' production data were collected in a word-list reading task and judged by eight L1 Spanish speakers in a three-alternative forced choice identification task format. The data were also analysed based on the waveform and spectrogram on Praat. Results revealed a significant increase in the participants' judgment scores by L1 listeners overall. When analysing the results for each contrast separately, there was a notable improvement only for the Spanish trill /r/.

Visual feedback on learners' pronunciation facilitates learning not only at the segmental level but also at the suprasegmental level. Hardison (2004) conducted a three-week pronunciation training on French prosody using a pitch display system that provided real-time feedback with L1 English speakers. Participants engaged in 13 sessions of 40-minute pronunciation training over the course of three weeks. In each session, participants read 30 sentences and compared the visual image of the pitch contour of their productions with a model pitch contour. Results revealed that participants' production improved at both the segmental and prosody levels, and the learning generalised to untrained sentences.

Hirata (2004) observed a significant improvement in the Japanese pitch accent (low vs high) and durational contrasts (short vs long) after pronunciation training using fundamental frequency contours for feedback. During training, L1 English learners of Japanese listened to the model speech of either a word in isolation, a word in a phrase, or a word in a sentence and

checked the contour of the speech. Then, they practised pronouncing the word, phrase, or sentence by trying to match the contour of their production to that of a model speech. The training consisted of 10 sessions: the first three sessions used word stimuli, the fourth used phrase stimuli, and the final six sessions used sentence stimuli. Results showed that learners significantly improved their pronunciation of pitch accent and durational contrasts in both isolated word and word-in-sentence contexts. Moreover, the degree of improvement was significantly greater than that of the control group, which did not undertake pronunciation training.

More recently, Thorin (2020), however, found a smaller effect of production training compared to previous studies when using Praat's visual images for feedback. In the study, L1 Dutch speakers were presented with the F1-F2 space of their productions as visual feedback on Praat to train them pronouncing the English /æ/ and /ɛ/ sounds. Participants received two 1.5-hour training sessions where they read aloud words containing the target phonemes. In each trial, participants received feedback that showed the F1-F2 space of their own speech production compared to the L1 speakers' model, along with the location of the L1 speakers' model of the corresponding phoneme. Before the production training, participants were also given pronunciation instructions on the target phonemes, including guidance on tongue position and mouth openness. After training, participants attained more target-like pronunciation in both trained and untrained words. However, their production development was not significantly different from that of the control group, in which participants received the same pronunciation instructions and training, except for the feedback. These findings suggest that production training is beneficial even without feedback. However, it is important to note that both groups received explicit pronunciation instructions, which may have obscured the specific effect of using Praat for feedback.

Although it is not as easily accessible as the technologies listed above, ultrasound imaging has been suggested as an effective tool for practicing L2 pronunciation. One of the obstacles to mastering difficult pronunciation is identifying and replicating the target tongue position. However, tongue position is internal and not easily visible. Ultrasound imaging addresses this issue by enabling L2 learners to visualise their tongue position by putting the transducer—an ultrasound probe—right under the chin (Bird & Gick, 2018). This helps learners understand where and how to move their tongue and aids in remembering the target position. Studies have shown that using ultrasound images is somewhat effective in learning

the target-like articulation of English consonants with L1 Japanese speakers (Bryfonski, 2023; Tateishi & Winters, 2015; Tsui, 2012; Yamane et al., 2015), English vowels with L1 French speakers (Kühnert & Pillot-Loiseau, 2022), and French vowels with both L1 Japanese (Antolík et al., 2019) and L1 English speakers (Inceoglu & Gnevsheva, 2020).

In Tsui's (2012) study, participants received four sessions of 45-minute production training on the English /l/ and /r/ contrast using ultrasound imaging. In addition to the production training, participants also practised reading word lists at home each day during the training period. After training, participants' production of the target phonemes became closer to L1 speakers' models, and the pronunciation learning generalised to untrained words. A delayed-posttest conducted two weeks later confirmed the retention of this learning.

Other studies, however, have shown a partial improvement. Tateishi and Winters (2015) investigated whether production training using ultrasound improves the pronunciation of the English /l/ and /r/ sounds with L1 Japanese speakers. During training, participants practised producing the target sounds in an isolated condition, CV syllables, and monosyllabic words while comparing their tongue positions and movements with L1 English speakers' models. After five half-hour training sessions, participants improved their pronunciation of the English /l/ sound based on acoustic analysis, but there was no significant improvement for the /r/ sound.

Similarly, Bryfonski (2023) investigated the effects of corrective feedback using ultrasound imaging. Participants were L1 Japanese speakers and received corrective feedback either in an oral condition or in an oral plus visual condition using ultrasound. They completed a speaking task and received corrective feedback on English /l/ and /r/ sounds during training. Their production data were collected in story-reading and word-list reading tasks. Results indicated that only participants who received ultrasound imaging feedback significantly improved production accuracy. However, this improvement was observed only in the story-reading tasks. A questionnaire survey revealed that participants found feedback using ultrasound imaging both meaningful and motivating. This positive response to using ultrasound imaging aligns with findings from Yamane et al. (2015). They reported that 88% of participants who underwent production training using ultrasound imaging responded that it helped them understand how each sound is articulated.

Kühnert and Pillot-Loiseau (2022) also reported partial improvements in production learning. They conducted production training using ultrasound imaging on two English vowel contrasts with L1 French speakers. After five training sessions, participants improved their

pronunciation of only one contrast. Similarly, Antolík et al. (2019) investigated the effects of ultrasound imaging on French vowels with L1 Japanese speakers. Participants received three 45-minute individual training using ultrasound imaging. Results revealed that three out of six participants improved their pronunciation of the target phonemes.

Expanding the exploration of ultrasound imaging, Inceoglu and Gnevshva (2020) introduced ultrasound imaging into a classroom setting. Participants were students in a French phonetics course at a university in Australia. One group started practicing using ultrasound imaging with the French /y/–/u/ contrast, which was the target contrast of the study, while the other group practiced with the French /e/–/ɛ/ contrast for counterbalancing purposes. Participants' pronunciation of /u/ became more accurate after practicing with ultrasound imaging. However, there was no significant change in /y/, possibly due to a ceiling effect.

While ultrasound imaging provides visual feedback on tongue placement, Automatic Speech Recognition (ASR) technology, or speech-to-text technology, provides visual feedback through written output, making it easy for learners to see how accurate their pronunciation is. There has been a growing interest in using ASR technology as a tool for production training due to its greater accessibility and ease of use. ASR recognises human's speech and responds accordingly based on the service. Many platforms utilise ASR technology, such as Google Docs and Dragon Dictation for typing and Apple's Siri, Amazon's Alexa, and Google Assistant for personal assistance. Due to its convenience and ubiquity, ASR is a useful tool for L2 learners to test the intelligibility of their pronunciation. While some might argue that ASR cannot recognise human speech as accurately as humans do, studies have shown that the technology has advanced significantly. For instance, a study in 2000 reported much lower recognition rates for ASR systems like Dragon Naturally Speaking compared to human raters (Derwing et al., 2000). However, more recent studies have demonstrated that ASR's identification accuracy generally aligns with human raters (Guskaroska, 2019; Inceoglu et al., 2023; McCrocklin & Edalatishams, 2020; McCrocklin et al., 2019).

Guskaroska (2019) investigated the accuracy of the ASR system of Apple's Enhanced Dictation on the pronunciation of English vowels (i.e., /i/–/ɪ/, /æ/–/ɛ/, /u/–/ʊ/, and /ɑ/–/ʌ/) by L1 English and Macedonian speakers. The study revealed that the ASR system accurately identified 100% of the target vowels spoken by L1 English speakers, with the exception of /ɛ/ (66%) and /ʌ/ (33%). Regarding Macedonian speakers' speech, the recognition rate was lower

in general, especially for the phonemes that they have difficulty producing accurately. In another comparison, McCrocklin et al. (2019) analysed the performance of different ASR systems and found that Google Voice Typing's recognition rate was higher than Windows Speech Recognition, leading them to conclude that Google's ASR technology is approaching the accuracy level of human raters.

To further investigate the accuracy of ASR, McCrocklin and Edalatishams (2020) examined the relationship between the output of the Google Voice Typing's ASR and that of human raters. Speech materials were recorded by three L1 groups: English, Mandarin Chinese, and Spanish speakers. Results revealed a significant overall relationship between Google's ASR outputs and human raters' evaluations. However, when analysed separately by group, only the L1 Chinese group showed a significant relationship with human raters in terms of intelligibility, comprehensibility, and accentedness scores, though not with segmental errors. Additionally, Google ASR was found to identify longer words more accurately, and shorter words (i.e., one-syllable words) were more likely to lead an error. Inceoglu et al. (2023) also examined how ASR recognised L2 English speakers' speech in comparison to human raters. In their study, L1 Mandarin Chinese speakers' speech was transcribed by Google Assistant and then evaluated by L1 English raters. The researchers found that overall, the recognition patterns were similar between ASR and human raters. However, there were differences in the recognition patterns for individual speakers, indicating that ASR and L1 raters do not always assess speech in exactly the same way.

As reviewed above, recent ASR technologies are mostly reliable and can generate similar outputs as human raters. Many researchers investigated whether such ASR technologies can be utilized as immediate feedback in production training. Studies have shown the effectiveness of ASR on overall pronunciation (Elimat & AbuSeileek, 2014; Mroz, 2020; Spring & Tabuchi, 2021; Wang & Young, 2015) and at the segmental level (Chen et al., 2020; García et al., 2020; Inceoglu et al., 2020; Liakin et al., 2013, 2015; McCrocklin, 2019; Spring & Tabuchi, 2022).

Spring and Tabuchi (2021) incorporated an ASR-based pronunciation practice tool into their weekly online English class for Japanese learners of English at elementary to intermediate proficiency levels. At the beginning of each class, students first practiced pronouncing eight words and sentences using an ASR-based practice tool called NatTos. In NatTos, learners could hear the model pronunciation of a word, repeat the word, and then

receive ASR-based feedback on accuracy and fluency. Then, their teacher provided instructions on how to pronounce challenging items based on students' mistakes. Finally, students practised the same items using NatTos again. After five weeks of training, students with initially lower intelligibility improved their pronunciation by 5%, while students with higher initial intelligibility showed a ceiling effect.

The positive effect of ASR on pronunciation learning has also been observed in school children. For example, Elimat and AbuSeileek (2014) investigated whether pronunciation training using ASR enhances pronunciation learning with nine-year-old students in Jordan. Students underwent pronunciation training in either an ASR-based method or a traditional instruction method. The ASR-based method was conducted using the Tell Me More Performance English program over eight weeks. Results showed that the ASR-based training significantly improved pronunciation learning compared to the traditional training style. In addition, it was also found that ASR-based training was more effective when students engaged in training individually rather than in pairs or groups.

A mobile-based ASR system was also reported to be effective for L2 pronunciation learning. For instance, in Mroz (2020), participants practiced L2 French pronunciation using mobile-based ASR system within Gmail. The study found that, based on a combination of L1 listener ratings and machine analyses, participants who practised French pronunciation with the mobile-based ASR system surpassed those who practised without it.

Wang and Young (2015) examined the effects of summative and formative feedback on college students and seventh graders using a CALL system with ASR technology. All participants underwent production practice using the CALL system at least twice a week and completed all units within eight weeks. They received two types of feedback—immediate feedback and summative feedback. Immediate feedback was provided in various formats, such as voice waveforms, scores, and model pronunciation. Summative feedback was provided at the end of each unit with a list of mispronounced words and a speaking portfolio. The results showed that only college students significantly improved their pronunciation. Questionnaire analyses revealed that college students found immediate feedback more useful, while seventh graders preferred summative feedback. These findings suggest that seventh graders may have struggled to utilise immediate feedback, such as waveforms, which could explain their lack of improvement. This highlights the importance of immediate feedback that are appropriate for the learners' developmental stages.

Studies have also reported positive effects of ASR-based training at the segmental level. Neri et al. (2006) employed an ASR system to investigate the effects of immediate feedback on immigrants learning L2 Dutch in the Netherlands. Participants received production training under three conditions: CAPT with feedback using ASR, CAPT without feedback, or no CAPT. Results indicated that participants who received ASR-based feedback improved their pronunciation of Dutch phonemes more than those who did not receive feedback.

Liakin and his colleagues conducted studies on the effectiveness of an ASR mobile application to practice French /y/. In Liakin et al. (2013, 2015), L1 English speakers who were beginners in French were divided into three groups: an ASR group, a Non-ASR group, and a control group. Participants in the ASR group practiced producing the French /y/ using Nuance Dragon Dictation for 100 minutes over five weeks. Those in the Non-ASR group practiced the target phoneme with the same words and phrases, but received immediate feedback from a teacher. Meanwhile, participants in the control group met with a teacher weekly to practice speaking French in general. Results revealed that only the ASR group significantly improved their production accuracy of /y/, suggesting that using ASR-based feedback enhanced production learning more efficiently than feedback from a teacher.

Some studies have observed pronunciation improvement only for specific target phonemes. McCrocklin (2019) compared two modalities of pronunciation training: face-to-face instruction and a hybrid of half face-to-face and half online practice using Windows Speech Recognition. Participants were ESL students from various L1 backgrounds, including Chinese, Korean, or Malay. Those in the hybrid group received instruction twice a week and completed homework assignments involving ASR-based practice. The pronunciation instruction was designed based on communicative pronunciation teaching. The target sounds were vowels (/ɛ/–/æ/, /ɑ/–/ʌ/, and /i/–/ɪ/) and consonants (i.e., /ɹ/, /θ/, /ð/, /ʒ/, and /dʒ/), and participants' productions were recorded in a dialogue reading task and a word-pair reading task before and after training. After three weeks of training, while participants in both groups improved their pronunciation of the target phonemes, there were no significant differences between the groups in terms of overall improvement. However, the degree of improvement varied depending on the phoneme.

Likewise, Chen et al. (2020) investigated whether using ASR was effective in improving the distinction of English vowel /i/–/ɪ/ and /ɛ/–/æ/ contrasts. Taiwanese students participated in six sessions of ASR pronunciation training over three weeks where they

practiced reading a short passage, sentences containing the target minimal pairs, and a word list of the minimal pairs. Results revealed that students were able to produce the /ε/–/æ/ contrast more accurately after training, although there was no significant improvement in their production of the /i/–/I/ contrast.

Using the same methodology as Chen et al. (2020), Inceoglu et al. (2020) investigated the effects of using ASR on the pronunciation practice of the English /i/–/I/ and /ε/–/æ/ contrasts with L1 Korean speakers. The results of acoustic analysis revealed that ASR production training was effective only for the vowel /i/, suggesting that the effectiveness of ASR production training varies depending on the difficulty of the vowel pairs. Inceoglu et al. argued that the lack of improvement for the other contrasts might be attributed to the absence of chances to listen to the model sounds, indicating that ASR production training for challenging phonemes might be more beneficial when combined with HPVT, where learners can repeatedly hear various models of target phonemes.

Spring and Tabuchi (2022) incorporated an ASR-based pronunciation practice tool into their English class for L1 Japanese learners of English. Students practiced English vowels and consonants for 15 minutes each week over a semester. In most weeks, students first received instructions on how to pronounce challenging phonemic contrasts for L1 Japanese speakers (e.g., /l–r/, /v–b/, /æ–ɑ/, and /i–I/) and then practised their pronunciation using NatTos. After a semester-long training, learners improved their overall intelligibility of target phonemes, especially vowels. However, none of the consonant contrasts exhibited a significant decrease in the number of errors. Spring and Tabuchi suggested that only vowels showed statistically significant learning outcomes possibly because consonants, except for /l–r/, were already relatively intelligible before the training began.

García et al. (2020) compared the effects of a 15-week ASR-based pronunciation training with teacher-led instruction. Participants were L2 Spanish learners, and the target sounds involved vowels and <g, j, h>. Although both types of training significantly improved pronunciation, ASR training improved comprehensibility, fluency, and confidence ratings significantly more than the teacher-led instruction for the <g, j, h> sounds. However, teacher-led instruction led to a better long-term learning than ASR training. Therefore, the researchers argued that supplementing teacher-led pronunciation instruction with ASR training would lead to the best learning outcomes.

While many studies have shown that phonetic training using ASR systems can benefit L2 pronunciation learning, Hincks (2003) found that ASR phonetic training was beneficial only for learners with heavy accents. In their study, middle-aged immigrants with various L1 backgrounds who were studying English in Sweden participated in pronunciation training using an ASR system. Overall, there was no significant improvement after training. However, when the results were analysed based on the participants' initial accentedness, it was found that only those with strong accents showed improvement. The researcher suggested that the lack of improvement among participants with better pronunciation could be due to the short training period, averaging 12.5 hours, or that mimicry practice alone might not be sufficient as learning progresses. Considering the training took place in 2001 and ASR technology has significantly advanced since then, it would be valuable to replicate the study to assess whether recent training using ASR systems still produce varying results based on learners' initial accentedness.

Lee (2021) reported results that differed from other studies investigating ASR-based pronunciation training. The study examined the effects of mobile-based phonetic training on the English /o/–/ɔ/ contrast with L1 Korean speakers. Participants practiced the target contrast mainly by shadowing model speech for five minutes per day. After a month of training, while participants improved their perceptual accuracy, the improvement in their production accuracy was relatively small, and it did not match the level of improvement observed in their perceptual accuracy.

Studies have explored not only the effectiveness of using ASR technologies for pronunciation learning but also learners' perceptions of this type of pronunciation training. Research has shown that learners find ASR training useful and beneficial for pronunciation learning. For example, Liakin et al. (2017) reported that participants who practiced French /y/ using Nuance Dragon Dictation found the ASR app helpful for improving their pronunciation and enjoyed using it. Similarly, Spring and Tabuchi (2021) found that L1 Japanese learners of English had positive perspectives on ASR-based pronunciation training with NatTos, with many believing that NatTos improved their pronunciation. They also indicated that NatTos was more effective for individual sounds (i.e., vowels and consonants) than for prosody (i.e., stress and rhythm).

Other studies, however, have found lower ratings on the usefulness of ASR, and some even reported negative perceptions. Chen et al. (2020) demonstrated that 55% of Taiwanese

participants rated ASR as a useful tool, none rated it as very useful, and a few expressed negative perceptions. Furthermore, only 22% of participants indicated an intention to use ASR for their future studies. In a similar line, Inceoglu et al. (2020) reported that 68% of L1 Korean participants found ASR helpful after six training sessions on English vowels, while 21% rated ASR below the average. As for the intention to use ASR for future pronunciation practice, only 58% of participants expressed a desire to keep using ASR.

A study conducted by Artieda and Clements (2019) with a large number of participants examined learners' perceptions of using ASR for word-level feedback in pronunciation practice. Their questionnaire survey, completed by 482 participants from China, Vietnam, Saudi Arabia, and Italy, revealed that 95% of the participants found ASR helpful. However, the study also indicated that the usage of ASR features varied depending on the learners' country of origin—Chinese and Vietnamese learners used more the retry option feature than those from other countries.

Research has also reported that ASR-based training can foster students' autonomous learning. McCrocklin (2016) investigated learners' autonomy after a three-week pronunciation workshop. There were three groups: conventional pronunciation training, conventional training including minimal ASR use, and hybrid training using ASR. Results from questionnaire surveys, learning logs, and interviews, indicated that only the groups introduced to ASR demonstrated a significant increase in beliefs about their autonomy after training. Furthermore, the hybrid group spent more time practicing pronunciation using ASR after the workshop compared to the other two groups. This suggests that consistent ASR practice throughout a language course can familiarise students with the tool and encourage them to continue using ASR outside the classroom. Mroz (2018) further supported that using ASR for pronunciation learning helps foster learner autonomy. Additionally, Mroz (2018) found that ASR training improved learners' awareness of intelligibility.

One reason why training using ASR fostered learners' autonomy could be simply because the training is enjoyable, as reported by participants in studies by Hincks (2003) and Moussalli and Cardoso (2016). However, some ASR-based studies revealed that learners expressed frustration towards ASR, claiming that ASR was not advanced enough to accurately detect their pronunciation (Chen et al., 2020; Inceoglu et al., 2020; Liakin et al., 2017; McCrocklin, 2016; Spring & Tabuchi, 2021).

As reviewed above, ASR's ability to recognise human speech is now approaching the level of L1 listeners, allowing L2 learners to receive feedback on their pronunciation. Previous literature has also shown that many L2 learners hold a positive perception of using ASR for pronunciation practice, though some have expressed frustration and doubt about its accuracy. Artieda and Clements (2019) even raised the possibility that perceptions of ASR may vary depending on the learners' country of origin. In Experiment 3 of this thesis, ASR was utilised as a tool for production training to investigate its effectiveness of ASR on the English /l-/r/ contrast with L1 Japanese speakers compared to other training types (i.e., perception-only training, perception-and-production-mixed training, and perception-to-production training). Experiment 3 also explores how learners perceive ASR as a pronunciation learning tool.

2.4.2 Effects of production training on perception

The connection between perception and production abilities in language learning has been discussed in previous sections, particularly in the context of the transfer from perception to production. Although most researchers agree that speech perception generally precedes speech production, some argue that the reverse can also be true. Evidence suggests that production accuracy can sometimes precede perceptual accuracy, as seen in cases where L1 Japanese speakers can accurately produce English /l/ and /r/ despite struggling to perceptually distinguish between them (Goto, 1971; Sheldon & Strange, 1982; Yamada et al., 1994). For example, Kissling (2014) explored the impact of explicit phonetics instruction on the perception and production of L2 Spanish sounds. The phonetics instruction included articulatory explanations, visual aids, and pronunciation practice focused on specific Spanish sounds that are typically challenging for English speakers. The study found that the participants who received pronunciation instruction not only significantly improved their production accuracy but also showed enhanced discrimination accuracy between Spanish and English sounds. This suggests that phonetics instruction can effectively enhance both perception and production in language learning.

Beyond phonetics instruction, research indicates that production training can enhance perceptual accuracy at both the segmental (Catford & Pisoni, 1970; Herd et al., 2013; Kartushina et al., 2015; Lacabex et al., 2008; Linebaugh & Roche, 2015; Thorin, 2020) and suprasegmental levels (Hirata, 2004). Herd et al. (2013) demonstrated that production training using waveform and spectrogram feedback led to improvements in perceptual

accuracy of Spanish intervocalic /r/ and /r/ sounds for American English speakers, supporting the idea that production training can enhance perceptual learning.

Kartushina et al. (2015) also showed a transfer from production learning to perceptual learning. In their study, L1 French speakers were divided into two groups: a treatment group which received immediate feedback, and a control group which did not receive immediate feedback. Following training, only the treatment group significantly improved in both production accuracy and perceptual accuracy. However, it is important to note that there was no significant correlation found between the two modalities.

Similarly, Thorin (2020) found that L1 Dutch speakers improved their perceptual accuracy of English /æ/ and /ε/ sounds through production training with feedback, suggesting that even brief production practice can positively impact perception. Moreover, the transfer effect from production learning to perceptual learning was observed at the suprasegmental level. Hirata (2004), conducted a study in which L1 English learners of Japanese underwent production training focused on Japanese pitch and vowel duration. During training, learners listened to model speech, looked at the contour of the speech, and practised pronouncing it while checking the contour differences between the model speech and their own speech. After training, learners significantly improved not only their production accuracy but also their perceptual accuracy. Interestingly, the study revealed that some learners showed greater improvement in production than in perception, while others exhibited the opposite improvement pattern—better learning in perception.

Furthermore, some studies have shown that production training can be more effective than perceptual training for perceptual learning. One of the earliest studies on this topic conducted by Catford and Pisoni (1970) reported that participants who received production training outperformed those who underwent perceptual training in both perception and production. Similarly, Lacabex et al. (2008) reported on a study with L1 Spanish teenagers who underwent either perceptual training in a discrimination task format, production training with instructor's feedback, or received no training on L2 English vowel reductions over three months. The results showed that both treatment groups significantly improved perceptual accuracy, but the production training group marginally outperformed the perceptual training group.

In a related study, Linebaugh and Roche (2015) also supported a transfer from production learning to perceptual learning. In their study, L1 Arabic speakers received

training on the English /æ/–/ʌ/, /ɜ/–/ɔ/ and /g/–/dʒ/ contrasts in either articulatory training or focused exposure training. Results revealed that while both training groups improved, the production training led to more robust and lasting gains, particularly for the consonants /g/ and /dʒ/, and the vowel pair /æ/–/ʌ/. Thus, these findings indicate that production training can have a more significant impact than perceptual training alone.

However, Sakai (2016) argued that these studies cannot conclusively claim the effects of production training on perception because their participants listened to model sounds of the target phonemes. To address this, Sakai more strictly controlled the design of production-only training by having participants listen to white noise during training sessions, allowing for a comparison between the production-only and perception-only groups. The results showed that the perception-only group improved their perception but not their production. In contrast, the production-only group showed moderate improvement in both perception and production.

On the other hand, some studies revealed no learning transfer from production to perception (Hattori, 2009; Liakin et al., 2013, 2015; Tateishi & Winters, 2015). For instance, Hattori (2009) conducted individual production training on the English /l/ and /r/ contrast with L1 Japanese speakers. Participants watched a video recording of model sounds and practiced producing the target phonemes. After 10 sessions of 30- to 40-minute training, although participants' productions became closer to the L1 speakers' models, their perceptual accuracy did not improve. Similarly, Tateishi and Winters (2015) showed that L1 Japanese speakers who underwent production training on the English /l/ and /r/ contrast using ultrasound imaging did not improve their perceptual accuracy, although they improved their pronunciation of the English /l/ sound.

Additionally, similar findings were observed in the learning of the French /y/ sound by L1 English speakers (Liakin et al., 2013, 2015). These studies found that learners who practiced the pronunciation with ASR improved accuracy after five sessions of 20-minute training. However, this improvement in production did not transfer to perceptual learning.

In contrast, Lee (2021) reported different findings regarding production training. The study investigated the effects of ASR-based pronunciation training using a mobile app with L1 Korean speakers. Results showed that while participants improved their perception of an L2 English vowel contrast, their production did not improve as much as their perception.

Overall, both perception and production training enhance perceptual learning, but production training generally leads to greater improvements in production accuracy. However, because only a limited number of studies have employed production-only training without explicit pronunciation instruction, further research is needed to explore this complex relationship, with a particular focus on production-only training.

2.5 Perception and production-mixed training

Although studies have shown that perceptual training alone can improve L2 learners' production, some studies reported that the gains are limited to around 15% (Bradlow et al., 1997; Flege, 1995b; Shinohara & Iverson, 2018). A possible explanation why the ceiling effect is observed is that HVPT facilitates category formation but not pre-categorical processing (Sadakata & McQueen, 2013). Shinohara and Iverson (2018) combined an identification task, focusing on category representations, and a discrimination task, targeting pre-categorical processing, in an attempt to increase the amount of perceptual gain. After five sessions of 30-minute training, L1 Japanese learners improved their perceptual accuracy of the English /l/–/r/ contrast, and the learning transferred to production learning. However, like in other studies, the improvement in identification accuracy was limited to 15%. Findings from production training studies are in the same line; production training alone does not appear to be adequate to improve both perception and production, as described in the previous section.

These findings on the lack of effectiveness of phonetic training pose a question as to whether combining two modalities by adding speech production to perceptual training can facilitate L2 learners' perception and production effectively. Some phonetic training studies have shown that perceptual learning transfers to production learning (e.g., Bradlow et al., 1999; Bradlow et al., 1997) and vice versa (e.g., Kartushina et al., 2015). These findings suggest that the two modalities positively influence each other's learning. Thus, training the two modalities in one training paradigm, more specifically, adding a production component in perceptual training, could lead to better learning than training in a single modality.

Moreover, adding production to perceptual training is a useful technique to explore the relationship between speech perception and production. Although Sakai (2016) strictly controlled the input of the to-be-learned sounds in production-only training by having participants hear white noise from headphones, it is almost impossible to eliminate any chance of being exposed to target sounds in production-only training because participants

hear their own speech when producing sounds. Sakai noted that some of her participants talked about the training to their families, implying that they produced the target sounds during the training period. Adding production to perceptual training and comparing the effect with perception-only training allows us to explore how the two modalities are related and the interactive effects of perceptual and production learning. Therefore, there has been an increased interest in the effect of perception-production-mixed training on L2 speech acquisition in recent years.

2.5.1 Production effect on vocabulary learning

Before L2 researchers started exploring the effect of perception-production-mixed training, the effect of production had been investigated on L1 vocabulary learning. Previous literature has suggested that reading words aloud helps memorise words more than reading words silently (e.g., Bodner et al., 2014; Fawcett, 2013; Forrin et al., 2012; Gathercole & Conway, 1988; MacLeod et al., 2010).

Gathercole and Conway (1988) explored how production affects vocabulary learning by comparing several learning conditions across four experiments, namely Silent reading, Read aloud, Listen and silent reading, Silent mouthing, Writing, and Invisible writing conditions. In the Writing condition, participants could see their own writing, while they could not in the invisible writing condition. In each experiment, participants learned a set of ten 4-letter words in one of the learning conditions, and all participants learned 30 words in three different conditions. After the learning phase, participants took a recognition test where they judged whether they had learned each word or not. They were also asked to indicate their confidence in their response. Across a series of experiments, Gathercole and Conway found that participants were better at recognising words that they had read aloud, while the read-only words showed the poorest recognition rate. The effects of mouthing and writing, neither visible nor invisible, were reliable. Their final experiment from the same study conducted in a between-subject design also showed similar findings—participants who read words aloud demonstrated the best recognition performance, followed by the Listen and silent reading condition. Gathercole and Conway (1988) concluded that acoustic information facilitated word memorisation, especially when learners orally generated the acoustic information.

MacLeod et al. (2010) coined this phenomenon the ‘production effect’ and further explored it by conducting a series of experiments. In their Experiment 1, participants read 40 real words aloud and another 40 read words silently. The yes/no recognition task results

revealed that participants learned words they read aloud more correctly than the words they read silently. This finding was replicated with nonwords¹ in their Experiment 6. However, their Experiment 2 revealed that the production effect was observed only in the within-subject design but not in a between-subjects design where participants read 80 words either aloud or silently, depending on which training condition they were assigned to. These findings were replicated in their Experiment 3 using a 2AFC recognition task for testing instead of a yes/no recognition task.

MacLeod et al. (2010) suggested that producing words that participants were studying and responding to the words enhanced learning. To test their hypothesis, in their Experiment 4, they had a group of participants press a key only for half the words. Another group of participants said 'yes' instead of pressing a key. The researchers found no production effect for either group, suggesting that the production effect was attributed to producing each word rather than to silent or oral responses. Their Experiment 5 further investigated the source of the production effect to examine whether vocalisation is necessary. Specifically, they used a within-subject design to compare a Silent mouthing condition, where participants mouthed words without articulating the sounds, and a Silent-reading condition. Participants memorised significantly more words that they read in either condition than unstudied words, but they recognised words read in the Silent-mouthing condition significantly better, showing a production effect. That is, mouthing words enhanced word memorisation even without vocalising words.

MacLeod's et al.'s (2010) final experiment investigated whether the production effect was observed because participants were not fully engaging with the task when reading words silently. To test the hypothesis, the researchers had participants answer if the word was living or nonliving to ensure they were focusing on reading each word. The yes/no recognition test results revealed a robust production effect, meaning that participants memorised the words that they read aloud better than the words read silently, and did so not because they only paid attention to the words they needed to read aloud.

Taken together, MacLeod et al.'s (2010) series of experiments found that orally producing or mouthing words, including nonwords, helped word memorisation. However, the

¹ The terms 'nonword' and 'pseudoword' are treated according to their distinct definitions: 'pseudoword' refers to letter strings that are orthographically valid and pronounceable but lack meaning, while 'nonword' refers to random letter strings that are both orthographically and phonologically invalid (Ziegler et al., 1997). While 'pseudoword' might be more appropriate in some contexts, this thesis adopts the term 'nonword' throughout for consistency, following the terminology used in many cited studies.

production effect was observed only in a within-subject condition. Because of the absence of the effect in a between-subjects condition and ruling out other factors, MacLeod et al. (2010) argued that the production effect is caused by the orally produced words being more distinctive compared to the words silently read. The difference in the distinctiveness could have brought about the benefit of production to memorise words.

To further investigate the production effect and the distinctiveness account, Forrin et al.'s (2012) experiments included nonspeech production, such as Spelling aloud, Writing, and Typing. Across all three conditions, participants learned 80 five-to-ten letter-long words, and each condition was compared with a Silent learning counterpart. In the Spelling condition, participants spelled words aloud, while those in the Silent learning counterpart spelled words silently in their mind. All participants completed a yes/no recognition test after a learning phase. Results revealed that all the production conditions demonstrated better learning than the Silent learning condition. There was, however, no significant difference in learning outcomes among the three production conditions. These findings indicated that the source of the production effect was not about acoustic information, but rather, the action of production itself, both acoustic and non-acoustic. To test this hypothesis, Forrin et al. (2012) compared each Writing, Silent mouthing, and Whispering condition to Read aloud and Silent reading conditions. They found that words read aloud were memorised better than the words learned in any other conditions. The Silent reading condition showed the least recognition performance. Thus, Forrin et al. argued that production did not have to be vocally generated, but vocal production enhanced learning the best because speaking brings about distinctiveness in two dimensions, motor and auditory, supporting the distinctiveness account.

However, the meta-analysis conducted by Fawcett (2013) showed that the production effect could also be observed in a between-subject design. This prompted Bodner et al. (2014) to re-examine the production effect in both within-subject and between-subject designs to better understand the source of the production effect. Participants learned 100 words either by reading aloud or silently. The recognition test results revealed that production helped memorise new words even in a between-subject design. Moreover, they found no difference in the recognition of produced items between within-subject and between-subject conditions. Therefore, these findings rejected the distinctiveness account proposed by MacLeod et al. (2010). Instead, Bodner et al. hypothesised that rather than that production enhanced learning in a within-subject design, production masked the learning effects for the words read silently compared to the words read aloud.

Although studies listed above reported a ‘production effect (MacLeod et al., 2010)’, other studies found mixed effects (Abbs et al., 2008) and reverse effects of production on vocabulary learning in L1 (Leach & Samuel, 2007) and in L2 (Dahlen & Caldwell-Harris, 2013; Kaushanskaya & Yoo, 2011). Abbs et al.’s (2008) study with L1 English speakers showed somehow mixed effects of production. Participants learned 12 trisyllabic nonwords based on English phonology by hearing target nonwords together with paired alien pictures. In a Production condition, participants repeated nonwords aloud, while those in a Perception condition did not pronounce nonwords. The results of an immediate recall test showed no significant difference in expressive recall performance between the two conditions. A recall test after a distraction task, however, revealed that participants who repeated nonwords aloud retained them better. Yet, a one-day delayed test found no benefit of production on long-term retention. These mixed findings suggested that production is not necessary to memorise L1 new words, but production might facilitate a better retention of newly learned words.

Abbs et al. discussed that the lack of production effect could be attributed to the unequal amount of exposure because participants in the production condition could hear their own production during the exposure phase. In another study, Abbs et al. (2008) controlled the amount of exposure by giving participants one extra time to hear each nonword. The results were consistent with their previous findings, except for a slight production effect based on power analysis.

Next, Abbs et al. (2008) examined whether the production effect was attenuated because of the semantic component (i.e., paired alien pictures) that participants needed to memorise in addition to words. A nonword repetition task was used for training so that participants focused purely on the phonetic component. In a Production condition, participants listened to a six-syllable nonword and repeated the word immediately after they heard it in each trial, while participants in a Perception condition were instructed to repeat the word in their mind. Results revealed no production effect.

In line with Abbs et al.’s (2008) immediate recall test findings, Leach and Samuel (2007) found negative effects of production. L1 American English speakers learned 12 nonwords with either the /f/ or /s/ sounds. For learning, a letter was presented on a screen, followed by an auditory presentation of a nonword. Participants pushed ‘yes’ or ‘no’ according to whether the letter appeared in the nonword they heard. Those who received training in a Production condition repeated the word after pushing the button. All participants

encountered each item 24 times per day and received training for four days. Results showed that participants who did not repeat nonwords aloud successfully memorised them, but those who repeated them did not show a statistically significant level of perceptual learning.

Next, Leach and Samuel (2007) added semantic information to the training. In the new training sessions, participants learned nonwords together with pictures. Two pictures were presented at the beginning of each trial, and participants listened to a nonword that was associated with one of the pictures. Then, they answered which picture was paired with the word they had heard. The findings agreed with the researchers' previous experiment that showed the null-production effect. Participants who learned without repeating nonwords demonstrated robust learning, while the learning of those who repeated aloud was unreliable.

Studies reviewed so far were based on L1 word learning. Kaushanskaya and Yoo (2011) found that the production effect disappeared when words contained unfamiliar sounds. In their first experiment, 48 novel words were invented using eight L1 English phonemes, and L1 English speakers learned these words along with their meanings. The learning session was conducted in a within-subject design; half the words were learned in a Production condition and the other half in a Silent reading condition. During the learning session, participants listened to a word twice and saw the meaning presented on the screen. Participants repeated the word aloud together with the meaning three times in the Production condition, while in the Silent reading condition, they did the same procedure silently. For testing, participants engaged in a recall task and recognition task right after and one week after the learning session. In the recall task, they named the meaning as they heard a word. In the recognition task, they listened to a word and chose the correct meaning from five options.

Results revealed that participants recalled orally produced words more accurately than heard words right after the learning session but not in the delayed recall tasks. Regarding recognition, participants scored higher for produced words both in immediate and delayed tasks. The researchers hypothesised that producing phonologically familiar words aloud facilitates the integration of new sounds into long-term memory, but does not do so for phonologically unfamiliar words, because articulatory ability is influenced by long-term phonological knowledge (Gathercole & Conway, 1988).

To test their hypothesis, Kaushanskaya and Yoo (2011) examined whether the production effect would be obtained with phonologically unfamiliar words in the same procedure. Contrary to their previous findings, participants recalled and recognised heard

words more accurately than produced words in the immediate task. Moreover, the more unfamiliar sounds a word contained, the less accurately participants recalled it. Kaushanskaya and Yoo (2011) claimed that with the phonologically unfamiliar words, participants could not utilise their L1 phonological system when learning because producing unfamiliar sounds enhanced the realisation of the differences between the sounds in the words they were learning and their L1 phonological system. In contrast, the differences were not highlighted in the Perception condition, facilitating the reliance on their L1 phonological system to memorise new words. The researchers also pointed out that production interfered with memorisation because the auditory stimuli and participants' articulated sounds might have conflicted due to their inaccurate pronunciation of unfamiliar sounds.

The null effect of production on L2 word learning was also observed in Dahlen and Caldwell-Harris (2013). L1 English speakers learned 20 Turkish words in either Production, Production with white noise, Subvocal production, and Articulatory suppression conditions. These conditions were designed to manipulate the availability of input rehearsal in their phonological loop (see more details about Phonological Short-Term Memory in Chapter 2.6). Turkish words containing English phonemes were selected for stimuli, and each word was paired with a picture. In each training trial, participants were auditory and visually presented with a Turkish word three times, followed by a rehearsal task that depended on the assigned condition. Participants in both the Production and Production with white noise conditions repeated the word aloud but the latter group could not hear their own voices because of the white noise. In the Subvocal production condition, participants repeated the stimulus in their mind without articulating it, which was similar to other studies' perception conditions. The researchers also included an Articulatory suppression condition, where participants read aloud a short magazine article right after hearing stimuli to prevent any rehearsal of the stimuli in their mind. After training, participants took a picture naming recall task and recognition task.

Results revealed that participants in the Subvocal production condition performed best in the recall task, followed by the Production condition. Participants in the Production with white noise condition did not recall as much as those in the Production condition did. The performance of those in the Articulation suppression condition was the worst. Regarding the recognition task, the Production and Subvocal production conditions yielded the best learning, followed by the Production with white noise condition. Again, the Articulation suppression condition showed the least learning. These findings suggested that production, regardless of whether it was internal or overt, helps L2 word memorisation for recognition. However,

mentally producing a word without any articulation was more effective for recall. Dahlen and Caldwell-Harris (2013) described that silent production was the most powerful way to memorise words probably because one can rehearse speech in one's mind faster than when articulating it, which allows learners to be exposed to the speech more times. Moreover, articulating speech while trying to rehearse it divides the learner's attention.

Research on child language development also showed a null effect of production. For instance, Zamuner et al. (2017) conducted two experiments to investigate whether production facilitates L1 vocabulary learning with L1 English children aged between four-and-a-half and six years old. In the first experiment, children learned four new CVC words paired with imaginary animal pictures twice per word, resulting in eight trials in total. They learned two words in a Perception condition and the other two in Production condition. In the Perception condition, children heard a word with a paired picture that was visually presented. Two seconds later, they were either prompted to remain silent and heard the stimulus again or orally repeated the stimulus. After training, the children saw two pictures and heard a word that was paired with one of the presented pictures. Eye-tracking data revealed that the children stared at the correct pictures longer for words they learned in the Perception condition than in the Production condition, suggesting that they learned new words better without repeating them aloud.

In the second experiment, Zamuner et al. (2017) further tested the effects of production by doubling the number of words and simplifying the training task. The main difference between Experiment 1 and Experiment 2 was that the Perception condition and Production condition were separated rather than randomised. First, children learned four words either in the Perception condition or in the Production condition. Next, they learned another set of four words in the other training condition. In addition to the recognition test, the children also took a recall test, in which they named the words they remembered after training. The results showed that children looked at the paired pictures more correctly and faster for the words learned in the Perception condition than in the Production condition. In the recall test, they remembered more words they learned in the Perception condition than in the Production condition. These findings indicated that producing words attenuated young children's L1 vocabulary learning.

Zamuner et al. (2017) labelled this counter effect the 'reverse production effect' on the contrary to the 'production effect' coined by MacLeod et al., (2010). They also raised two

circumstances where the production effect can be attenuated. One is when the to-be-learned items are composed of foreign or infrequent sounds. The other is when production is embedded in another task. Vihman et al. (2014) explained that young children tend to learn words with well-known meanings and/or sounds because it poses fewer cognitive loads and frees up resources for sound-meaning mapping. In contrast, processing unfamiliar phonological words takes more cognitive resources, leading to interference in learning. Zamuner et al. (2017) hypothesised that, in their experiments, learning a new word and articulating the word simultaneously was cognitively demanding for children, possibly leading to the disruption of learning. Since learners can better learn in less demanding tasks with familiar sounds, the researchers hypothesised that it might be recommended to first engage in perceptual learning before moving on to production, ‘freeing up resources for the creation of phonological, lexical, and semantic representations (p.10).’

In sum, most vocabulary learning research has shown that production facilitates memorisation for words that contain familiar L1 sounds. However, when learners memorise words with unfamiliar L2 sounds, the production effect is attenuated, and learning can even be disrupted. Also, some researchers analysed that the production effect might be related to how high the cognitive cost of training is.

2.5.2 Effects on perceptual learning

As reviewed above, L1 vocabulary studies suggested that the production effect would not be observed when processing words containing unfamiliar sounds. Similarly, studies on L2 perceptual learning have shown that production during perceptual training does not bring about extra benefits to improving L2 phonetic perception ability at the segmental level (Thorin et al., 2018) and suprasegmental level (Lu et al., 2015), except for one study that measured perceptual learning using neurophysiological measures and found a positive production effect (Thorin, 2020). Other studies, however, found that producing target sounds in perceptual training may disrupt L2 perceptual learning at the segmental level (Baese-Berk, 2010, 2019; Baese-Berk & Samuel, 2016) and suprasegmental level (Wright, 2021).

Baese-Berk (2010) conducted three experiments to investigate the effects of producing target phonemes in an incidental perceptual training regimen. L1 American English speakers received Perception-only training in the first experiment and Perception+production training in the second experiment. Target sounds were synthesised pre-voiced /d/ and short-lag /t/. In Perception-only training, participants listened to a stimulus, saw a paired picture, and pressed

a button to move on to the subsequent trial. In Perception+production training, they listened to a stimulus, saw a paired picture, imitated the stimulus, and pressed a button to play the next stimulus. Both types of training consisted of 768 tokens in total over a two-day period with one hour of training and 384 tokens each day. The results of AX discrimination tests and the 6-alternative forced choice categorisation tests showed that participants who received Perception-only training learned the distinction between the two sounds, and the post-categorisation accuracy was 76.97%. Those who repeated the stimuli aloud in perceptual training failed to learn discrimination of the target sounds, and the post-categorisation accuracy was 64.09%.

In Baese-Berk's third experiment, Perception-only training and Perception+production training sessions were conducted using the same stimuli with two new groups of participants from the same population to investigate whether disruption of perceptual learning alleviates by adding one more day for training. The results showed that the Perception+production training group caught up with the Perception-only group; no significant difference in discrimination performance was observed between the two groups. These findings indicate that the third day of training alleviated the disruption to perceptual learning. In the categorisation tests, however, the Perception-only training group still outperformed the Perception+production training group even after the third day, suggesting that adding production to perceptual training disrupts perceptual learning, but it can be partially alleviated if learners keep being exposed to the to-be-learned sounds longer.

Baese-Berk and Samuel (2016) replicated Baese-Berk (2010) using different phonemes, the Basque /s/-/ʃ/, with L1 Spanish speakers. In Experiment 1, participants received either Perception-only training or Perception+production training. The training sessions were composed of 390 trials in total and were conducted on two days. The Perception-only training was conducted in an ABX task format, and immediate feedback was given, unlike in Baese-Berk (2010). The Perception+production training also employed an ABX task format, but the participants repeated the final stimulus aloud right *after* hearing it (i.e., like shadowing) and before making a perceptual judgment. The results showed that the Perception-only group learned the distinction between the two sounds, but not the Perception+production group. This result supported the findings of Baese-Berk (2010) that producing target sounds disrupts perceptual learning.

Baese-Berk and Samuel (2016) conducted Experiment 2 to investigate whether learners' prior exposure to a target language alleviates the disruption. The research design was

the same as Experiment 1, except that participants were intermediate learners of Basque. The result of Perception-only training replicated that of Experiment 1. Perception+production training showed a slight improvement in distinction, but there was no significant difference between the results of Experiments 1 and 2. This suggested that prior exposure to the target language might alleviate the disruption of perceptual learning by producing the target sounds, but the degree of alleviation was not significant.

One possible source of the negative effect of the production training is that participants had to listen to their inaccurate L2 speech during the training, which disrupted L2 perceptual learning. To access this hypothesis, Baese-Berk and Samuel (2016) conducted another experiment in which they examined whether the disruption was caused by producing an inaccurate version of the target sounds or articulation itself caused the disruption. The participants were recruited from the same population as Experiment 1. They completed the same ABX task with the same stimuli, but they produced a letter visually shown on the screen instead of repeating the stimuli. Results revealed that the degree of their perceptual learning was between that of the Perception-only training and the Perception+production training groups from Experiment 1. The researchers concluded that producing sounds itself disrupted perceptual learning, but the disruption was more severe when producing target sounds.

Thorin et al. (2018) proposed that the Naming-a-letter condition posed a lighter cognitive load than the Repeating-the-stimulus condition in Baese-Berk and Samuel (2016), and that the disparity of cognitive load between the two conditions might have, therefore, brought less disruption of perceptual learning for the Naming-a-letter condition. To further investigate the effects of producing sounds in perception training, Thorin et al. (2018) conducted perception training on the English /æ/–/ε/ with L1 Dutch speakers, whose English proficiency was the upper-intermediate to advanced levels and compared two training conditions—a Related production condition and Unrelated production condition—to make sure the degree of cognitive load was equalised. Participants listened to a series of words and chose which word they heard last from two visually presented words on the screen. After immediate feedback, a word was presented on the screen, and they read the word aloud. In the Related production condition, they read aloud the last word they heard in each trial and a similar word that did not contain the target phonemes in the Unrelated production condition. Participants received four training sessions over ten days, each consisting of 200 trials (40 trials x 5 blocks). The results revealed that their perception training significantly improved

their identification scores—no disruption was observed. The improvement was generalised to novel speakers and novel words except for those starting with a novel consonant. With regard to the two conditions, there was no difference. These findings suggested that production practice during perceptual training does not bring about extra benefits in perceptual learning.

Thorin (2020) further analysed the behavioural data of perception and production from Thorin et al. (2018) using the auditory mismatch negativity (MMN), a fine-grained tool ‘in measuring automatic auditory change detection even in the absence of attention (p. 50).’ Results of the MMN data revealed that only the Related production condition showed an MMN response in the post-test passive listening task, where participants heard words containing to-be-learned minimal pairs while focusing on watching a silent movie. This indicates that repeating stimuli during perceptual training might lead to better perceptual learning than perception-only training.

Thorin et al. (2018) pointed out that the negative effect that Baese-Berk and Samuel (2016) observed may have stemmed from the training order of perception and production. In the Perception+production training, participants listened to three words and repeated the last word aloud *before* answering which of the first or second orally presented words matched the last word for the discrimination task. This means that they had to store the phonetic information of the orally presented words in memory *before* completing the discrimination task. This may have potentially placed an increased cognitive load on the participants’ short-term memory and negatively affected their performance on the discrimination task.

However, Baese-Berk (2019) rejected the possibility that disruption is attributed to the training order of perception and production. Instead, she explained that the disruption could be due to the dual nature of task involving both perception and production of training. To further explore what had caused the disruption to perceptual learning, she conducted two experiments on the vowels in the prevoiced and short-lag stops with L1 American English speakers. The participants took training in either a Perception-only or a Perception+production condition in an implicit learning paradigm, where participants learned sounds with the help of paired presented pictures. In the Perception-only training, a token was auditorily presented with a paired picture to participants, and they pressed a button to proceed to the subsequent trial. The procedure of the Perception+production training was identical to that of the Perception-only training, except that participants repeated the stimulus before pressing the proceeding button. Participants were presented with 384 trials each day, with

participants in Experiments 1 and 2 receiving training for two and three days, respectively. Participants' perception was assessed at pre- and post-tests with a discrimination task.

Experiment 1 found that although both groups showed improvement after the 2-day training, the Perception-only group learned the distinction significantly more than the Perception+production group. In contrast, in Experiment 2, the Perception+production training group began to show an improvement in discriminating the target contrasts after the third day of training. However, there were individual differences in learning progress between perception and production; some participants in the Perception+production group attained robust learning in perception. Baese-Berk (2019) hypothesised that the disruption might have been attributed to the nature of the dual-task training regime. In combined training, participants switch tasks between perception and production, which means the processing of production begins right after or even during perceptual learning. If the two modalities share the same processing resources, it is possible that perceptual learning was interfered with by switching to the next task before completing the process of perceptual learning.

To test this hypothesis, Baese-Berk and Samuel (2022) conducted two experiments and investigated whether the disruption was caused by 'switching costs' that were built up when switching between perception and production tasks in each trial. In their first experiment, they developed a new training design with a lower switching cost by creating a pause between the tasks so that participants could prepare for the next task. L1 Spanish speakers received training on the /sa-/ /ja/ contrast in either a Short-delay condition or a Long-delay condition. The data of the Perception-only (No-delay) condition and the Perception+production (No-delay) condition from Baese-Berk and Samuel (2016) were also included in the analysis. The stimuli and basic procedures were the same as those in Baese-Berk and Samuel (2016). Participants in the No-delay group repeated the last stimulus right *after* they heard it. Conversely, participants in the Short-delay and Long-delay groups were prompted to produce the stimulus 500 to 3500 ms and 2500 to 5500 ms after they heard, respectively.

The results showed that the participants in the Short-delay and Long-delay groups became able to discriminate the target contrast as well as the Perception-only group, while the No-delay group did not. Moreover, those in the Long-delay group showed significantly better discrimination than the Short-delay group. These findings suggest that a delay before repeating the stimulus alleviates disruption in perceptual learning. Baese-Berk and Samuel

claimed that a 4-second delay allowed participants to finish their perceptual learning and successfully switch to the next production task.

In Experiment 3, Baese-Berk and Samuel's (2016) tested whether production itself hindered perceptual learning by conducting training where participants produced a visually presented letter instead of repeating the stimulus. They found that although the Letter-naming training was significantly more effective than the training that required repetition of stimuli, Perception-only training was the most effective. To untangle this complex matter, Baese-Berk and Samuel's (2022) second experiment investigated the impact of production and dual-task nature on perceptual learning by changing the second task from producing stimuli to non-production button pushing.

L1 Spanish speakers underwent training either in the Button push condition or Delayed button push condition. For analysis, the data of the Perception-only (No-delay), Perception+production (No-delay) condition, and Spoken letter (No-delay) condition from Baese-Berk and Samuel (2016) were also included. During training, participants in the Button push groups pushed the button of the letter that was shown on the screen when the last stimulus was playing. Those in the Delayed button push group pushed the button 500 to 3500 ms after they heard the last stimulus, as in the Short-delayed condition in Experiment 1. Results revealed that all groups involving the second 'distractor' task showed significantly smaller learning than the Perception-only group, while all groups with the letter-naming task showed a small improvement.

These findings suggested that the combination of switching tasks (from speech to production) and the use of target sounds in the second task affected perceptual learning; therefore, delaying the repetition of the stimulus promoted perceptual learning. Thus, Baese-Berk and Samuel (2022) explained that the disruption of perceptual learning was caused by the increased cognitive load of switching tasks, particularly when the second task involved speech production. Regarding the null effect of delay observed in the Button push condition, the researchers postulated that this was attributed to the inability to access the stored sound through production.

The disruption associated with producing target sounds has also been observed in perceptual learning of suprasegmental features. Wright (2021) conducted a one-day incidental perceptual training with 192 trials on four Thai tones with L1 English speakers. He compared the following three training conditions: Perception-only, Familiar-syllable production, and

Unfamiliar-syllable production conditions. In the Perception-only and Familiar-syllable production conditions, the tones appeared in the /ma/ syllable during training, and in the Unfamiliar-syllable production condition, the /mu/ syllable was used to test whether phonetic familiarity affects training outcomes. The stimuli were recorded by an L1 Thai female speaker. During training, participants in all conditions heard five tokens with the same tone and clicked the box that contained X out of four empty boxes. Those in the Perception-only condition proceeded to the next trial after clicking X, whereas those in the Production conditions were prompted to record their imitation of the stimulus they heard before moving on to the next trial.

Results revealed that participants who underwent training in the Perception-only condition showed the most robust learning to untrained tokens. The two production conditions also demonstrated learning, but the degrees were smaller than those of the Perception-only condition. Reaction times further supported these results. Indeed, reaction times in the Perception-only condition became faster during the course of training, which was the fastest among the three conditions, followed by the Familiar-syllable condition. Reaction times in the Unfamiliar-syllable condition did not become faster during training. Findings also revealed that the learning in the Perception-only condition generalised to Untrained tokens, but none of the conditions exhibited learning generalisation to new voices.

These findings suggested that extra production practices during perceptual training did not lead to better perceptual learning. Also, the familiarity of the syllable in which to-be-learned tones appeared did not affect perceptual learning. Write (2021) argued that disruption in perceptual learning had been potentially caused because participants could not pay much attention to audio-visual reinforcement, where they learned which tone was associated with which box as they started preparing to produce the stimulus. Furthermore, he pointed out that repeating the stimulus right after perceptual learning in each trial interfered with storing and processing the stimulus in the phonological loop, thus diminishing the chances to learn the target tones. This point is supported by the findings in Baese-Berk and Samuel (2022), which demonstrated that delaying production during training alleviated the disruption in perceptual learning.

In contrast, the study on the suprasegmental feature conducted by Lu et al. (2015) showed a null effect of production but not disruption. They conducted Perception-plus-production training on the perception of Mandarin tones with L1 American English speakers and compared the training effect with that of Perception-only training. The target

pronunciation items were three Mandarin tones: T1 (high-level), T2 (high-rising), and T4 (high-falling). In both groups, feedback on their perceptual answers was given as well as both the replay of the first stimulus and the visual image of the tone. After listening to the replay while looking at the image, only those in the Perception-plus-production training group imitated the stimulus. Then, the second stimulus was replayed with the visual image of the tone, and they were prompted to imitate the stimulus again. The participants in the Perception group were asked to utter 'next' instead of imitating the stimulus. The training session consisted of 192 trials and lasted about an hour. The results of AX discrimination tasks showed that participants in both groups significantly improved their tone discrimination accuracy, and the learning was generalised to untrained stimuli. There was no significant interaction between the groups, suggesting that repeating stimuli did not bring about an additional benefit or disrupt perceptual learning.

Although these studies have observed either disruption or null-effects of production during training, studies conducted by Iino and his colleagues (2020) found positive effects of repeating stimuli aloud in perceptual learning. They asked Japanese learners of English to repeat each stimulus aloud immediately after hearing it and before selecting an answer during perceptual training on the English /l/ and /r/ sounds. The treatment lasted ten weeks, with three training sessions, each containing 200 items. Additionally, participants watched videos of an L1 English speaker producing the target phonemes in the first week. The results showed significant improvement in perception from pre- to post-tests. Also, the degree of improvement was larger than perception-only training by Iino and Thomson (2018). The findings suggested that repeating stimuli aloud during perceptual training facilitates improvement in perception.

The studies mentioned until now all implemented a mix-training design whereby participants perceive and produce target sounds in each trial. It is worth now to examine studies that have investigated the production effect by having the same group of participants undergo perceptual training and production training separately. Herd et al. (2013) conducted combined training by having participants engage in both perception and production training, but in separate sessions. In their study, L1 American English learners of Spanish at the intermediate level underwent Perceptual training on the Spanish intervocalic /d-r-r/ sounds. The learners were divided into four groups: Perception-only training, Production-only training, Combined training, and Control groups. Training stimuli were composed of three

contrasts /r-r/, /d-r/, and /r-d/ recorded by six L1 Spanish speakers. Ninety minimal pairs were created using both real words and nonwords.

Perception-only training was conducted in a 2AFC identification task format. Feedback was provided in each trial, and the stimulus was replayed regardless of whether the answer was correct or not. Each training session lasted 20 to 30 minutes, and all participants completed six sessions. In Production training, immediate feedback was given using Praat for visual images such as waveform and spectrogram of the model speech recorded by L1 speakers. In each trial, participants first inspected an L1 speaker's production of a word, produced the same word, and compared the waveform and spectrogram of their own speech with those of an L1 speaker. Each training session lasted 35 to 45 minutes, except for the first one which took over an hour due to a tutorial on identifying the target phonemes with the visual images. Combined training comprised three perception training sessions and three production training sessions, switching from perception to production training in different sessions. During pre- and post-tests, perception data were collected in a 2AFC test.

Overall, Perception-only training and Production-only training groups showed significantly more improvement in perceptual accuracy than the Control group, but the Combined training group did not. When analysed separately for each contrast, the Production-only training group significantly improved the identification of the /r/-r/ contrast, and the Perception-only training group improved the /d/-r/ contrast. The Production-only training group exhibited learning generalisation to both new speakers and words with the /r/-r/ contrast, while the Perception-only training group only did to new words. Herd et al. assumed that the Combined training did not make gains in perception because they received half the perceptual training compared to the Perception-only training group.

In a similar context, Fujiyuki et al. (2017) examined the order effect of perception and production effect. L1 Japanese speakers received training on English /l/ and /r/ for five to six days. Participants were divided into three groups: Perception-to-production, Production-to-perception, and an Alternating group. Perceptual training was conducted in a 2AFC task format with 49 minimal pairs spoken by five L1 American English speakers. Each training was composed of 980 trials and lasted 75 minutes. Participants underwent six sessions in total. Production training consisted of a repetition task and translation-naming task. Participants received 980 trials in total over two days. In the Perception-to-production group, participants received perceptual training on the first three days, followed by two days of production training. In the Production-to-perception group, participants completed production

training for two days first, then took three-day perceptual training. Those in the Alternating group started with perceptual training and ended with production training on each training day. In addition to pre- and post-tests on perceptual accuracy, a mid-test was conducted before starting training in the other modality (e.g., production training in the Perception-to-production group). In the Alternating group, participants completed a mid-test on the fourth day of their training.

Results showed that participants in the Perception-to-production group improved their /l/ and /r/ comprehension significantly more than those in the other two groups after perceptual training at mid-test, reaching 77.8% accuracy. However, the accuracy score did not increase after production training. The other two training groups demonstrated similar learning trajectories—participants gradually increased their perceptual score and reached 73.1% accuracy in the Production-to-perception group and 72.1% in the Alternating group. There was no significant difference among the three groups at post-test.

The study on the English vowels /ɑ/, /ʌ/, and /ɔ/ with L1 Korean speakers conducted by Hyunsu and Joo-Kyeong (2022) showed contradictory findings. Participants were divided into three groups: Perception-to-production, Production-to-perception, and Control. The two training groups received nine training sessions over ten weeks. On each training day, participants underwent training in both modalities. Hyunsu and Joo-Kyeong's (2022) Perception-to-production group was, therefore, similar to the Alternating group from Fujiyuki et al. (2017). Perceptual training was conducted in a 2AFC task format, and each session was composed of 120 trials. In production training, participants received pronunciation instruction in a small group where they repeated the model pronunciation and practised pronunciation in minimal pairs. Participants produced 30 to 40 words in each session, lasting 15 to 20 minutes. The results of pre- and post-identification tests showed that participants in both training groups improved their identification accuracy significantly more than the Control group, and the perceptual learning generalised to untrained words. Taken together with Fujiyuki et al.'s study (2017), the order effect demonstrated no statistically significant impact on perceptual learning. However, it is still unknown whether alternating training (i.e., perception-to-production on each training day) improves perceptual accuracy more than perception-only training.

In summary, the literature on perception and production mixed training studies suggests that production practice during perceptual training does not bring about extra benefits on perceptual learning in many cases in general, while Iino et al. (2020) found a

positive effect of production on perceptual learning. Although Thorin (2020) found a positive production effect on perceptual learning based on the neurophysiological measures, behavioural measures (i.e., identification test) did not show evidence of the positive production effect.

2.5.3 Effects on production learning

As reviewed above, production during perceptual training can either cause disruption or does not yield any further benefits to perceptual learning. In contrast, previous studies reported that production and perception mixed training was as effective as perception-only training on production learning despite the absence of benefits for perceptual learning (Thorin et al., 2018). Further, some studies found that the degree of learning from perception and production mixed training was larger than perception-only training (Baese-Berk, 2010, 2019) and production-only training (Herd et al., 2013).

Thorin et al. (2018), who investigated the effect of production during perceptual training on perceptual learning, also examined the effects on production learning. The pre-test reading task consisted of trained words and the post-test contained trained and untrained words. The recordings obtained in the reading task were analysed using an ASR system, and acoustic analysis was also performed. Results revealed that participants in both groups—Perception-only training and Perception-and-production-mixed training—attained more target-like pronunciation, and that there was no significant difference in training conditions. Acoustic analyses found that the distance between F1 and F2 became significantly larger after training. As for the generalisation to untrained words, while the distance for novel words was significantly larger than the trained words of the pre-test, it was significantly smaller than the trained words of the post-test. The durational distinction also became significantly larger at post-test. The analysis using an ASR system also confirmed the improvement in production. Although both modalities significantly improved, there was no correlation between the degree of learning in perception and that of learning in production.

In contrast to Thorin et al. (2018), Baese-Berk (2010) found a more significant production improvement by learners who received the Perception+production training than those who underwent the Perception-only training. L1 American English speakers received perceptual training on pre-voiced /d/ and short-lag /t/ either in the Perception-only or Perception+production training. The production data were collected in repetition and naming tasks and analysed acoustically. After training, participants in the Perception-only training

group produced the two target sounds more distinctively from each other than before training in the repetition tasks. Although the Perception+production training group did not show perceptual learning, they exhibited production learning in the repetition tasks, and the learning was more robust than that of the Perception-only group. Neither of the training groups showed production changes in the naming tasks.

Baese-Berk (2010) extended these two types of training by adding another training day to see if any changes can be observed after extra training. The participants in the Perception+production training group produced the two sounds more distinctively on Day 3 than on Day 2, but the Perception-only training group did not show significant changes in production performance. Her research replicated some previous studies that the Perception-only training was effective both on perception and production to some extent. In contrast, she found that producing stimuli disrupted perceptual learning, but the oral repetition of the stimuli enhanced production learning.

These findings from Baese-Berk (2010) were further supported by Baese-Berk (2019). She also collected production data using a repetition test, and the collected productions were acoustically analysed based on the voice onset time and the duration of the whole word. After the two-day training, only the Perception+production group exhibited a robust improvement in production. The Perception-only group started learning the distinction between the contrast, but the learning by the Perception+production group was significantly more robust. In addition, she analysed the correlation between perceptual learning and production learning per group. The Perception-only group showed that two modalities were related, while the Perception+production group did not.

The findings by Herd et al. (2013) further revealed that combined training was more effective than production-only training. Participants' production data were collected in a word-list reading task and judged by eight L1 Spanish speakers in a 3AFC identification task format. They were also analysed based on the waveform and spectrogram on Praat. The results revealed that the Combined training group and the Production-only group showed a significant increase in their judgment scores by L1 listeners overall. Analysing each contrast separately, the three experimental groups (i.e., the Perception-only group, Production-only group, and Combined group), but not the Control group, increased their score for only the Spanish trill /r/. With regard to the results of acoustic analysis, participants in the Combined training group produced the trill /r/ significantly more closely to L1 speakers' production than those in the Control group. Surprisingly, the Combined training group exhibited more

considerable gains than the Production-only training group, where participants practised producing the target sounds in all the six sessions. These findings led Herd et al. to argue that training both modalities might be essential to improve production.

The study by Hyunsu and Joo-Keaong (2022), who investigated the order effect of perception and production training on perceptual learning, also examined the order effect on production. During pre- and post-tests, all participants read aloud a list of words in a carrier sentence as a production test. Three L1 English speakers judged participants' speech in the 2AFC task format. Contrary to the results of the perceptual learning, the Production-to-perception training group showed a significantly larger improvement in pronunciation accuracy than the Perception-to-production group, suggesting that receiving production training before perceptual training is more effective than the reverse order. However, it is important to note that the production training involved explicit instructions and participants were asked to listen to an L1 speaker's model before repeating the stimuli. Therefore, it is still unclear which led to these findings: production training preceding perceptual training or receiving explicit instructions before being exposed to target sounds.

Overall, studies on L1 and L2 vocabulary learning and L2 phonetic acquisition demonstrated mixed findings. Therefore, it is still unclear (1) whether production in perceptual training causes disruption in perceptual learning and (2) whether perception and production mixed training enhances production learning more than training on a single modality (either perception or production). Examining these issues is theoretically and empirically important, given that they may show how closely production and perception are interrelated and how effective training the two speech modalities in the same time course are for L2 phonetic acquisition.

2.6 Phonological Short-Term Memory (PSTM)

Several variables influence phonetic learning outcomes, with individual differences playing a crucial role. Among these, PSTM has recently emerged as a significant factor in phonetic learning research. PSTM is a component of the Working Memory model initially proposed by Baddeley and Hitch (1974) and further refined by Baddeley (1992). The model comprises the central executive, which is the main component that coordinates resources, and two subcomponents: the visuospatial sketch pad and the phonological loop. The visuospatial sketch pad is responsible for processing visual images, while the phonological loop handles

speech-based information. PSTM is directly associated with the phonological loop, emphasizing its role in processing and retaining verbal information.

Within the phonological loop, there are two parts: a phonological store and an articulatory control process (rehearsal mechanism). Perceived phonological information sent from the central executive is stored in the phonological store, which can retain information for a few seconds. Subvocal repetition is necessary to retain the information for a longer duration; otherwise, it decays. As rehearsing sounds allows for holding phonological information in the phonological loop, the interruption of rehearsal decays the information, leading to the drastic reduction of memorising the new information (Baddeley et al., 1975). Indeed, Dahlen and Caldwell-Harris (2013) found that participants who learned words by rehearsing them recalled and recognised them better than those whose rehearsal of the words was interrupted during the learning session. Studies also found that the phonological loop can store phonologically familiar items more than phonologically unfamiliar items (Gathercole & Baddeley, 1990; Storkel, 2001), suggesting that L2 learners can store new words with sounds that they can correctly perceive better than words with sounds they have difficulty perceiving.

According to Baddeley's Working Memory model, the information stored in the phonological loop is eventually connected to the previous knowledge via the 'episodic buffer' and eventually stored in the long-term memory (Baddeley, 2000). Some studies showed that covert rehearsal can transport new information from the phonological loop to the longer-term memory better than overt rehearsal, suggesting better long-term retention (Dahlen & Caldwell-Harris, 2013; Kaushanskaya & Yoo, 2011). Baddeley et al. (1988) found that PSTM plays a crucial role in successful long-term linguistic learning for unfamiliar items, such as nonwords or foreign words, but not for associations between familiar items.

Studies have reported that PSTM capacity expands from infancy to early adolescence (Gathercole et al., 2004). Starting around the age of seven, children begin to use subvocal rehearsal processes and the phonological loop similarly to adults (Gathercole, 1998). In contrast, Ottem et al. (2007) argued that PSTM capacity is fully developed by age four and does not increase with age. Rather, it increases along with the improvement in linguistic abilities. In either case, PSTM capacity does not change after a certain age. Indeed, French and O'Brien (2008) support these findings. Their study with 11-year-old French-speaking children revealed that the children's performance in a PSTM task did not expand over an intensive five-month English course when the nonwords in the task were created based on the Arabic language. However, their performance improved with English nonwords after the

course. These findings indicate that an individual's PSTM capacity remains the same after adolescence, but their performance in the PSTM task can improve as they learn the language of nonwords used.

2.6.1 Relationship between PSTM and perception

Previous research has found that PSTM is closely related to vocabulary learning (Gathercole & Baddeley, 1990; Speciale et al., 2004). Some studies also showed that PSTM influences learning in L2 grammar (French & O'Brien, 2008), L2 speech fluency (O'Brien et al., 2007), L2 speech production (O'Brien et al., 2006), L2 lexical encoding (Daidone & Darcy, 2021), and L2 listening (Kondo, 2019). More relevant to the present study, studies have found that PSTM can predict L2 speech perceptual learning of English consonants (Lengeris & Nicolaidis, 2014; MacKay et al., 2001), English vowels, (Cerviño-Povedano & Mora, 2011, 2015; Cerviño-Povedano et al., 2011; Safronova, 2016), French nasal vowels (Inceoglu, 2019), and suprasegmental features (Darcy et al., 2015).

MacKay, Meador, and Flege (2001) investigated the relationship between PSTM and L2 perceptual accuracy in a naturalistic learning setting. Participants were L1 Italians who immigrated to Canada and had been exposed to English for at least 18 years. Their age of arrival varied between seven and nineteen. Their perceptual accuracy was tested in a 5-alternative forced-choice identification task consisting of stimuli of 19 English consonants embedded in noise. To assess PSTM, participants were given a nonword repetition test composed of nonwords based on Italian CV syllables. The number of syllables ranged from two to five. Results revealed significant negative correlations between the PSTM scores and the error percentage of the identification of L2 English consonants. The researchers argued that L2 learners with a bigger PSTM capacity might encode L2 phonemes and store them in long-term memory more successfully.

Lengeris and Nicolaidis (2014) replicated MacKay, Meador, and Flege (2001) and found positive correlations between PSTM and L2 perception of English consonants with L1 Greek speakers. The participants' English proficiency was Cambridge FCE to CPE (B2 to C2 level in the CEFR). An identification test was conducted in three conditions: a quiet background, a competing speaker where another speaker's voice competed with the target speech, and an 8-speaker babble (the latter two were noise conditions). PSTM was measured using a serial nonword recognition (SNWR) task in Greek. The stimuli were composed of eight pairs of sequences for five, six, and seven CV syllables respectively. In the task,

participants listened to two sets of stimuli and answered whether the order of the two nonwords was the same. It was found that correlations between PSTM and identification accuracy in all three conditions, suggesting that PSTM can become a predictor of L2 perception in a realistic listening environment.

Cerviño-Povedano and Mora (2011) investigated the relationship between PSTM and Catalan-dominant Catalan-Spanish bilingual's perception of the English /i:/-/ɪ/ contrast and the impact of linguistic knowledge on the PSTM task performance. Their perceptual accuracy was tested in an identification task format. PSTM was measured using an SNWR test with 24 pairs of CVC syllabic nonword sets stimuli. The test included sets of nonwords, each containing between five and seven items, and featured three different types of nonwords: Catalan, English and Russian. The researchers used a serial nonword recognition task over a nonword repetition task to minimise the effects of articulatory demands, which might obstruct one's task performance, and lexical influences. Results showed that the higher the PSTM score was, the more target-like their perception was. There was a significant main effect for the test language (Russian > Catalan > English), but the test scores between the three languages were correlated. This finding suggests that although languages affected the performance of PSTM tasks, there were not many individual differences in PSTM across test languages.

Cerviño-Povedano, Mora, and Aliaga-García (2011) further investigated the relationship between PSTM and perceptual accuracy and its language effect with bilingual speakers. They conducted SNWR tasks in Catalan and Russian nonwords with Catalan-Spanish bilinguals of high and low Catalan use. The number of nonwords varied from five to seven. Results revealed that the SNWR score was moderately correlated with the percentage of Catalan use on the SNWR task in Catalan, and the Catalan-dominant speakers got higher scores than the Spanish-dominant speakers. On the Russian SNWR, however, there was no difference between the two populations, but the Catalan-dominant performed better when the number of nonwords was seven. The authors hypothesised that Catalan-dominant recalled nonwords of an unknown language better because Catalan's larger vowel system had the advantage of processing Russian sounds. With regard to the relationship between the SNWR scores and L2 perceptual accuracy, both tasks showed that the larger the PSTM was, the more correctly they identified L2 sounds. Thus, they suggested that language dominance played a role in the SNWR task performance for bilinguals. To minimise the effects of language dominance, using an unknown language might be more reliable. Cerviño-Povedano and Mora

(2015) partially replicated their study and confirmed the relationship between L2 perception of the English /i:/-/ɪ/ contrast and PSTM by using Danish nonwords in an SNWR task with L1 Spanish speakers.

Safronova (2016) looked at the relationship between PSTM and L2 perception as part of a study exploring the influences of cognitive abilities on L2 perception. The participants were Catalan-Spanish bilinguals whose English proficiency was low-intermediate to advanced. To measure their perceptual accuracy of five American English vowel contrasts, a perceptual assimilation task and an ABX task were conducted. For PSTM, an SNWR in the same design as one in Cerviño-Povedano and Mora (2011) was used. Results showed that PSTM significantly correlated with L2 sound categorisation and discrimination accuracy.

The positive association of PSTM and L2 perception is also replicated by French nasal vowels with L1 English speakers. In Inceoglu (2019), participants were L1 Australian English speakers, and their French proficiency varied from A2 to C1 in the CEFR. Participants completed a perception test in audio-only, audio-visual and visual-only modalities. In the audio-only and audio-visual modalities, they saw a recorded image of a speaker's head on the screen. In every modality, they chose a vowel out of the three vowels after each stimulus. To assess their PSTM, they took a nonword repetition task consisting of 16 pairs of English nonwords. The number of nonwords increased from three to eight. They were asked to repeat the nonwords two seconds later in each trial. The mean percentage of the PSTM test was 54.09%. Results of linear regression analysis showed that the PSTM score was a significant predictor of L2 vowel perception in all three conditions.

PSTM may be linked to L2 speech perception not only at the segmental level but also the suprasegmental level. Darcy, Park, and Yang (2015) explored individual differences in L2 phonological perception. In their study, L1 Korean speakers in the US completed forward digit, backward digit, forward nonword, and backward nonword recall tasks in English and Korean to measure their PSTM. In those tasks, they verbally repeated the stimuli. They also had a sentence repetition test in English and Korean and a paired-associated learning task in Korean. To measure their phonological processing ability, participants underwent an ABX task with CV(C)CVC nonwords of English /i:/-/ɪ/, /u:/-/ʊ/, /æ/-/ɛ/, /p/-/f/, and /r/-/l/ contrasts, a sequence repetition with nonwords of the /g/-/b/ contrast focussing on word stress, and speeded lexical decision tasks consisting of 33 English nonwords. The researchers found that the scores of tasks for WM, including PSTM, were related to overall scores of L2 phonological perception, including segmental and suprasegmental levels, except for the scores

of a paired-associated learning task. They also found that complex tasks for WM (sentence repetition test) were more strongly correlated to the L2 phonological scores.

More broadly, Kondo (2019) tested whether PSTM influences L2 listening skills. L1 Japanese speakers whose major was not English participated in the study. Their average English proficiency was A2 in the CEFR. Their English listening skills were measured using a dictation test. In the test, participants listened to 18 English sentences and then transcribed them verbatim. To assess PSTM, a forward-digit memory test and a nonword memory test were carried out. In the forward digit memory test, participants heard a set of digits in Japanese and verbally reproduced them. The test had 16 items, and the number of digits was between six and eleven. The nonword memory test consisted of 17 sets of nonwords based on the Japanese mora system. Each nonword was composed of three moras, and the number of moras ranged from two to five. Participants listened to the nonwords and reproduced them as they heard. Results revealed that the dictation scores positively correlated with the two tests' scores, but the effect size was small. Comparing the two test results, the forward digit memory test scores were a better predictor than those of the nonword memory test.

In a slightly different vein, Daidone and Darcy (2021) explored what variables, including PSTM, impact the accuracy of L2 lexical encoding. In their study, they conducted a lexical decision task to measure lexical encoding accuracy and an SNWR task for PSTM. The participants were L1 English speakers whose major or minor was Spanish, and the target phonemes were the Spanish /r/-/r/, /r/-/d/, /r/-/d/, and /f/-/p/ contrasts. In the lexical decision task, participants listened to a stimulus and answered whether it was a real Spanish word or not. The SNWR task was composed of sequences of Russian CVC nonwords, and the number of nonwords varied between four and seven. It was found that PSTM was a significant predictor of lexical encoding accuracy only for the /r/-/r/ contrast. The authors argued that PSTM was related to lexical encoding accuracy only for one contrast because only the /r/-/r/ contrast can be assimilated to a single L1 phonetic category, suggesting that storing both sounds finely is more essential than other contrasts. Hence, PSTM, which retains phonetic information, plays a role in lexical encoding.

Contrary to the studies reviewed above, other studies did not find such a positive association between PSTM and L2 perceptual accuracy. Safronova and Mora (2012) found that neither PSTM nor acoustic memory (AM) was related to L2 perception. Their study

investigated the relationship between PSTM and AM with Spanish-Catalan bilinguals whose English proficiency was intermediate to advanced levels. AM is an auditory sensory memory store used before phonological encoding to retain speech. Therefore, AM works when processing speech at the acoustic and phonetic level and PSTM at the phonological level. The participants completed a vowel identification task of the English /i:/-/ɪ/, a vocabulary size task, an SNWR task for PSTM, and an SNWR task with rotated speech for AM. The SNWR task for AM consisted of rotated speech of 144 Catalan nonwords used in the SNWR task for PSTM. The procedure was the same for both tasks. Results revealed that the PSTM and AM scores were not correlated, which means these two tasks measured different memory capacities. Also, the scores of neither of the tasks were related to their perceptual accuracy scores. The authors explained that neither PSTM nor AM related to perceptual accuracy because the participants were homogeneous in terms of their L2 proficiency and phonological training.

The studies on the link between PSTM and L2 perception reviewed so far are all cross-sectional in nature. Yet, there are a couple of studies that examined whether PSTM can predict the degree of L2 perceptual learning by conducting perceptual training (Aliaga-Garcia et al., 2011; Ghaffarvand Mokari & Werner, 2019). Aliaga-García, Mora, and Cerviño-Povedano (2011) conducted HVPT targeting 5 British English vowels /i, ɪ, a, ʌ, ɑ/ with L1 Catalan speakers whose English proficiency is advanced level. Participants completed ten sessions of an hour-long audio-visual HVPT with 176 trials, categorisation and AX discrimination tests before and after HVPT, and an SNWR task in Catalan. After training, the group who received high PSTM scores performed better in perception tests at pre-test, and their gain scores were also higher than those who got low PSTM scores.

In contrast, Ghaffarvand Mokari and Werner (2019) found no positive relationship between PSTM and L2 perceptual learning. They conducted five HVPT sessions on English vowels with Azerbaijani speakers in Tabriz, Iran. An SNWR task consisted of Azerbaijani nonwords to measure PSTM, following Cerviño-Povedano and Mora (2011). Perceptual accuracy of 11 British English vowel contrasts was tested in an AX task format before and after five sessions of 4-alternative-forced choice HVPT with 14 vowels. Each HVPT session had 225 trials and took 45 to 60 minutes. Results revealed that PSTM was not correlated with the gain scores of the AX task. The researchers claimed that the discrepancy between their findings and previous studies on the relationship between PSTM and L2 perception might be

attributed to the different degree of difficulty in target phonemes, or PSTM might not be related to L2 segmental learning.

2.6.2 Relationship between PSTM and production

Studies have also suggested a positive association between PSTM and L2 production learning (Kondo, 2012; Nagle, 2013). Nagle (2013) conducted a study to explore the factors that can predict the accentedness of L2 learners' production. Participants were L2 Spanish learners whose L1 was English. They had experience studying Spanish for at least six semesters. As part of the experiment, participants took a nonword pairs repetition test to measure their PSTM. The test was composed of 16 pairs of nonwords whose syllable lengths ranged between two to five. The nonwords were based on Russian and aligned with the English and Spanish phonotactics. Participants repeated each pair of nonwords immediately after hearing them in every trial. Their accentedness of production was recorded in a word-list reading task, a paragraph-reading task, and a free-response task and assessed by L1 Spanish speakers on a Likert scale. Results showed that their PSTM scores significantly predicted their production scores.

Kondo (2012) showed a positive correlation between verbal phonological memory and L2 production of suprasegmental features with L1 Japanese learners of English. L2 production skills were measured in a word repetition task to test the place of primary stress and the number of syllables, and a sentence repetition task was used for intonation and the number of words reproduced. In a verbal phonological memory task, participants heard a set of Japanese colour names and chose the correct sequence out of two. The study revealed that verbal phonological memory was positively related to L2 production accuracy.

Some studies partially confirmed an association between PSTM and L2 speech production depending on language backgrounds (Mora & Darcy, 2016) and L2 language proficiency (Moorman, 2017; Zahler & Lord, 2022). Mora and Darcy (2016) investigated the relationship between PSTM and L2 pronunciation. Monolingual L1 Spanish speakers and bilingual Spanish-Catalan speakers participated in the study. Their production data of the target contrasts (i.e., /i:/–/ɪ/ and /ʃ/–/tʃ/) were collected in a delayed sentence repetition task and measured by acoustic analyses and perceptual judgement. PSTM was measured in an SNWR task using Danish words. Results showed that only the production accuracy of

monolingual speakers was related to PSTM, indicating that learners' linguistic backgrounds, i.e., monolingual vs bilingual, influenced the relationship.

Zahler and Lord (2022) found mixed findings depending on learners' language proficiency in the L2. Participants were L1 English learners of Spanish in their fourth year at university or in a graduate program. They all completed a spontaneous speaking task for vowel elicitation. An SNWR task with CVC Russian words and nonwords measured their PSTM capacity. Results showed that PSTM was correlated with L2 vowel naiveness only for the fourth-year students.

In addition to the cross-sectional studies above, a longitudinal study conducted by Moorman (2017) exhibited an association between PSTM and L2 phonetic accuracy, but only for advanced learners. L1 English speakers of Spanish, at both the intro and advanced levels, took a speaking task three times over the course of a semester. The focus was on the Spanish /e/ and /o/ vowels, and the duration and the F1 and F2 values were analysed. PSTM was measured in an SNWR task. Results showed that PSTM could predict the production development of the /o/ sound only for advanced learners, whereas PSTM was a negative predictor of the production development of the /e/ sound for both the intro and advanced learners. These findings were opposite from Serafini and Sanz's (2015) study reporting that PSTM and L2 Spanish morphosyntactic learning were only strongly related at the initial stage of language learning.

Some studies, however, revealed that PSTM does not predict L2 production learning. Slevc and Miyake (2006) explored individual factors in L2 language learning with Japanese immigrants in the US. Participants completed several tasks for individual differences in addition to tasks testing their receptive and productive accuracy of English. PSTM was measured in a nonword-repetition task with English nonwords. Results showed that PSTM correlated with their receptive accuracy but not their productive accuracy.

Similarly, Venkatagiri and Levis (2007) also explored which factors contribute to individual differences in L2 pronunciation learning. In their study, EFL speakers completed numerous tasks to test their phonological awareness and PSTM. The PSTM tasks comprised a nonword association task using pictures and a nonword repetition task. In addition, participants' English comprehensibility was evaluated using a passage reading task and a picture narration task. Results revealed that PSTM was not correlated with their comprehensibility.

Kissling (2014) explored which factors contribute to individual differences in L2 pronunciation learning by conducting phonetics instructions. Spanish learners at a university in the US participated in the study. Half of them received modules of phonetics instruction on Spanish consonants over two weeks, and the other half underwent similar modules without explicit instructions on phonetics. The target sounds were eight Spanish consonants. Participants completed pre-, post-, and delayed-post word-list reading tasks for production data. PSTM was measured in a nonword repetition task. Results revealed no strong correlation between PSTM and production improvement, both with and without phonetics instruction.

In summary, many studies have shown that PSTM can predict L2 phonetic learning. However, other studies only partially replicated those studies or did not observe such a positive association between PSTM and L2 perception learning. In addition, most studies were cross-sectional, and only a few studies have explored the relationship between PSTM and the degree of learning by conducting phonetic training. Moreover, fewer studies have been conducted on production learning, and more research is necessary to fully unveil the relationship between PSTM and L2 phonetic learning.

2.7 English /l/ and /r/ sounds for L1 Japanese speakers

Research has shown that perceiving the English /l/–/r/ contrast is notoriously difficult for Japanese learners of English (e.g., Aoyama et al., 2004; Goto, 1971; Iino, 2018; Iverson et al., 2003; MacKain et al., 2008; Masaki et al., 1996; Saito, 2011; Sheldon & Strange, 1982; Shinohara & Iverson, 2021; Strange, 1992). One reason for this difficulty is the presence of a liquid phoneme in the Japanese language, which is an apico-alveolar tap or flap /r/ (Vance, 2008). When pronouncing the Japanese flap /r/ in the word-initial position, the tip of the tongue makes brief contact with the alveolar ridge, which is located just behind the upper front teeth. There is a small gap between the tip of the tongue and the upper teeth, especially when compared to the English /l/. The sound is produced with a very short duration, without the prolonged articulation observed in other similar sounds, such as the English /l/ and /r/ sounds. In terms of its quick contact, which is why it is described as tap or flap, the Japanese /r/ is closer to the flapped /d/ sound in American English than to the English /l/ or /r/ (Price, 1981).

In contrast, the English /l/ and /r/ sounds involve different articulatory gestures from those of the Japanese flap /r/. The English /l/, a lateral approximant, is produced by raising the

tip of the tongue to touch the alveolar ridge, which is directly behind the upper front teeth. Simultaneously, the sides of the tongue are lowered, allowing air to flow around the tongue (Sproat & Fujimura, 1993). The English /r/ sound, a postalveolar approximant, is produced without any contact between the tip of the tongue and the alveolar ridge, and the tongue is positioned further back in the mouth (Carley & Mees, 2019; Ogden, 2017).

L2 speech models provide a theoretical explanation for why the presence of flap /ɾ/ causes difficulty in acquiring a robust English /l/ and /r/ distinction. According to SLM (Flege, 1995a), the most difficult pattern for learning L2 sounds occurs when a similar sound to the target sound exists in learners' L1. In this case, the Japanese flap /ɾ/ is so similar to both the English /l/ and /r/ that it hinders the learning of these English sounds. Additionally, PAM (Best, 1995; Best & Tyler, 2007) also predicts the difficulty of learning this contrast, suggesting that when two phonemes from a target contrast can be assimilated into a single L1 sound, the contrast becomes particularly challenging to perceive accurately. Since the English /l/–/r/ contrast can both be assimilated to the Japanese flap /ɾ/, this contrast is considered to be extremely difficult for Japanese learners of English to master.

In addition, there are significant differences between how L1 English speakers and L1 Japanese speakers perceive the English /l/ and /r/ sounds. L1 English speakers distinguish the English /l/–/r/ contrast primarily based on the differences in F3 frequency—when F3 is higher, the phoneme is perceived as /l/, and when lower, it is perceived as /r/ (Ladefoged, 2003; Saito & Lyster, 2012). The auditory distinction is supplemented by closure duration and transition duration cues, leading to accurate identification of the sounds. However, L1 Japanese speakers tend to rely more heavily on F2, which can lead to unreliable perception of the contrast (Iverson et al., 2003). The reliance on F2 rather than F3 might explain why Japanese learners often struggle to differentiate between these English sounds.

Additionally, Lotto et al. (2004) examined the speech productions of Japanese learners of English and found that the Japanese flap /ɾ/ has a shorter duration compared to both English /l/ or /r/, and the formant ranges for F2 and F3 often overlap with those of English /l/ and /r/. This overlap leads to poorly segregated distributions in the productions of Japanese learners, where clear separation on the F3 continuum is lacking, causing confusion for L1 English speakers.

Despite such difficulty, according to the functional load theory, the English /l/–/r/ contrast is high functional load—ranked as maximal importance—and acquiring the contrast is crucial for successful communication (Brown, 1988; Munro & Derwing, 2006b).

Functional load refers to the importance of a phonetic contrast in distinguishing meaning within a language. Therefore, if learners cannot hear the distinction or pronounce the contrast intelligibly, it is likely to lead to confusion or miscommunication. Given that class time is generally limited, teaching high functional phonetic contrasts should be prioritised in pronunciation instruction.

However, many L1 Japanese speakers cannot attain the level of robustness in perception and production of highly proficient speakers, even after intensive perceptual training (Bradlow et al., 1999; Bradlow et al., 1997; Shinohara & Iverson, 2018; Ueda & Hashimoto, 2019) and explicit pronunciation instruction (Spring & Tabuchi, 2022). One exception is a study by Hattori (2009) which showed a robust learning after pronunciation instructions, although no improvement was observed in perception. Therefore, exploring and developing the effective way to learn the English /l/–/r/ contrast is important.

2.8 Current study

This thesis investigates the interactive effects of perception and production training on L2 phonetic learning, with a particular focus on how production during perceptual training affects phonetic learning. To achieve this, three experiments were conducted to enhance our understanding of L2 phonetic acquisition. Although perception and production training are both recognised for their respective contributions to L2 perception and production, the interaction between these two training methods and their combined impact on L2 phonetic learning remains underexplored. In addition, the experiments in the current thesis also examined the relationship between PSTM and phonetic learning.

Experiment 1 aimed to further investigate the findings of Baese-Berk and Samuel (2016) and Thorin et al. (2018). As discussed, the negative effect of production training observed in Baese-Berk and Samuel's study may be due to the cognitive demands of the task rather than a disruption of caused by to perceptual learning (Thorin et al., 2018). One factor that complicates the comparison between their results is the difference in the languages used—Baese-Berk and Samuel used L2 Basque, while Thorin et al.'s used L2 English. The incompatible results between these studies may be attributed to cross-linguistic differences. Given the possibility that production during perceptual training might disrupt perceptual learning, it is important to explore and extend these findings to provide further evidence regarding whether production practice negatively affects the improvement of perceptual

ability. Experiment 1 directly compared production and perception (discrimination) training methods with Japanese learners of English to address the cross-linguistic issue.

Experiment 2 extended the findings of Experiment 1 by employing an identification training method. While Experiment 1 used discrimination training, Experiment 2 was designed to further investigate the findings of Experiment 1 by using a different method to provide strong evidence about whether production practice disrupted perceptual learning. Additionally, Experiment 2 examined the effects of identification training on L2 perceptual learning. Previous studies have shown that identification training brings about better learning outcomes than discrimination training (Jamieson & Morosan, 1986; Logan et al., 1991; Strange, 1992; Strange & Dittmann, 1984). Thus, identification training may potentially counteract the negative effects of production practice during perceptual training.

Experiment 3 investigated whether the formation of phonetic categories through exposure to perceptual training facilitated production practices. Previous studies have shown that production practices do not necessarily improve L2 perceptual ability. However, Baese-Berk and Samuel (2016) found that prior exposure to the target language may alleviate the negative or null effect of production during perceptual training. This suggests that being exposed to target sounds might allow L2 learners to form phonetic categories, and such formation may be necessary for production practices to positively impact L2 perceptual learning. Experiment 3 explored this issue by comparing four types of training: perception-only training, perception-and-production-mixed training, production-only training, and perception-to-production training. Additionally, Experiment 3 was designed to examine whether and how these four training methods affected the improvement of L2 production ability, addressing the controversial issues discussed in Literature Review: i.e., whether perception and production training effectively improves production and perceptual abilities, respectively.

Chapter 3 Experiment 1: The effects of production during discrimination training and the relationship between L2 perception and PSTM

3.1 Introduction

Previous research has found that perceptual training is effective for both perceptual and production learning (e.g., Aliaga-Garcia & Mora, 2009; Bradlow et al., 1999; Huensch & Tremblay, 2015; Iverson et al., 2005; Lambacher et al., 2005; Lopez-Soto & Kewley-Port, 2009; Motohashi-Saigo & Hardison, 2009; Shinohara & Iverson, 2018; Thomson, 2011). However, some researchers claimed that the learning gained from such perceptual training is not robust for many L2 learners even after a long period of training (e.g., Bradlow et al., 1999; Shinohara & Iverson, 2018). Given that perception and production skills are considered to be closely tied based on shared phonetic categories or representations (Best & Tyler, 2007), adding a production component (i.e., repeating stimuli aloud) to perceptual training is assumed to facilitate perceptual learning. However, some studies have found that producing target sounds during perceptual training did not facilitate perceptual learning (Lu et al., 2015; Thorin et al., 2018), while others revealed that production disrupted perceptual learning (Baese-Berk, 2010, 2019; Baese-Berk & Samuel, 2016; Wright, 2021). These researchers raised several possible reasons for why production causes disruption to perceptual learning, such as listening to their own ‘inaccurate’ speech and switching tasks between perception and production. However, perception and production mixed training have not yet been explored extensively.

To fill the gap in the literature, this experiment was adapted from Baese-Berk and Samuel (2016) and Thorin et al. (2018). As discussed in the previous chapter, the negative effect of production training observed in Baese-Berk and Samuel (2016) may be due to their task being cognitively costly rather than L2 phonetic perceptual learning being disrupted (Thorin et al., 2018). Another possible factor leading to the incompatible results between these studies is cross-linguistic differences, that is, Baese-Berk and Samuel’s (2016) use of L2 Basque compared to Thorin et al.’s (2018) use of L2 English. For a better understanding of the nature of the production effect during perceptual training on L2 perceptual learning, it is important to extend their studies and provide further evidence about whether production practice negatively affects the improvement of L2 perceptual ability. The current study directly compares their training methods (i.e., the different timings of production) by dealing with the cross-linguistic issue with a single group of Japanese L2 learners of English.

In addition, it is also important to examine what factor contributes to individual differences in perceptual training effects, such as learners' PSTM capacity. Although the relationship between PSTM and L2 perception has been getting increasing attention lately, previous studies have shown mixed findings: positive correlation (Cerviño-Povedano & Mora, 2011; Cerviño-Povedano et al., 2011; Inceoglu, 2019; Kondo, 2019; Lengeris & Nicolaidis, 2014) and no correlation (Safronova, 2016; Safronova & Mora, 2012). Moreover, many of the studies were cross-sectional, and little research has been conducted on whether PSTM can be a predictor of the degree of perceptual learning after phonetic training. Even among the few studies, contradicted findings have been shown (Aliaga-Garcia et al., 2011; Ghaffarvand Mokari & Werner, 2019).

Accordingly, the purpose of this study is twofold: (1) to investigate the production effect during perceptual training on the English /l-/r/ contrast with L1 Japanese learners of English, and (2) the relationship between PSTM and phonetic training outcomes. The research questions are as follows:

1. Does perceptual training using discrimination tasks improve L2 phonetic perception?
2. Does production during perceptual training in a discrimination task format disrupt L2 phonetic perceptual learning?
- 3a. Is perceptual accuracy prior to phonetic training related to learners' PSTM?
- b. Is the degree of perceptual learning by phonetic training related to learners' PSTM?

3.2 Method

3.2.1 Participants

3.2.1.1 Japanese learners of English

Data collection for this study was conducted at a university in Japan. In total, 74 Japanese learners of English participated in this phonetic training study. However, those whose first language was not Japanese and who did not complete the pre- and post-tests of the discrimination and identification tasks and all the training sessions properly were excluded from the analysis. Thus, the data of 49 participants were initially used for analysis. The number further decreased to 46 after the analysis of the pre-test data because three participants received scores over 90 out of 100 on both discrimination and identification tasks. The threshold score of 90 was chosen to exclude participants who demonstrated near-ceiling

performance prior to training following previous studies such as Okuno and Hardison's (2016) study. This decision was made because such participants are unlikely to benefit from training in a measurable way, and including these participants could have skewed the analysis. In addition, four participants who did not complete the second PSTM task properly (e.g., recording errors) were excluded from analysis for the relationship between PSTM and perceptual learning.

Participants were 18- to 19-year-old undergraduate students. Their majors were School Nursing, Art and Calligraphy Education, Mathematics Education, Home Economics Education, and Music Education. All participants were taking an English course covering the same topics and using the same learning materials facilitated by the same instructor. Their English proficiency was A2 to B1 in the CEFR, which is a beginner to lower-intermediate level. None of them reported that they had lived in English-speaking countries for more than three months. The participants who completed all the training sessions and pre- and post-training tasks received extra credits for the English course.

3.2.1.2 L1 English speakers for stimulus recording

In addition to the Japanese participants, seven L1 Australian English speakers were recruited for the preparation of the audio stimuli. Because of too much static noise, the recording of one male speaker was discarded. Thus, the recordings by the rest of the six speakers (female = 3, mean age = 28.67, *SD* of age = 5.72) were used.

3.2.2 Materials

3.2.2.1 English stimuli for training sessions and tests

The target phonemes investigated in the current study are the English /l/–/r/, which are known to be difficult to perceive and produce by L1 Japanese speakers (Lambacher, 1999). For the discrimination and identification tasks, 20 minimal-pair words starting with the English /l/–/r/ (e.g., light/right) and five minimal pairs that do not contain the target phonemes (for training practice) were created and recorded by six L1 Australian English speakers (see Appendix A).

The recording of the stimuli was conducted during the COVID-19 lockdown period. Therefore, the L1 speakers recorded the stimuli on their own following the instructions given by the researcher. The speakers used one of the following voice recording applications on their smartphones: the Voice Memos app on iPhone, Voice Recorder app on Galaxy, and

Recorder app on OPPO. They were provided a list of words and asked to read each one twice in a row. Their first token of each word was extracted as stimuli using Praat (Boersma, 2001).

The training tasks were composed of stimuli with 20 minimal-pair words with target phonemes spoken by five speakers. Pre- and post-tests were composed of stimuli with four types of familiarity shown in Table 3.1. The first stimulus type consisted of 20 trained minimal word pairs spoken by one of the trained speakers (TWTS). The second type grouped 20 trained minimal pairs spoken by an untrained speaker whose speech was not used in training (TWUS). The third type included five untrained pairs spoken by a trained speaker (UWTS). Finally, the fourth type contained five untrained pairs spoken by an untrained speaker (UWUS). The first type was used to test whether the participants become able to discriminate and identify trained stimuli. The rest of the three types were used to test whether their perceptual learning can be generalised to untrained stimuli that did not appear in the training sessions.

Table 3.1 Experiment 1: Stimulus Types of Different Word and Speaker Familiarity

	Word familiarity	Speaker familiarity	Minimal pairs
TWTS	Trained	Trained	20
TWUS	Trained	Untrained	20
UWTS	Untrained	Trained	5
UWUS	Untrained	Untrained	5

3.2.2.2 Stimuli for the PSTM tasks

The PSTM task was designed following Kondo’s Non-Word Memory Test (Kondo, 2019). In the current experiment, two sets of 65 nonwords were created based on the Japanese phonological system (see Appendix B). Two sets were created because some learners conducted the first set more than once. All the nonwords were three-mora words (i.e., CVCVCV) and did not appear in the Super Daijirin Japanese dictionary (Sanseido, 1999). Both sets of nonwords were recorded by a female L1 Japanese speaker using the Voice Memos app on iPhone 11.

3.2.2.3 Language background questionnaire

The questionnaire asked participants’ gender, age, first language, onset age of studying English, experiences of staying outside of Japan for more than three months, and languages that they can speak besides Japanese and English (see Appendix C).

3.2.3 Procedure

The experiment was conducted online from December 2020 to February 2021 (during the COVID-19 pandemic) and included a pre-test, four training sessions, and a post-test. The pre- and post-training tasks and the perception and production training tasks were created using Ibex farm, a web-based experimental presentation platform. For communicating with participants, online tools such as email, Zoom, Google Forms, and Google Sites were used. Communication between L1 Japanese participants about the research, including instructions for each task, was conducted in their first language, Japanese. This procedure was reviewed and approved by ANU Human Ethics Committee, and participants provided consent.

The timeline of the data collection is summarised in Figure 3.1. Participants were randomly divided into three groups, Production-first group (PROD-FIRST), Production-last group (PROD-LAST), Perception-only group (PERC-ONLY). First, all L1 Japanese participants completed a language background questionnaire, a PSTM task, a discrimination task pre-test, and an identification task pre-test. The discrimination task took about 15 minutes, the identification task 10 minutes, the PSTM task 5 minutes, and filling the questionnaire 5 minutes. Then, participants received four sessions of different training based on their group within 24 days and did not complete more than one session per day. The training sessions were spread over time instead of being conducted in a one-day single session to avoid fatigue effects. Each training session was designed to be completed in 15 minutes so that participants could concentrate on learning throughout training. Another reason why the training sessions were spread over time was to employ a spaced learning design. Second language acquisition research suggests that spaced learning, where learning is spread across a certain period of time, has better learning results than massed learning, where all the learning procedure takes place in a single day (Nakata, 2015). Once they had completed all the training sessions, the participants completed a PSTM task, a discrimination task post-test, and an identification task post-test within 12 days of the last day of training.

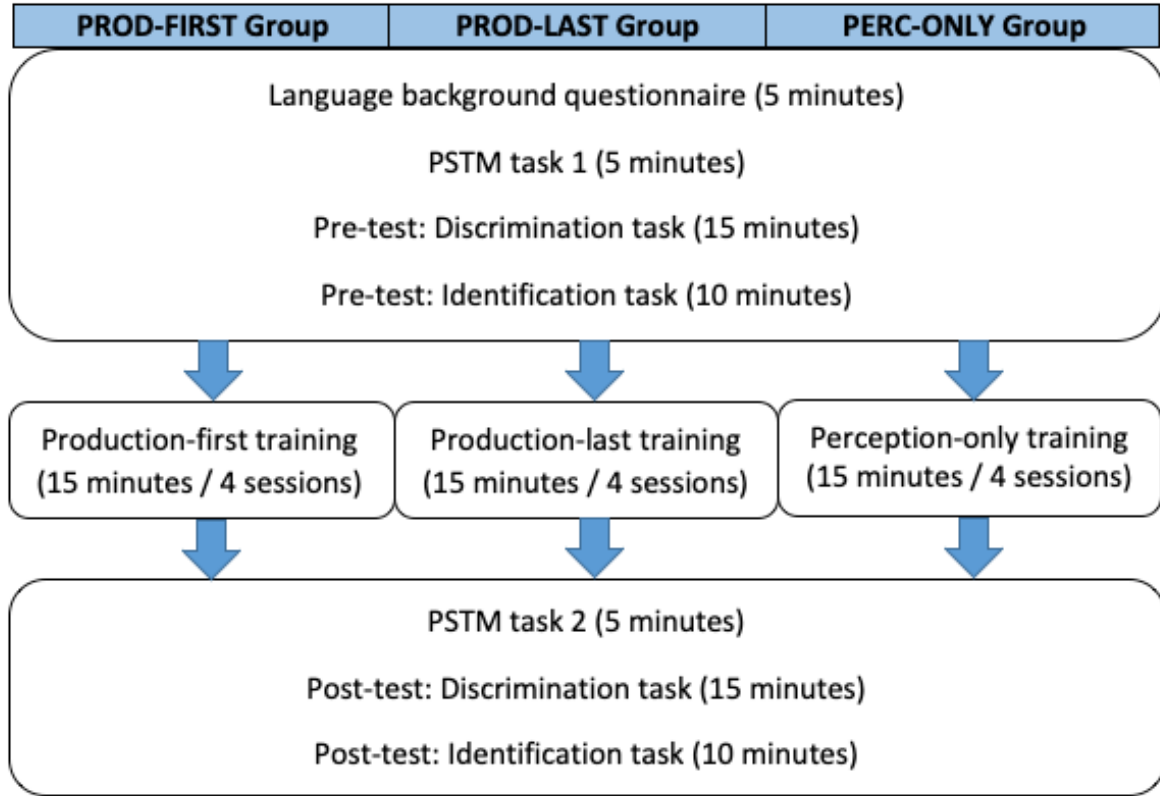


Figure 3.1 Experiment 1: Timeline of Data Collection

3.2.3.1 Phonetic training

The training was designed in an ABX task format with an interval of 300 milliseconds between A–B and B–X. The PROD-FIRST group received the perception and production training, which was designed based on Baese-Berk and Samuel’s (2016) study (i.e., participants listened to three words, repeated the last word aloud and answered which of the first or second aurally presented words was the same as the last word). The PROD-LAST group received training following Thorin et al.’s (2018) training format (i.e., participants listened to three words and judged which of the first or second aurally presented words was the same as the last word before repeating the last word aloud). The PERC-ONLY group was treated as a control group and completed only the perception training.

Each training session was composed of 100 trials. Since there were 200 training tokens in total (40 words × five speakers), participants encountered the same token twice during training. The order of presenting the stimuli was randomised. Each session started with five practice trials using words that do not contain /l/ and /r/. Participants in all groups received immediate feedback on their responses during these training sessions. Each training session lasted about 15 minutes.

3.2.3.2 PSTM task

In the task, participants listened to the recording of sets of Japanese nonwords and orally repeated the words in the same order as much as they could. As in Kondo's Non-Word Memory Test (Kondo, 2019), the PSTM task of the current experiment consisted of 17 trials followed by two practice trials. The first trial was composed of a set of two nonwords. The second to seventh trials were composed of three nonwords. Then, there were eight sets of four nonwords, followed by two sets of five nonwords.

Participants' oral repetition of the nonwords was video-recorded on Zoom using the recording system. Their repetition was assessed by an L1 Japanese speaker following Kondo (2019); that is, the score of each trial was scored from 0 to 10 by calculating using the following equation: the number of accurate reproduction $\times$ (10 / the number of nonwords). The maximum score was 170.

In the initial study design, participants were supposed to complete the PSTM task only once before training. However, because some participants attempted the task more than once, and there was no guarantee that they had submitted the video of their first attempt, they were asked to complete the task with a different set of nonwords after training once more to ensure that every participant completed the second PSTM task only once. For analysis, data of the second PSTM task were used.

3.2.3.3 Discrimination task

In the discrimination task, participants listened to three English words consecutively. The first two words contained either /l/ or /r/, and the third word matched one of the two. Participants were asked to answer whether the third word matched the first or second by pressing the designated key on the keyboard (i.e., 1 for /l/ and 2 for /r/). No feedback was given, and there was no time limit for answering. The task was composed of 100 trials (50 words $\times$ two speakers), which included five minimal pairs and one speaker not used as trained stimuli. The order of presenting the stimuli was randomised. Participants had ten practice trials in each pre- and post-training task before the test part. There was no break between trials.

3.2.3.4 Identification task

In the identification task, participants listened to a word containing either /l/ or /r/ and then judged which of the two orthographically presented probe words corresponded to the aurally presented word. Similar to the discrimination task, no feedback was given, and the stimulus order was randomised. There were 100 trials with 40 trained and ten untrained words

spoken by a trained speaker and an untrained speaker. Likewise, the task started with ten practice trials, and participants did not have a break between trials.

3.2.4 Statistical analysis

As mentioned in the participants' section, three participants who scored higher than 90% on the pre-test both in discrimination and identification tasks were excluded from data analysis entirely to avoid a ceiling effect (Cramer & Howitt, 2004). Thus, the number of participants in each group was 14 in the PROD-FIRST group, 17 in the PROD-LAST group, and 15 in the PERC-ONLY group.

Statistical analyses were conducted in R (R Core Team, 2023). First, a logistic mixed-effects model was fitted using the lmerTest package (Kuznetsova et al., 2017) to the discrimination task and identification task data separately with the binary score (correct or incorrect) as the dependent variable and Group (PROD-LAST, PROD-FIRST, and PERC-ONLY), Time (pre-test and post-test), and Stimulus type (TWTS, TWUS, UWTS, and UWUS) as fixed effects. Regarding random effects, the maximal model included Participants and Words as random intercepts. Random slopes were also considered, but because the model failed to converge, they were not included. The reference level was set to the PERC-ONLY group's pre-test of TWTS stimuli. Odds ratios and 95% confidential intervals were calculated using the sjPlot package (Lüdtke, 2024).

To determine the significant fixed effects, each model was pruned by excluding the most non-significant fixed effect in a stepwise fashion. As post-hoc analyses, Tukey's pairwise comparison tests were performed using the emmeans function (Lenth, 2023) for further analyses. An alpha level of .05 was used for all statistical tests.

With regard to the PSTM task, four participants who did not complete the second PSTM task properly were excluded from analysis, as explained in the participants' section (3.2.1.1). The following are the numbers of participants in each group: PROD-FIRST group = 13, PROD-LAST group = 15, and PERC-ONLY group = 14.

To answer Research Question 3a on the relationship between PSTM and pre-training perceptual accuracy, a linear regression model was fitted to the pre-test data with accuracy scores with all stimuli combined as the dependent variable and the PSTM score as the independent variable. The discrimination tasks and identification tasks were analysed separately.

To answer Research Question 3b, exploring the relationship between PSTM and the degree of perceptual learning, a linear regression model was fitted to the discrimination and identification task data separately with the gain score as the dependent variable and the PSTM score, Pre-test score, and Group as independent variables. Gain scores were calculated by subtracting the pre-test scores from post-test scores for each participant. Pre-test scores were included in the model to control for participants' initial performance levels and to examine whether participants with different initial perceptual accuracy benefited differently from phonetic training based on their PSTM capacity.

3.3 Results

3.3.1 Discrimination task

3.3.1.1 Descriptive statistics

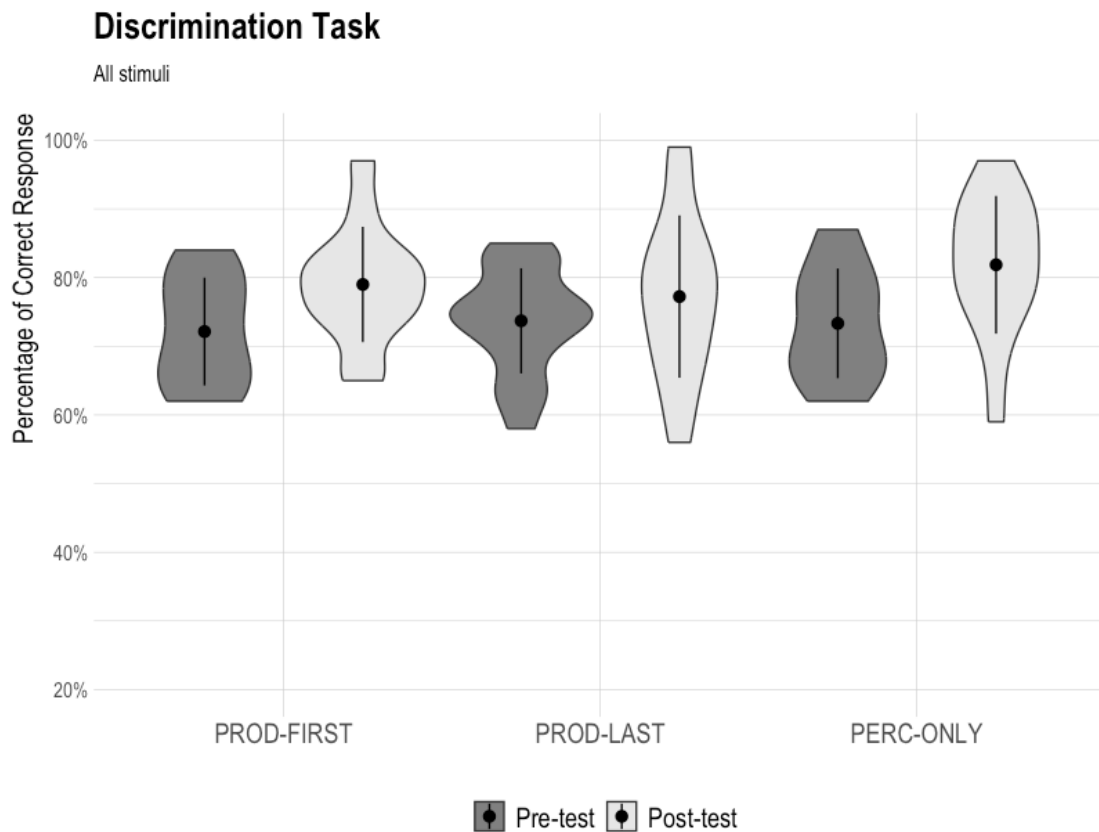
Table 3.2 shows the mean number of correct responses, standard deviations, and percentage of correct responses in discrimination tasks per training group and familiarity type of stimuli. The maximum score of all the stimuli was 100, which combined the scores of four stimulus types. The TWTS (trained words spoken by a trained speaker) stimuli and TWUS (trained words spoken by the untrained speaker) stimuli had a maximum of 40 points each, and the UWTS (untrained words spoken by a trained speaker) stimuli and UWUS (untrained words spoken by the untrained speaker) stimuli had a maximum of 10 points each. First, the results of all stimuli combined are presented, followed by each stimulus type one by one below.

Table 3.2 Experiment 1: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-Tests of the Discrimination Task

		PROD-FIRST		PROD-LAST		PERC-ONLY	
		Mean (S.D.)	Percentage	Mean (S.D.)	Percentage	Mean (S.D.)	Percentage
All stimuli	Pre-test	72.14 (7.85)	72.1%	73.71 (7.65)	73.7%	73.33 (7.99)	73.3%
	Post-test	79.00 (8.39)	79.0%	77.24 (11.82)	77.2%	81.87 (10.03)	81.9%
Trained words x Trained speaker	Pre-test	26.36 (3.39)	65.9%	27.06 (3.82)	67.6%	26.33 (3.37)	65.8%
	Post-test	29.07 (3.58)	72.7%	28.41 (6.00)	71.0%	30.20 (4.66)	75.5%
Trained words x Untrained speaker	Pre-test	30.71 (4.08)	76.8%	31.47 (3.87)	78.7%	31.53 (3.66)	78.8%
	Post-test	33.00 (4.21)	82.5%	33.00 (4.80)	82.5%	34.80 (3.99)	87.0%
Untrained words x Trained speaker	Pre-test	7.43 (1.55)	74.3%	6.88 (1.50)	68.8%	7.20 (1.42)	72.0%
	Post-test	8.07 (1.44)	80.7%	7.53 (1.46)	75.3%	7.93 (1.71)	79.3%
Untrained words x Untrained speaker	Pre-test	7.64 (1.69)	76.4%	8.29 (1.21)	82.9%	8.27 (1.75)	82.7%
	Post-test	8.86 (1.03)	88.6%	8.29 (1.69)	82.9%	8.93 (1.39)	89.3%

Before training, all groups answered more correctly to the untrained words stimuli than trained words stimuli. Additionally, all groups performed better scores on the stimuli spoken by an untrained speaker for both trained words. Thus, the stimuli used during training were found to be the most challenging before training, compared to other types of stimuli.

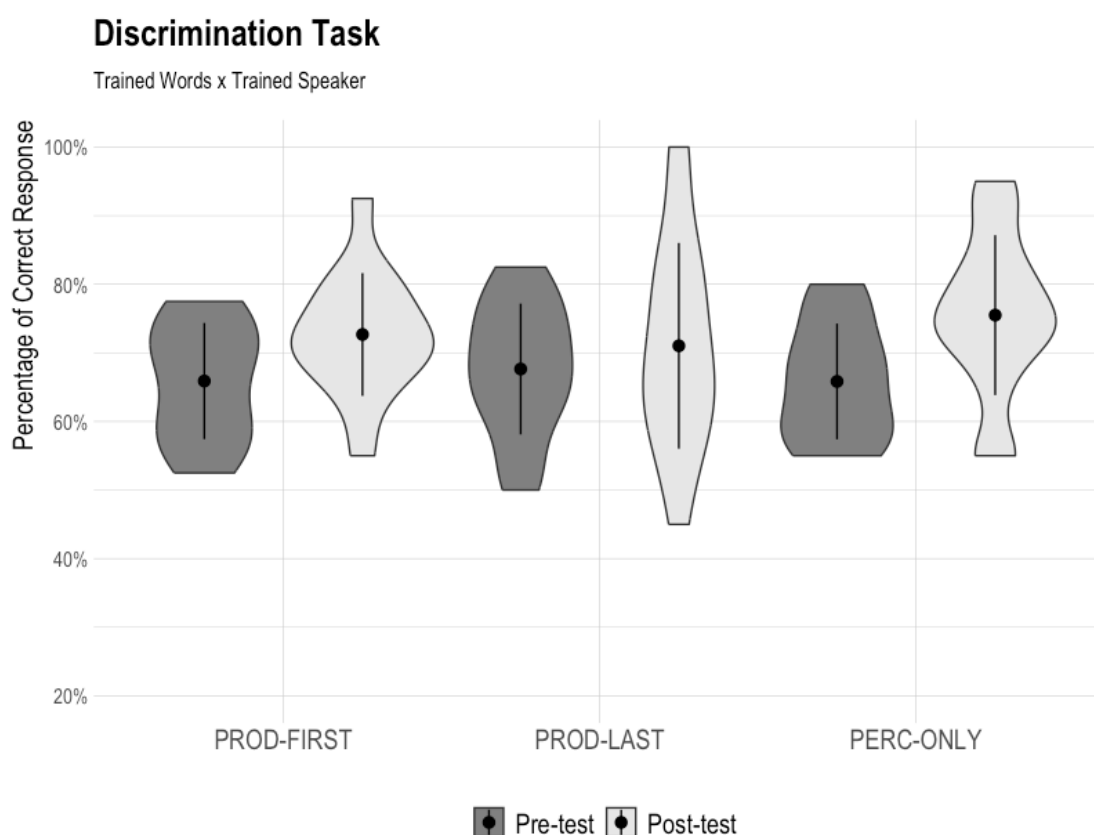
Figure 3.2 illustrates the pre- and post-test results of the discrimination tasks combining all the stimuli. The three groups started with close scores in the range of 72.1% to 73.7%. After training, all groups improved their score from pre-test to post-test. The PERC-ONLY group improved 8.6% and achieved 81.9%, which was the highest among the three. The PROD-FIRST group increased by 6.9%, and the PROD-LAST group, 3.5%. Both production groups did not reach 80% accuracy after training.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 3.2 Experiment 1: Discrimination Task Results with All Stimuli

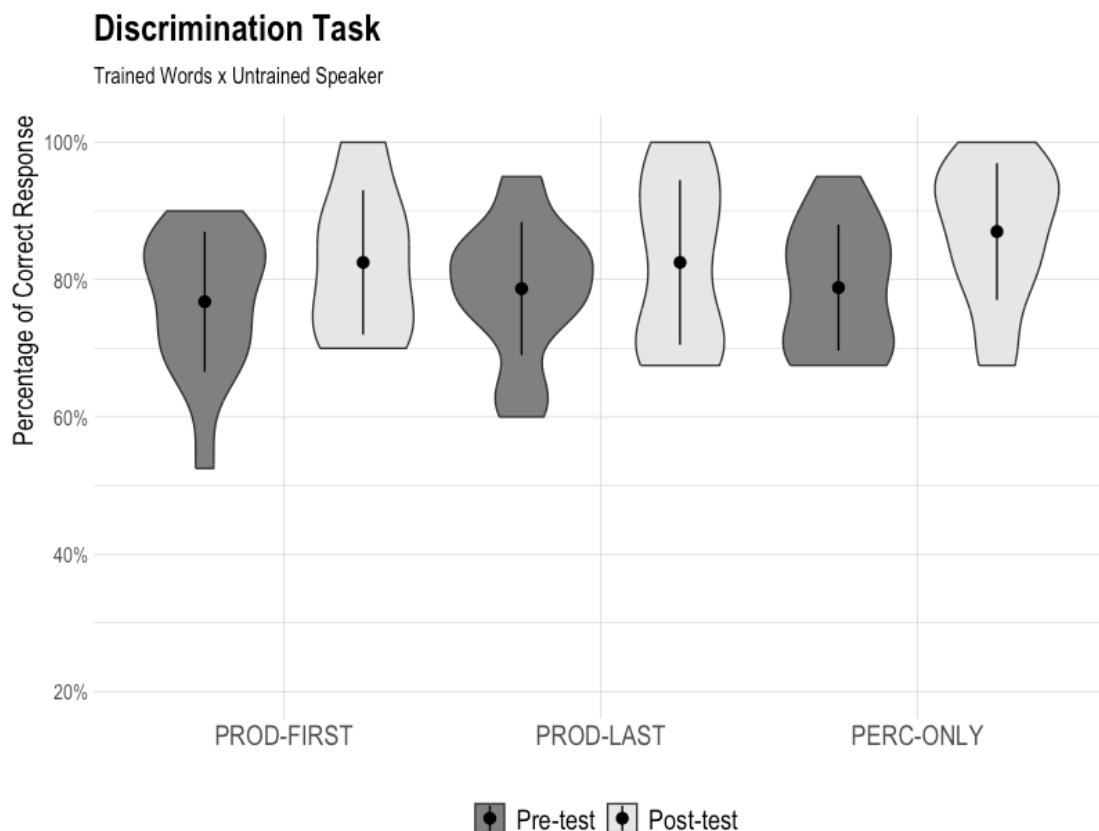
Figure 3.3 visualises the discrimination task results of the TWTS stimuli, and it shows that all groups reached higher than 70% accuracy at post-test. Before training, the PROD-FIRST group and PERC-ONLY group scored very close at 65.9% and 65.8%, respectively. The PROD-LAST group's pre-test score was a bit higher than the other two groups, which was 67.6%. After training, The PERC-ONLY group improved 9.7% and achieved the highest accuracy rate (75.5%), followed by the PROD-FIRST group's 72.7% (6.8% increase). The PROD-LAST group received the lowest score, 71.0%, and improved only 3.4%, although they attained the highest pre-test score.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 3.3 Experiment 1: Discrimination Task Results with the TWTS Stimuli

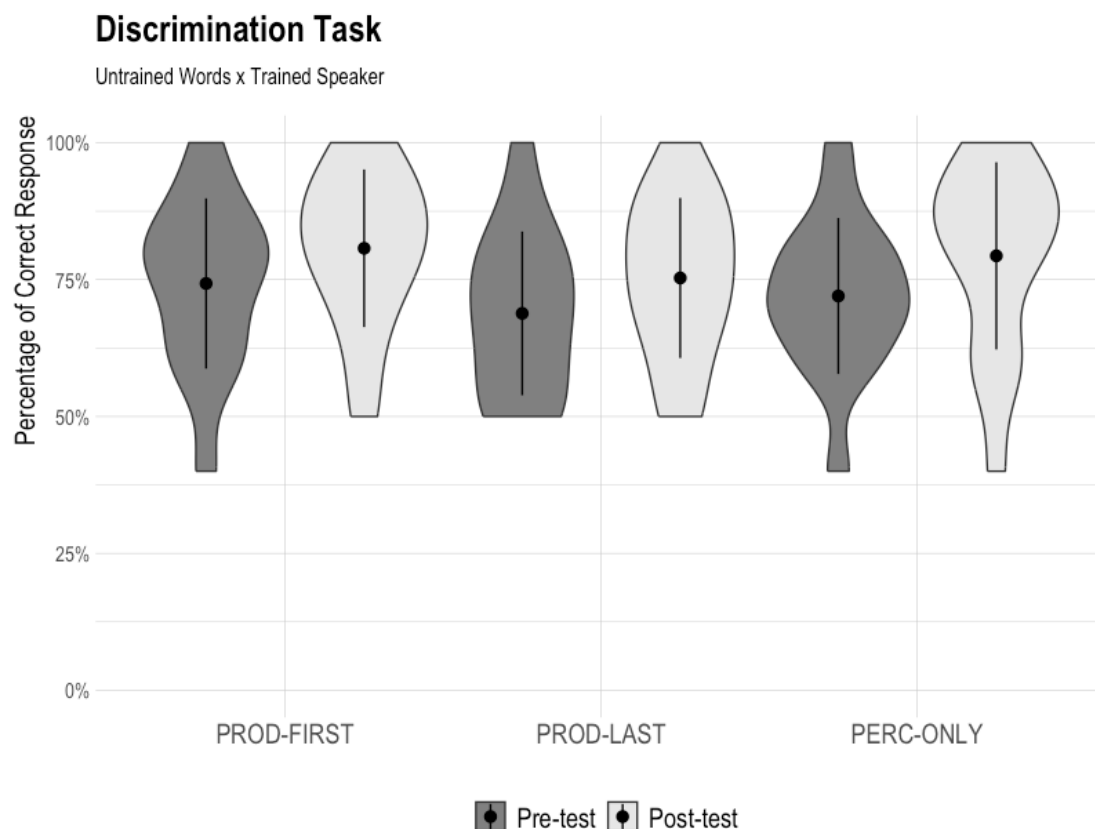
Figure 3.4 shows the discrimination task results of the TWUS stimuli. The PROD-LAST group and PERC-ONLY group attained a very close accuracy score, 78.7% and 78.8%, respectively. The PROD-FIRST group scored 76.8% at pre-test, which was lower than the other two groups. After training, the PERC-ONLY group improved 8.2% and achieved 87% accuracy, while the PROD-FIRST group (5.7% increase) and PROD-LAST group (3.8% increase) received 82.5%.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 3.4 Experiment 1: Discrimination Task with the TWUS Stimuli

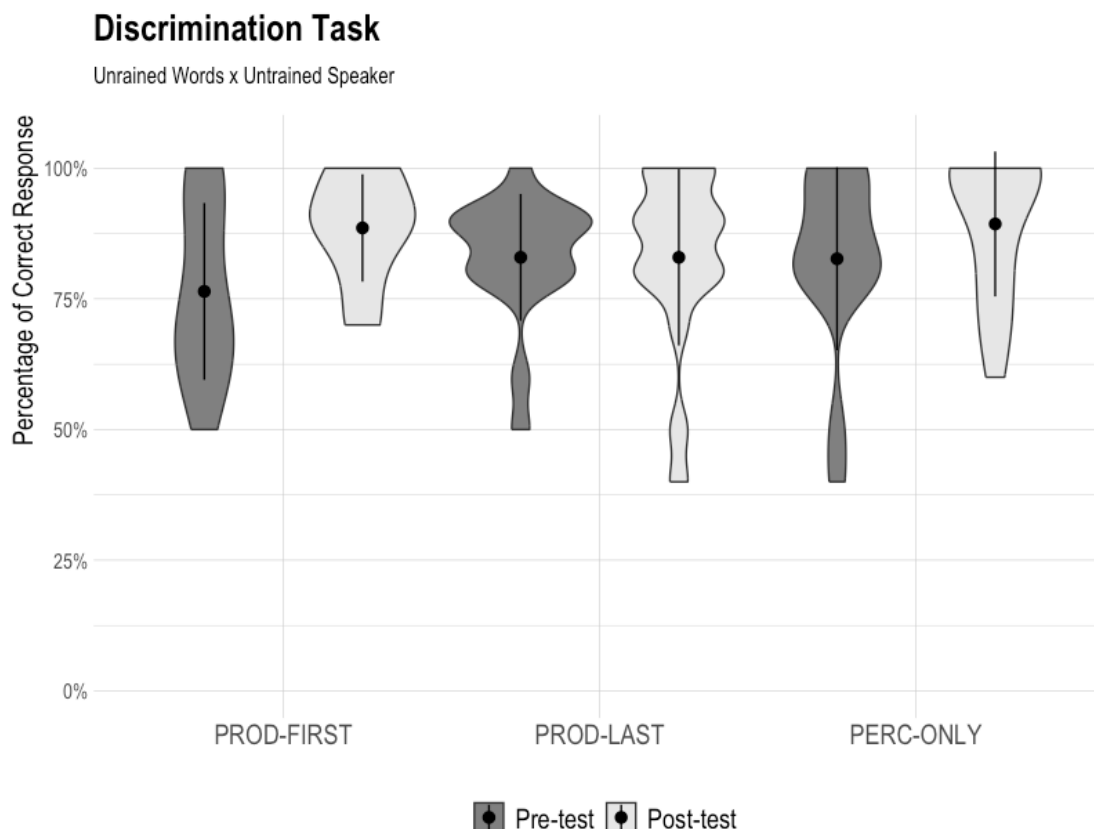
Figure 3.5 depicts the discrimination task results of the UWTS stimuli. The PROD-FIRST group received the highest score (74.3%), followed by the PERC-ONLY group (72%). The PROD-LAST group attained the lowest score (68.8%). After training, all group increased their accuracy rate by around 7%. The PERC-ONLY group increased the score by 7.3%, the PROD-LAST group by 6.5%, and the PROD-FIRST group by 6.4%.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 3.5 Experiment 1: Discrimination Task Results with the UWTS Stimuli

Figure 3.6 shows the discrimination task results of the UWUS stimuli. The PROD-FIRST group started with the lowest score among the three groups (76.4%). The PROD-LAST group and PERC-ONLY group received a very close accuracy rate at pre-test (82.9% and 82.7%, respectively). After training, The PROD-FIRST group increased their accuracy by 12.2%, which was the largest improvement among the three groups, followed by the PERC-ONLY group's 6.6%. The PROD-LAST group's accuracy rate remained the same.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 3.6 Experiment 1: Discrimination Task Results with the UWUS Stimuli

3.3.1.2 Inferential Statistics

Table 3.3 shows the final model of a logistic mixed-effects analysis with all stimuli when the reference level was the PERC-ONLY group's pre-test with the TWTS stimuli. The three-way interaction was not significant and was excluded from the final model. The model revealed that an interaction between Time and Group was significant ($p = 0.008$), and Time was also a significant predictor ($p < 0.001$), suggesting that the PERC-ONLY group

significantly improved their discrimination accuracy, and the improvement was significantly larger than that of PROD-LAST group but not that of PROD-FIRST group. There was no significant difference between groups at pre-test.

Table 3.3 Experiment 1: Model Output of a Logistic Mixed Effects Model for the Discrimination Task

<i>Predictors</i>	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	2.09 ***	1.56 – 2.81	<0.001
time [Post-test]	1.70 ***	1.42 – 2.04	<0.001
group [PROD-FIRST]	0.93	0.63 – 1.36	0.696
group [PROD-LAST]	1.02	0.71 – 1.48	0.896
stimuli [TWUS]	1.93 ***	1.73 – 2.16	<0.001
stimuli [UWTS]	1.32	0.98 – 1.79	0.068
stimuli [UWUS]	2.38 ***	1.73 – 3.26	<0.001
time [Post-test] × group [PROD-FIRST]	0.87	0.68 – 1.13	0.302
time [Post-test] × group [PROD-LAST]	0.72 **	0.57 – 0.92	0.008

Random Effects

σ^2	3.29
τ_{00} word	0.13
τ_{00} participant	0.22
ICC	0.10
$N_{\text{participant}}$	46
N_{word}	50
Observations	9200
Marginal R^2 / Conditional R^2	0.041 / 0.133

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Neither the Time × Stimuli interaction nor the Group × Stimuli interaction was significant, indicating that participants consistently answered more correctly to the stimuli spoken by an untrained speaker both before and after training, and this pattern was consistent across all groups. Stimuli as a main effect was a significant predictor, suggesting that all groups answered significantly more correctly to the stimuli spoken by the untrained speaker.

Table 3.4 shows the results of pairwise comparison analyses between Time and Group. The analyses revealed that the PERC-ONLY group ($p < 0.001$) and PROD-FIRST group ($p < 0.001$) significantly increased their accuracy score after training. In contrast, the PROD-LAST group did not show significant differences between their pre-test and post-test results ($p = 0.133$). Additionally, there was a significant difference between the PROD-FIRST group's pre-test score and the PERC-ONLY group's post-test score ($p = 0.026$), highlighting that the high score obtained by the PERC-ONLY group after training, given that there was no significant difference between their pre-test scores.

Table 3.4 Experiment 1: Summary of Pairwise Comparisons between Time and Group for the Discrimination Task

<i>Pre-test</i>	<i>Pre-test</i>	β	<i>SE</i>	<i>z</i>	<i>p</i>
PROD-FIRST	PROD-LAST	0.024	0.183	0.131	1.000
	PERC-ONLY	-0.075	0.192	-0.390	0.999
PROD-LAST	PERC-ONLY	-0.099	0.186	-0.530	0.995
<i>Post-test</i>	<i>Post-test</i>				
PROD-FIRST	PROD-LAST	0.090	0.189	0.473	0.997
	PERC-ONLY	-0.206	0.197	-1.049	0.901
PROD-LAST	PERC-ONLY	-0.296	0.187	-1.580	0.612
<i>Pre-test</i>	<i>Post-test</i>				
PROD-FIRST	PROD-FIRST	-0.391	0.091	-4.323	0.000
	PROD-LAST	-0.302	0.187	-1.610	0.592
	PERC-ONLY	-0.598	0.195	-3.068	0.026
PROD-LAST	PROD-FIRST	-0.292	0.189	-1.551	0.631
	PROD-LAST	-0.203	0.082	-2.470	0.133
	PERC-ONLY	-0.499	0.186	-2.675	0.080
PERC-ONLY	PROD-FIRST	-0.316	0.194	-1.632	0.577
	PROD-LAST	-0.227	0.184	-1.230	0.822
	PERC-ONLY	-0.523	0.091	-5.746	<.0001

Table 3.5 summarises the results of pairwise comparisons between Time and Stimuli. They revealed that the TWTS stimuli were scored significantly lower than the TWUS stimuli ($p < 0.001$) and the UWUS stimuli ($p = 0.001$) at pre-test. These significant differences remained the same at post-test.

Table 3.5 Experiment 1: Summary of Pairwise Comparisons between Time and Stimuli for the Discrimination Task

<i>Pre-test</i>	<i>Pre-test</i>	β	<i>SE</i>	<i>z</i>	<i>p</i>
TWTS	TWUS	-0.624	0.077	-8.093	<.0001
	UWTS	-0.255	0.173	-1.472	0.822
	UWUS	-0.808	0.183	-4.414	0.000
TWUS	UWTS	0.368	0.176	2.099	0.415
	UWUS	-0.184	0.185	-0.995	0.975
UWTS	UWUS	-0.552	0.161	-3.423	0.014
<i>Post-test</i>	<i>Post-test</i>				
TWTS	TWUS	-0.695	0.084	-8.257	<.0001
	UWTS	-0.311	0.180	-1.724	0.672
	UWUS	-0.937	0.197	-4.766	0.000
TWUS	UWTS	0.385	0.184	2.093	0.419
	UWUS	-0.241	0.200	-1.208	0.930
UWTS	UWUS	-0.626	0.181	-3.451	0.013
<i>Pre-test</i>	<i>Post-test</i>				
TWTS	TWTS	-0.327	0.074	-4.410	0.000
	TWUS	-1.023	0.083	-12.404	<.0001
	UWTS	-0.638	0.179	-3.557	0.009
	UWUS	-1.264	0.196	-6.456	<.0001
TWUS	TWTS	0.296	0.079	3.751	0.004
	TWUS	-0.399	0.087	-4.606	0.000
	UWTS	-0.014	0.181	-0.080	1.000
	UWUS	-0.641	0.198	-3.241	0.026
UWTS	TWTS	-0.072	0.174	-0.414	1.000
	TWUS	-0.768	0.178	-4.314	0.000
	UWTS	-0.383	0.157	-2.432	0.226
	UWUS	-1.009	0.176	-5.742	<.0001
UWUS	TWTS	0.480	0.184	2.613	0.151
	TWUS	-0.215	0.187	-1.149	0.946
	UWTS	0.170	0.168	1.011	0.973
	UWUS	-0.457	0.185	-2.470	0.208

Pairwise comparisons between pre-test and post-test scores for stimulus types revealed significant improvements in discrimination accuracy for the TWTS stimuli ($p = 0.000$) and the TWUS stimuli ($p = 0.000$) but not for the UWTS stimuli ($p = 0.226$) or the UWUS stimuli ($p = 0.208$). These results suggest that participants increased their scores more for the trained word stimuli compared to the untrained word stimuli. Overall, the findings

indicate that training in this experiment effectively enhanced participants' discrimination accuracy for words that appeared during training more than for untrained words.

When comparing each stimulus type to the other stimulus types, the pre-test score of the TWTS stimuli was significantly lower than the post-test scores of all other stimulus types. For the TWUS stimuli, the pre-test score was higher than the post-test score of the TWTS stimuli and the UWUS stimuli. In addition, the post-test score of the TWUS stimuli was significantly higher than the pre-test score of the UWTS stimuli. The pre-test score of the UWUS stimuli was not significantly different from the post-test scores of any other stimulus types.

3.3.2 Identification task

3.3.2.1 Descriptive statistics

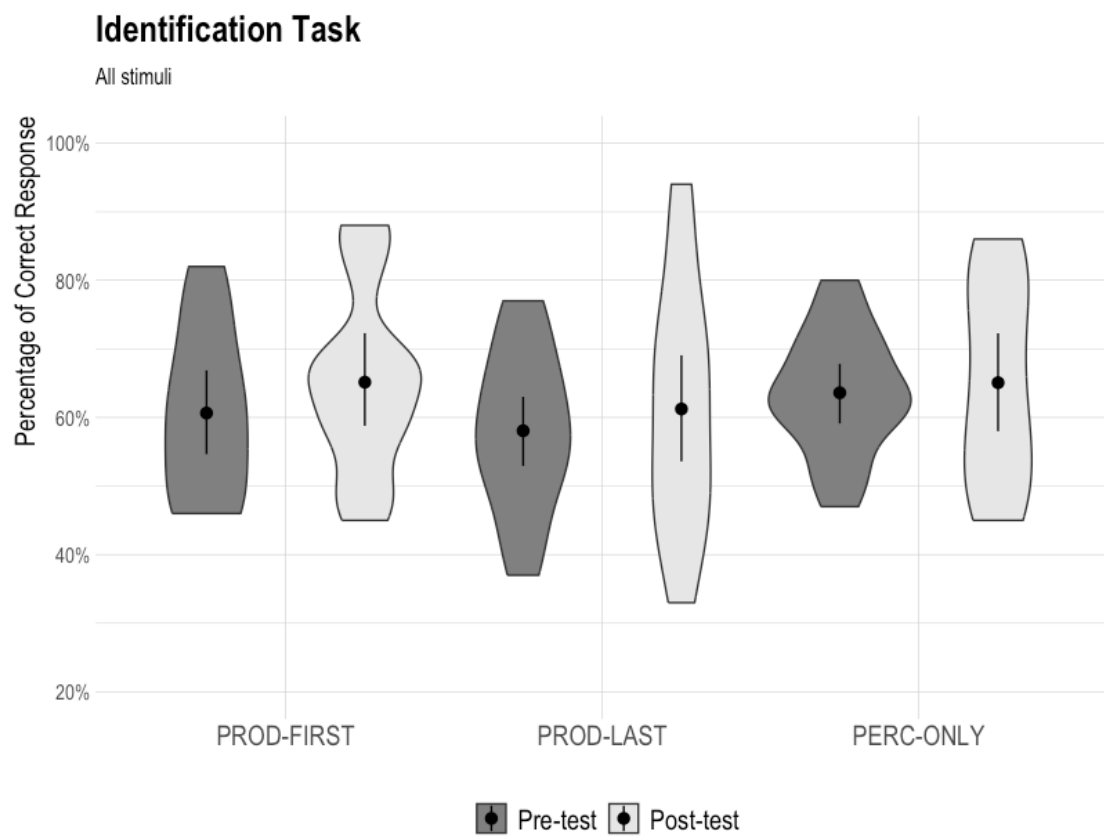
Table 3.6 shows the mean scores of correct responses, standard deviations, and the percentage of correct responses in the identification tasks per training group and familiarity type of stimuli. The identification score was out of 100 and combined the scores of four stimulus types. The TWTS stimuli and the TWUS stimuli had both a maximum of 40 points, and the UWTS stimuli and the UWUS stimuli both had a maximum of 10 points. First, the results of all stimuli combined are presented, followed by each stimulus type one by one below.

Table 3.6 Experiment 1: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-Tests of the Identification Task

		PROD-FIRST		PROD-LAST		PERC-ONLY	
		Mean (<i>S.D.</i>)	Percentage	Mean (<i>S.D.</i>)	Percentage	Mean (<i>S.D.</i>)	Percentage
All stimuli	Pre-test	60.64 (12.22)	60.6%	58.06 (11.62)	58.1%	63.60 (8.97)	63.6%
	Post-test	65.14 (13.29)	65.1%	61.24 (17.64)	61.2%	65.07 (14.50)	65.1%
Trained words x Trained speaker	Pre-test	22.71 (5.25)	56.8%	22.18 (5.14)	55.4%	24.07 (3.56)	60.2%
	Post-test	25.79 (4.51)	64.5%	23.47 (7.17)	58.7%	25.80 (4.51)	64.5%
Trained words x Untrained speaker	Pre-test	25.14 (5.86)	62.9%	23.82 (5.25)	59.6%	26.60 (4.70)	66.5%
	Post-test	27.00 (6.25)	67.5%	25.06 (7.81)	62.6%	27.13 (8.27)	67.8%
Untrained words x Trained speaker	Pre-test	6.57 (1.60)	65.7%	6.29 (1.65)	62.9%	6.73 (1.16)	67.3%
	Post-test	6.00 (2.22)	60.0%	6.35 (2.18)	63.5%	6.20 (1.93)	62.0%
Untrained words x Untrained speaker	Pre-test	6.21 (1.63)	62.1%	5.76 (1.82)	57.6%	6.20 (1.66)	62.0%
	Post-test	6.36 (2.06)	63.6%	6.35 (2.42)	63.5%	5.93 (1.94)	59.3%

Similar to the discrimination task results, participants performed better on the untrained stimuli compared to trained stimuli that were used for training. Moreover, all groups achieved better scores on stimuli spoken by an untrained speaker, suggesting that the stimuli spoken by the untrained speaker was more challenging than the trained speaker that appeared during training. This consistency across both tasks highlights the general difficulty of the trained stimuli that were used during training prior to training.

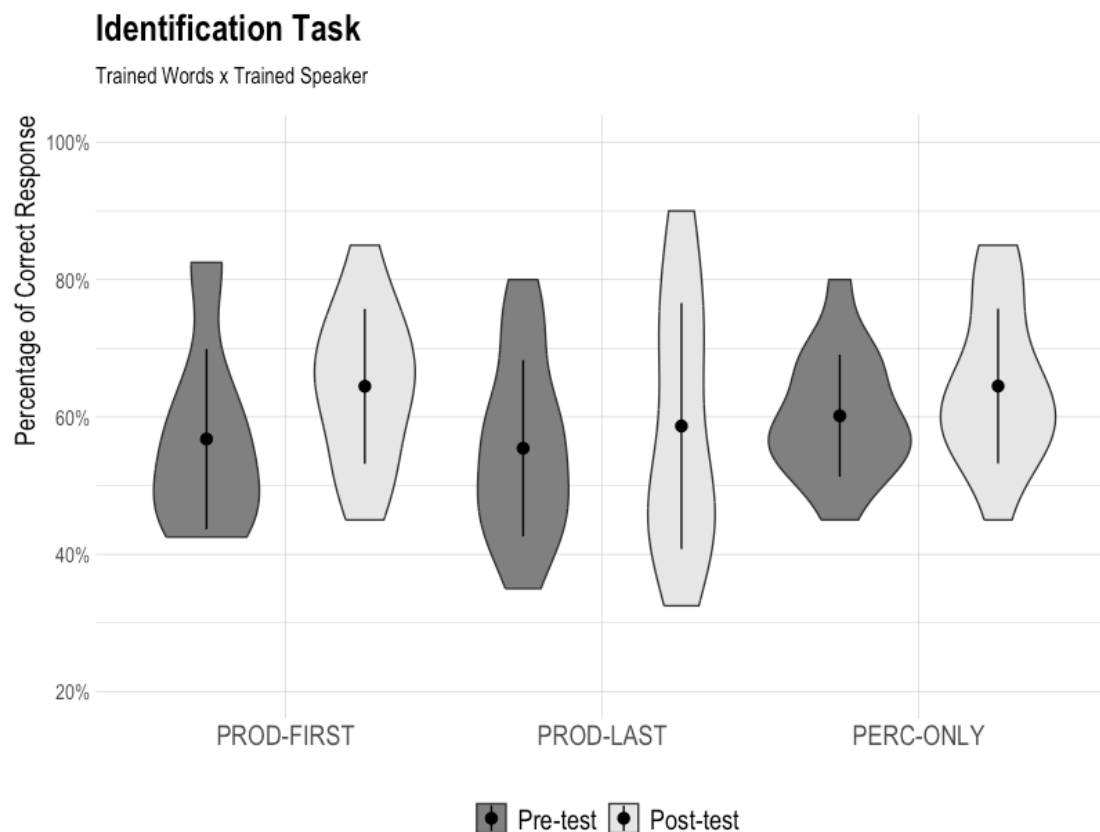
Figure 3.7 illustrates the identification task results of the three groups. The PERC-ONLY group received the highest score (63.6%) among the three groups at pre-test. The PROD-FIRST group and PROD-LAST group received close scores, 60.6% and 58.1%, respectively. After training, the PROD-FIRST group increased by 4.5%, which was the largest improvement of the three groups. The PERC-ONLY group also reached 65.1% accuracy at post-test, but they only increased by 1.5%. The PROD-LAST group increased 3.1%.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 3.7 Experiment 1: Identification Task Results with All Stimuli

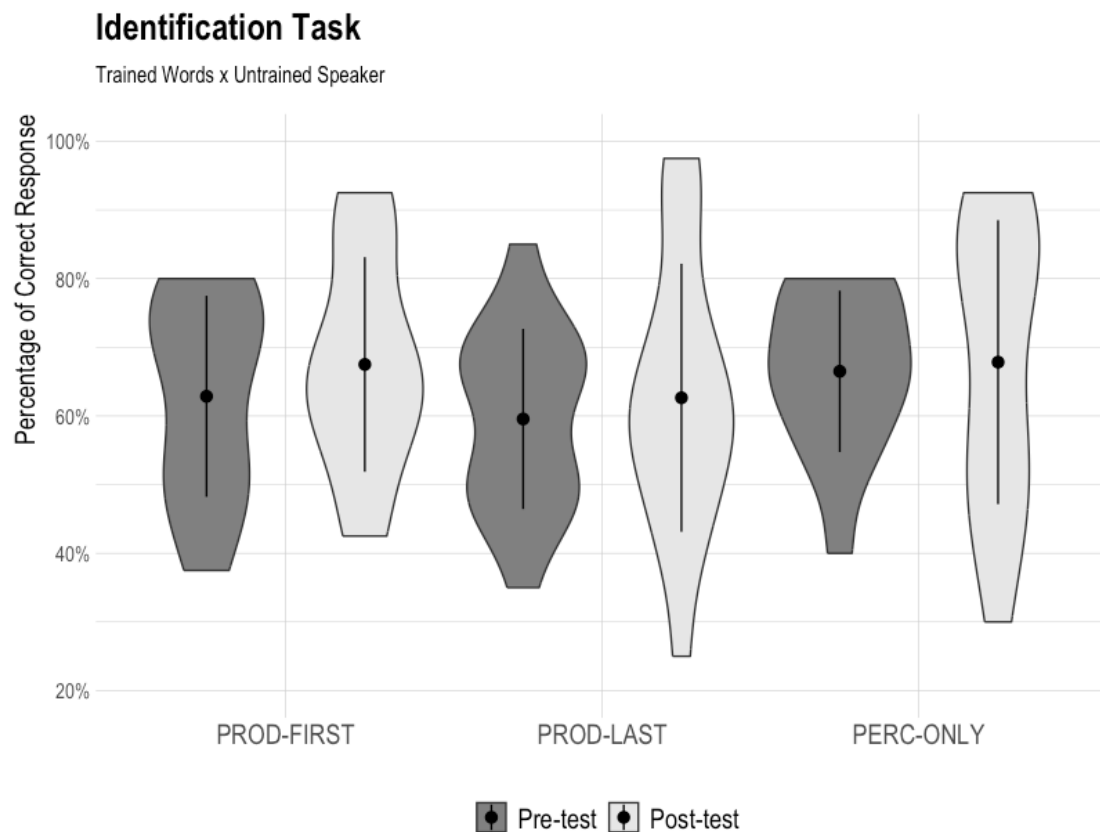
Figure 3.8 describes the pre- and post-test results for the TWTS stimuli. At pre-test, the PERC-ONLY group received 60.2% accuracy, which was the highest among the three groups. The PROD-FIRST group attained 56.8% accuracy, which was followed by the PROD-LAST group's 55.4%. Both the PROD-FIRST group and PERC-ONLY group achieved 64.5% at post-test. The PROD-FIRST group increased by 7.7%, which was a larger improvement than the PERC-ONLY group's 4.3% improvement. The PROD-LAST group increased by 3.3% from pre-test to post-test.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 3.8 Experiment 1: Identification Task Results with the TWTS Stimuli

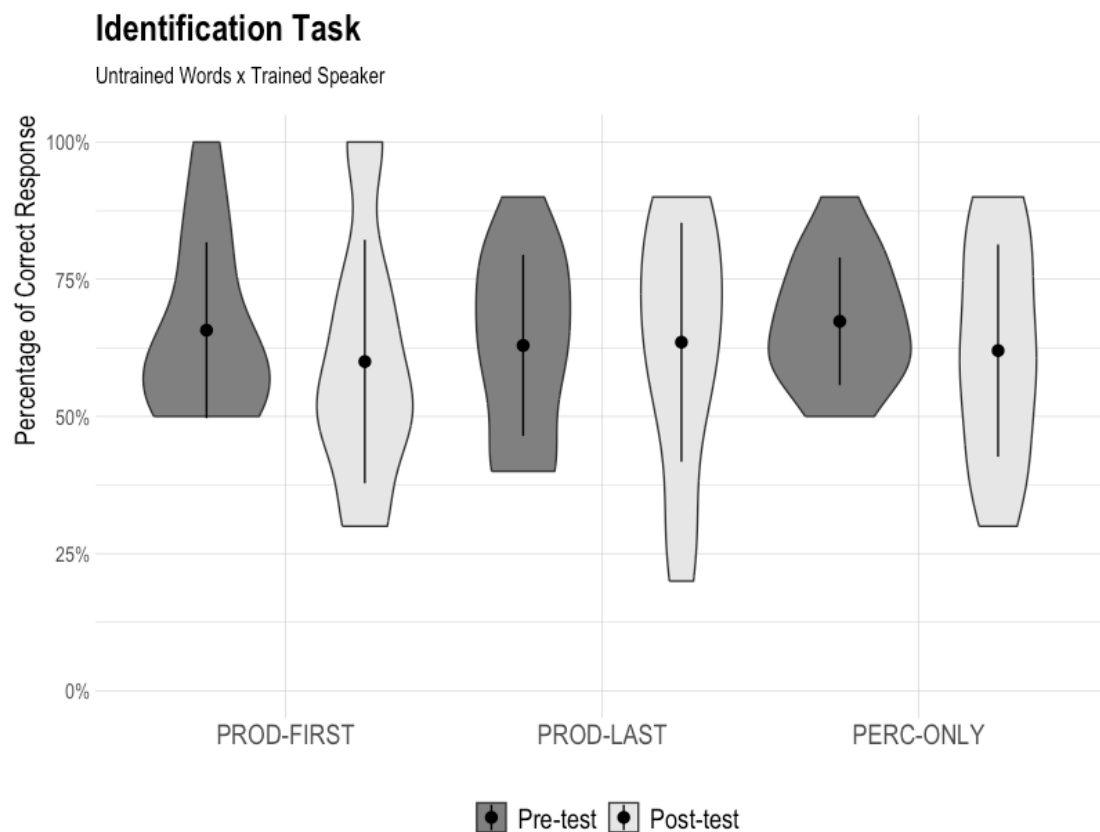
Figure 3.9 depicts the identification test results of the TWUS stimuli. The PERC-ONLY group received 66.5% accuracy at pre-test, which was the highest of the three groups. The PROD-FIRST group attained 62.9%, followed by the PROD-LAST group's 59.6% accuracy. After training, the PROD-FIRST group's accuracy increased the most among the three groups (4.6% increase). The PROD-LAST group improved accuracy by 3%. The PERC-ONLY group improved only 1.3%.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 3.9 Experiment 1: Identification Task Results with the TWUS Stimuli

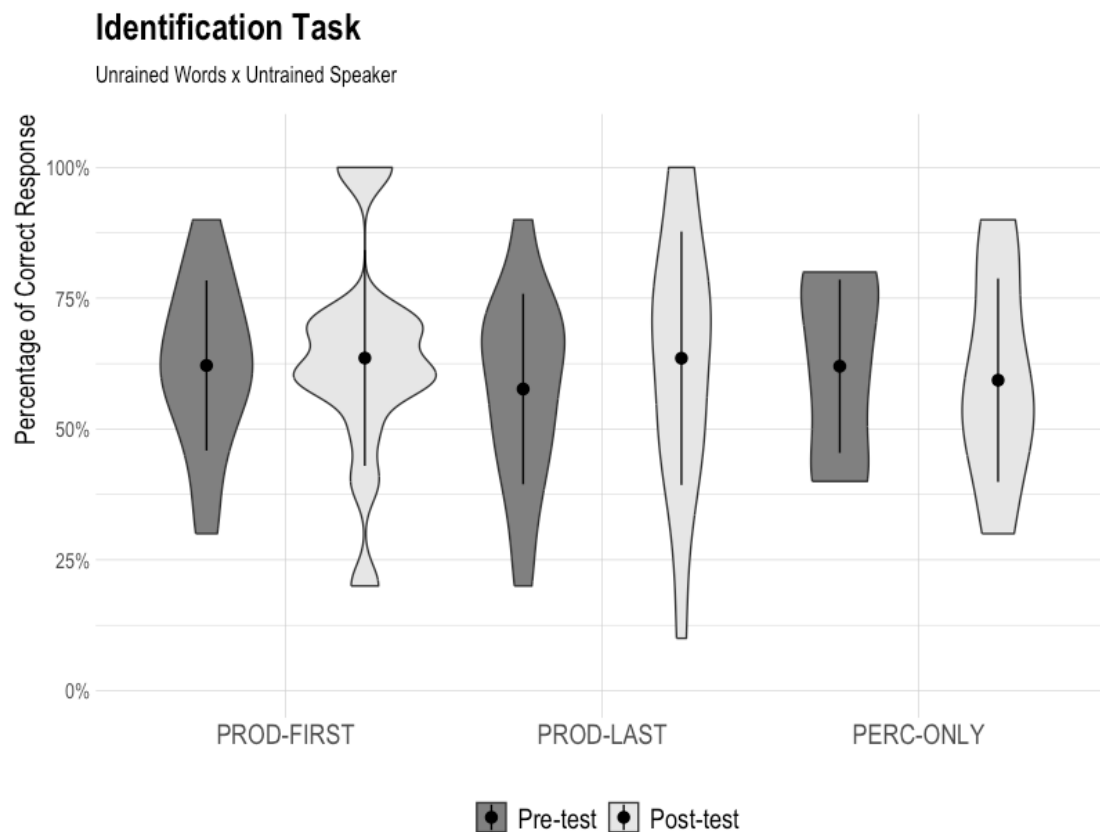
Figure 3.10 illustrates the identification test results of the UWTS stimuli. The PERC-ONLY group attained 67.3% accuracy, followed by the PROD-FIRST group (65.7%). The PROD-LAST group received the lowest pre-test score (62.9%). After training, the PROD-LAST group's accuracy score increased by 0.6%. In contrast, the PROD-FIRST group and PERC-ONLY group decreased their accuracy scores by 5.7% and 5.3%, respectively.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 3.10 Experiment 1: Identification Task Results with the UWTS Stimuli

Finally, Figure 3.11 demonstrates the identification test results of the UWUS stimuli. The PROD-FIRST group and PERC-ONLY group achieved close accuracy scores (62.1% and 62.0% respectively). The PROD-LAST group received the lowest pre-test score (57.6%). After training, the PROD-LAST group's accuracy score increased 5.9%, followed by the PROD-FIRST group's 1.5%. The PERC-ONLY group decreased their accuracy score by 2.7%.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 3.11 Experiment 1: Identification Task Results with the UWUS Stimuli

3.3.2.2 Inferential statistics

Table 3.7 displays the model output of a logistic mixed effects model with the PERC-ONLY group as the reference level for all stimuli. The three-way interaction was not significant and was removed from the final model. A Time and Group interaction was not significant either, but an interaction between Time and Stimuli was significant (UWTS: $p = 0.017$) and Time ($p = 0.001$) was a significant predictor, suggesting that participants as a

whole significantly improved their identification accuracy for the TWTS stimuli, and the improvement was significantly larger than for the UWTS stimuli. Stimuli was a significant predictor (TWUS: $p < 0.001$; UWTS: $p = 0.040$), indicating that participants answered significantly more correctly to the TWUS and UWTS stimuli than the TWTS stimuli at pre-test.

Table 3.7 Experiment 1: Model Output of a Logistic Mixed Effects Model for the Identification Task

<i>Predictors</i>	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	1.39 **	1.11 – 1.75	0.004
time [post]	1.25 **	1.09 – 1.44	0.001
stimuli [TWUS]	1.28 ***	1.12 – 1.47	<0.001
stimuli [UWTS]	1.45 *	1.02 – 2.08	0.040
stimuli [UWUS]	1.16	0.81 – 1.65	0.423
time [post] × stimuli [TWUS]	0.92	0.75 – 1.12	0.405
time [post] × stimuli [UWTS]	0.68 *	0.50 – 0.93	0.017
time [post] × stimuli [UWUS]	0.87	0.63 – 1.18	0.364

Random Effects

σ^2	3.29
τ_{00} word	0.16
τ_{00} participant	0.32
ICC	0.13
$N_{\text{participant}}$	46
N_{word}	50
Observations	9200
Marginal R^2 / Conditional R^2	0.005 / 0.132

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

An interaction between Group and Stimuli was not significant, indicating that all groups performed similarly across all stimulus types. Group as a main effect was not a significant predictor either, indicating that there were no significant differences in identification performances among groups.

Because Time was a significant predictor, pairwise comparisons between Time and Group were performed and revealed that none of the groups showed a significant improvement from pre-test to post-test (PROD-FIRST: $\beta = -0.147$, $SE = 0.074$, $z = -1.995$, $p = 0.345$; PROD-LAST: $\beta = -0.221$, $SE = 0.082$, $z = -2.581$, $p = 0.102$; PERC-ONLY: $\beta = -0.070$, $SE = 0.079$, $z = -0.875$, $p = 0.952$).

Table 3.8 presents the results of pairwise comparisons between Time and Stimuli. The results revealed a significant difference between the pre-test scores of the TWTS stimuli and that of the TWUS stimuli ($p = 0.010$), indicating that participants answered more correctly to the trained words stimuli spoken by the untrained speaker compared to those spoken by a trained speaker before training. Additionally, there was a significant improvement from pre-test to post-test only for the TWTS stimuli ($p = 0.029$), as well as a significant score difference between the pre-test scores of the TWTS stimuli and the post-test scores of the TWUS stimuli ($p < 0.001$). These findings suggest that participants significantly improved their identification accuracy for the stimuli they practiced with, and they achieved a higher score for the stimuli spoken by an untrained speaker compared to the pre-test score of those spoken by a trained speaker.

Table 3.8 Experiment 1: Summary of Pairwise Comparisons between Time and Stimuli for the Identification Task

<i>Pre-test</i>	<i>Pre-test</i>	β	<i>SE</i>	<i>z</i>	<i>p</i>
TWTS	TWUS	-0.250	0.071	-3.533	0.010
	UWTS	-0.374	0.182	-2.053	0.446
	UWUS	-0.145	0.181	-0.801	0.993
TWUS	UWTS	-0.124	0.182	-0.680	0.998
	UWUS	0.105	0.181	0.581	0.999
UWTS	UWUS	0.229	0.144	1.588	0.758
<i>Post-test</i>	<i>Post-test</i>				
TWTS	TWUS	-0.166	0.072	-2.305	0.291
	UWTS	-0.216	0.181	-1.196	0.933
	UWUS	0.000	0.181	0.000	1.000
TWUS	UWTS	0.176	0.182	0.971	0.979
	UWUS	0.166	0.182	0.914	0.985
UWTS	UWUS	-0.010	0.144	-0.072	1.000
<i>Pre-test</i>	<i>Post-test</i>				
TWTS	TWTS	-0.227	0.071	-3.212	0.029
	TWUS	-0.393	0.071	-5.503	<.0001
	UWTS	0.010	0.181	0.057	1.000
	UWUS	-0.227	0.181	-1.253	0.916
TWUS	TWTS	0.023	0.071	0.322	1.000
	TWUS	-0.143	0.072	-1.984	0.493
	UWTS	0.033	0.181	0.184	1.000
	UWUS	0.023	0.181	0.127	1.000
UWTS	TWTS	0.147	0.182	0.806	0.993
	TWUS	-0.019	0.183	-0.103	1.000
	UWTS	-0.072	0.143	-0.502	1.000
	UWUS	0.147	0.145	1.015	0.972
UWUS	TWTS	-0.082	0.181	-0.454	1.000
	TWUS	-0.248	0.181	-1.369	0.872
	UWTS	0.157	0.145	1.087	0.960
	UWUS	-0.082	0.143	-0.573	0.999

3.3.3 Summary of the results on perceptual learning

The Time $\times$ Group interaction was significant in the discrimination task. The PERC-ONLY group showed a significantly larger improvement in discrimination accuracy from pre-test to post-test compared to the PROD-LAST group. Pairwise comparisons revealed that

the PERC-ONLY group and the PROD-FIRST group significantly improved their discrimination accuracy after training, whereas the PROD-LAST group did not show significant improvement.

Regarding performance differences among stimulus types, significant improvements were observed for the TWTS and TWUS stimuli after training, while improvements for the UWTS and UWUS stimuli were not significant. In addition, participants generally performing better on stimuli spoken by an untrained speaker compared to a trained speaker both before and after training across all groups.

In the identification task, the Time $\times$ Group interaction was not significant, but there was a significant Time effect, indicating that participants as a whole significantly improved after training. A post-hoc analysis showed none of the groups significantly improved from pre-test to post-test. However, the Time and Stimuli interaction was significant, and pairwise comparisons revealed that only the TWTS stimuli were significantly improved after training. Similar to the discrimination task, participants generally performed better on stimuli spoken by an untrained speaker both before and after training.

3.3.4 Relationship between PSTM and perceptual learning

First, an overview of participants' performances in the PSTM task is provided. Next, the results of a linear regression model examining the relationship between PSTM and discrimination accuracy prior to training are presented. Following this, the results of the second linear regression model with the gain score from the discrimination task as the dependent variable are discussed. Finally, the results of the identification task data are provided in the same order.

3.3.4.1 PSTM task scores

Table 3.9 shows the mean scores, standard deviations, and minimum and maximum scores of the PSTM task, and Figure 3.12 depicts the PSTM score of each group. The mean score of the PROD-FIRST was 93.0, the PROD-LAST group 96.6, and the PERC-ONLY group 91.1, out of a maximum of 170 points. A linear regression model with the PSTM score as the dependent variable and Group as the independent variable revealed that there was no significant difference among the three groups ($\beta = -1.979$, $SE = 8.132$, $t = -0.243$, $p = 0.809$).

Table 3.9 Experiment 1: Mean Scores, Standard Deviations, and Minimum and Maximum Scores of the PSTM Task

	Mean	<i>S.D.</i>	Min.	Max.
All groups	93.6	20.7	53.3	144.0
PROD-FIRST	93.0	18.5	65.0	126.5
PROD-LAST	96.6	21.1	64.2	144.0
PERC-ONLY	91.1	23.3	53.3	135.5

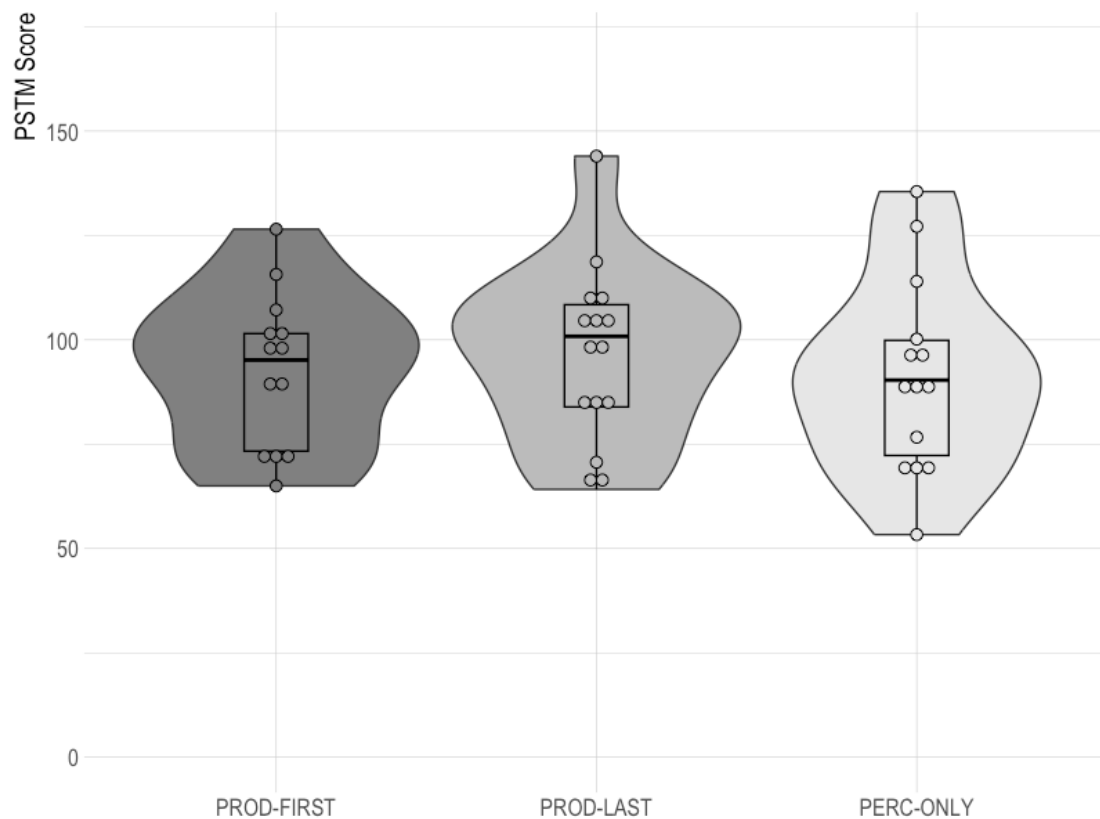


Figure 3.12 Experiment 1: PSTM Score per Training Group

3.3.4.2 PSTM and discrimination task score

Figure 3.13 shows the correlation between the PSTM score and the discrimination task score at pre-test. A linear regression model was fit to pre-test data with the discrimination task score as the dependent variable and the PSTM score as the independent variable. The model showed that PSTM was not a significant predictor of the discrimination accuracy before training ($\beta = -0.034$, $SE = 0.059$, $t = -0.573$, $p = 0.570$).

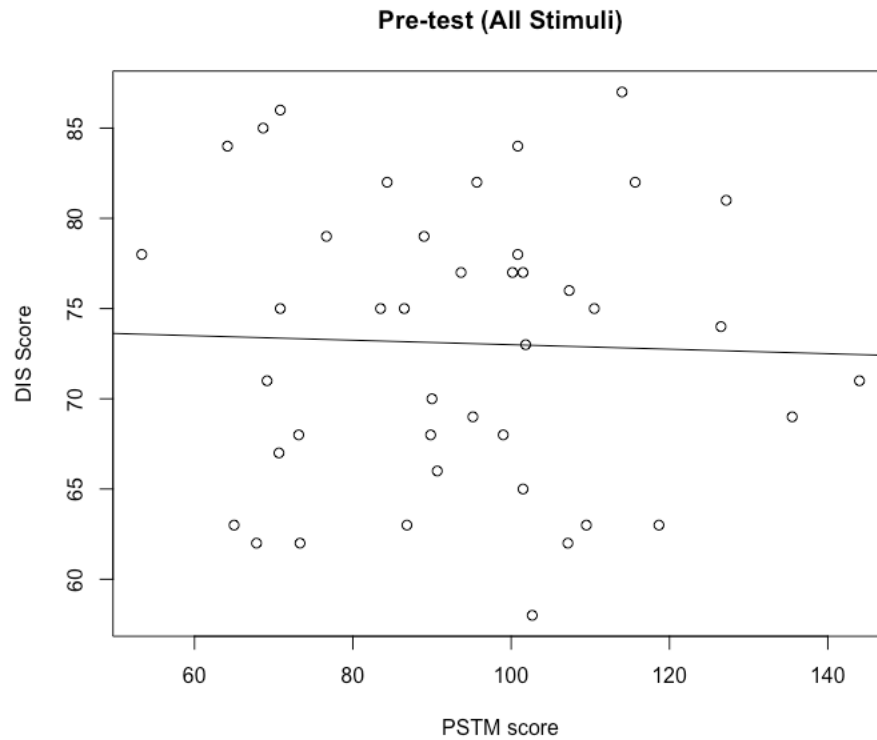


Figure 3.13 Experiment 1: Correlation between the PSTM Score and the Discrimination Task Pre-Test Score

Next, Figure 3.14 depicts the correlation between the PSTM score and the gain score of all stimuli across groups. A linear regression model was fit with the gain score as the dependent variable and the PSTM score, Pre-test score, and Group. The model revealed that neither the main effect of PSTM score nor any interactions involving PSTM were significant ($p > 0.05$). This suggests that PSTM score was not a significant predictor of gain score over time across any group, and that the relationship between PSTM and discrimination accuracy did not change as a result of training.

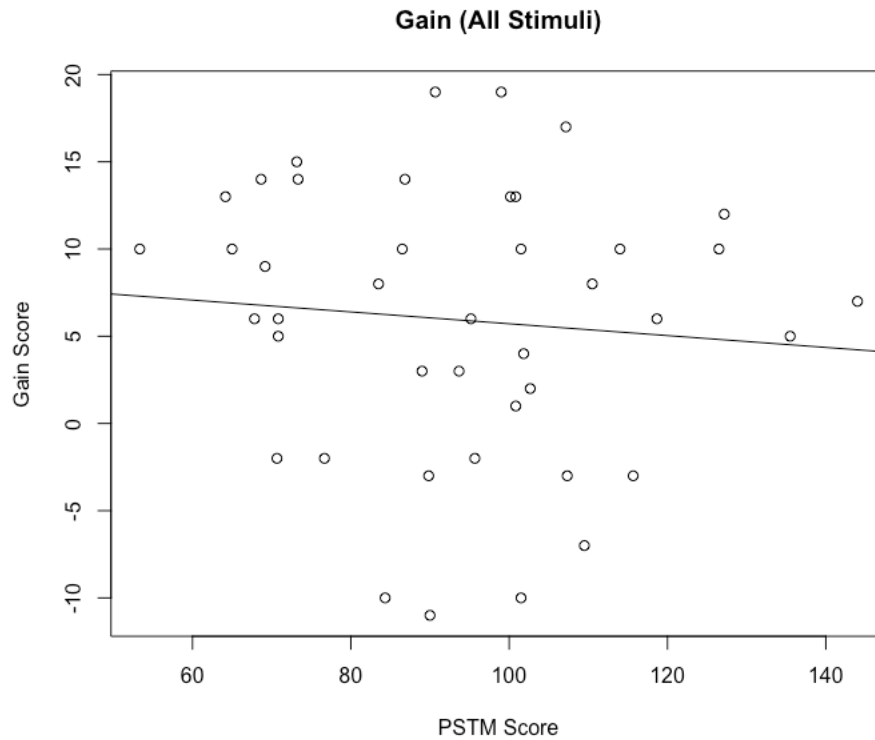


Figure 3.14 Experiment 1: Correlation between the PSTM Score and the Gain Score of the Discrimination Task

3.3.4.3 PSTM and identification task score

Figure 3.15 shows a correlation between the PSTM score and the identification task score at pre-test. A linear regression model was fit to pre-test data with the identification task score as the dependent variable and the PSTM score as the independent variable. The model showed that the PSTM was not a significant predictor of the identification accuracy before training ($\beta = -0.025$, $SE = 0.086$, $t = -0.295$, $p = 0.769$).

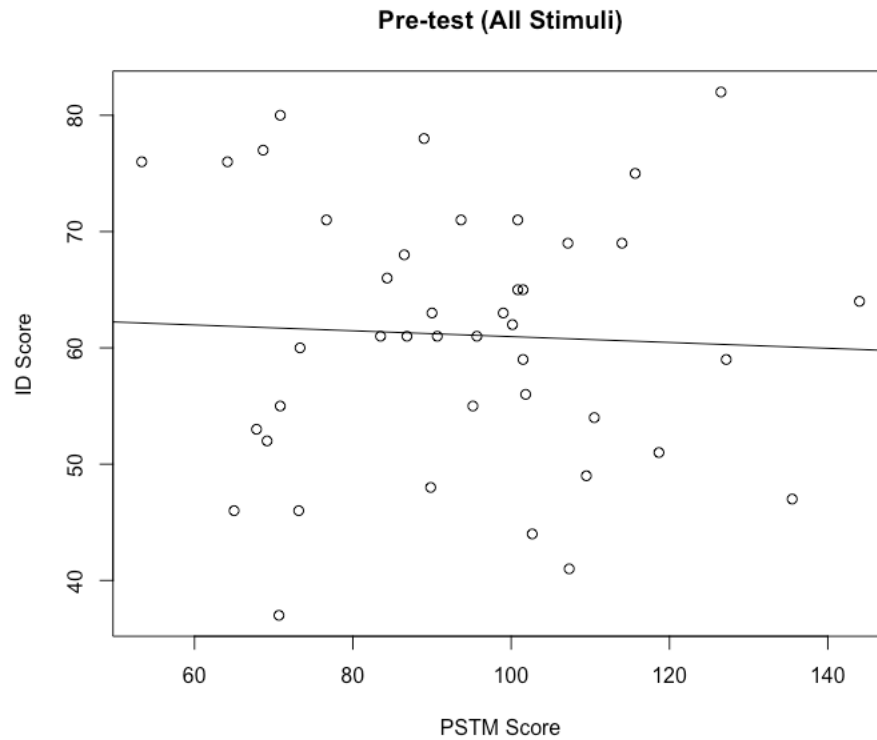


Figure 3.15 Experiment 1: Correlation between the PSTM Score and the Identification Task Pre-Test Score

Next, Figure 3.16 depicts a correlation between the PSTM score and the gain score of all stimuli across groups. A linear regression model was fit with the gain score as the dependent variable and the PSTM score, Pre-test score, and Group as predictors. The model revealed that there were no significant three-way or two-way interactions between the predictors, suggesting that the relationship between PSTM and gain score did not depend on participants' initial perceptual accuracy or training type.

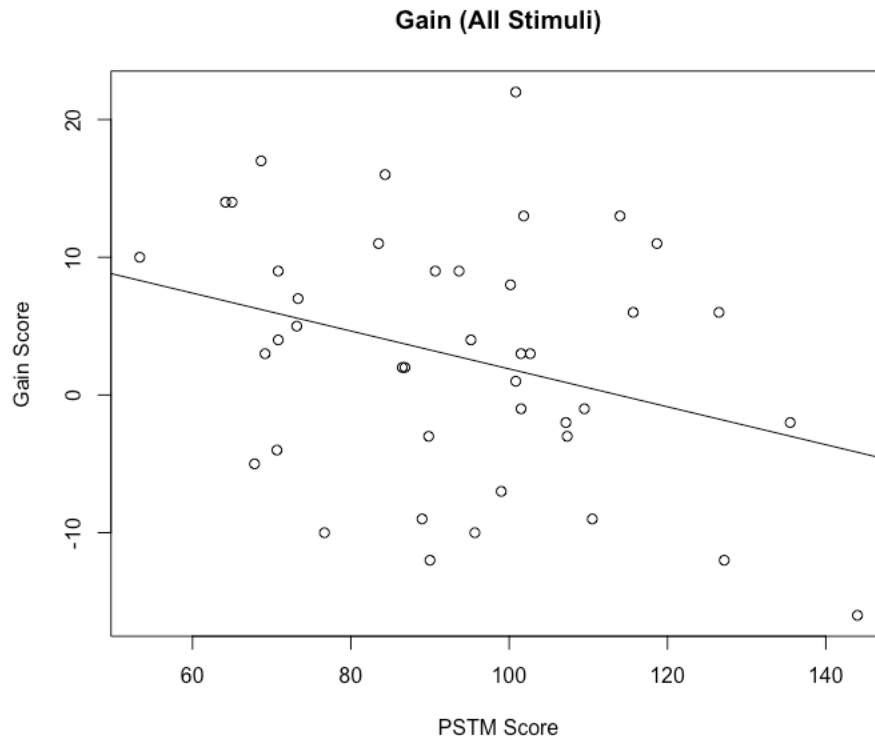


Figure 3.16 Experiment 1: Correlation between the PSTM Score and the Gain Score of the Identification Task

Furthermore, the main effects of the pre-test score and Group were not significant, indicating that neither participants' initial perceptual accuracy nor their training type was a strong predictor of the gain score. This suggests that differences in initial identification accuracy or training type did not significantly influence participants' improvement after training.

The only significant predictor was the PSTM score ($\beta = -0.138$, $SE = 0.065$, $t = -2.114$, $p = 0.041$). However, it was a negative relationship, suggesting that participants with higher PSTM scores did not benefit as much from training compared to those with lower PSTM scores.

Because this negative correlation might have attributed to the data from a participant who received a high PSTM score (144 points) but a negative gain score (-16 points), another model was fit without the outlier². The model showed no significant interactions or predictors including the PSTM score ($\beta = -0.097$, $SE = 0.069$, $t = -1.399$, $p = 0.170$), suggesting that

² This outlier was identified through a visual inspection of the scatter plot showing the relationship between the PSTM score and the gain score of the identification task, which is a common way to identify outliers in second language research (Larson-Hall, 2015).

PSTM was not a significant predictor of the degree of perceptual improvement in identification of the English /l/–/r/ contrast.

3.4 Discussion

This section begins with a summary of the results, followed by a discussion of the findings presented in the previous section and a comparison with those of previous studies on L2 speech learning. The discussion addresses each research question in turn.

In the discrimination test, the PERC-ONLY group significantly improved discrimination accuracy after training and this improvement was significantly larger compared to that of the PROD-LAST group, which did not show significant improvement. The PROD-FIRST group also improved significantly. In terms of stimulus type, significant gains were observed for the TWTS and TWUS stimuli, while no significant improvements were observed for the UWTS and UWUS stimuli. Across all groups, participants generally performed better on stimuli spoken by an untrained speaker both before and after training.

In the identification task, there was no significant interaction between Time and Group, but there was a significant interaction between Time and Stimuli, along with a main effect of Time. However, none of the groups showed significant improvements from pre-test to post-tests, suggesting overall improvement in identification of the TWTS stimuli after training. Again, participants generally demonstrated higher accuracy for stimuli spoken by an untrained speaker.

Regarding the relationship between PSTM and perceptual learning, PSTM did not significantly predict discrimination and identification accuracy before training. However, after training, a significant negative relationship between PSTM and identification accuracy was observed. This negative relationship, however, disappeared after excluding an outlier that might have skewed the results.

(1) Does perceptual training using discrimination tasks improve L2 phonetic perception?

The results of the discrimination task showed that four sessions of PERC-ONLY training in a discrimination task format can improve L2 learners' accuracy in discriminating the English /l/–/r/ contrast. These findings are in line with many previous studies (e.g., Lively et al., 1994; Nishi & Kewley-Port, 2007; Shinohara & Iverson, 2018) and suggest that exposure to target phonemes can improve perceptual learning. However, it is important to

note that the degree of their learning was around 9% (all stimuli: 8.6%, TWTS: 9.7%, and TWUS: 8.2%), which was lower than other studies. Shinohara and Iverson (2018) mentioned that the average of perceptual learning after perceptual training is about 15%. Moreover, their post-test score of the discrimination task was 81.9%, suggesting more room for improvement and indicating that participants could not construct robust phonetic categories of the English /l/ and /r/ sounds. A possible explanation for this might be a lack of sufficient exposure to the target sounds.

Some studies showed similar findings due to the small number of trials in training. For example, Ueda and Hashimoto's (2019) study, where training consisted of 288 trials in total and lasted an hour for six phonemes, did not observe significant improvement for the English /l/ and /r/ sounds. Qian et al.'s (2018) study, where participants received 10 to 100 minutes of training on 12 phonetic contrasts, did not show robust improvement in untrained stimuli. Conversely, studies with longer training showed more robust learning. Bradlow et al. (1997) conducted perceptual training on the English /l/-/r/ for 45 sessions and found that the trained group significantly outperformed the control group at the post-test with the untrained stimuli recorded by a trained speaker and untrained speaker. In the current study, training consisted of 400 trials and took about an hour for two phonemes, which means training was less intensive than Bradlow et al.'s. Therefore, the participants in this study might not have been exposed to the target phonemes sufficiently enough to construct robust phonetic categories and the learning effect was not as strong as other perceptual training studies.

Despite the improvement in discrimination accuracy, the results showed that PERC-ONLY training using a discrimination task format did not improve L2 learners' identification accuracy. These findings indicate that although the participants' perception significantly improved for trained words spoken by both trained untrained speakers when measured with the discrimination task, such improvement was not observed in the identification task. These findings are in line with some previous studies (e.g., Grenon et al., 2018; Law et al., 2019).

One possible reason for this is that the participants only became able to detect the difference between the target sounds but not to label the stimuli. In a discrimination task, participants can answer correctly as long as they can distinguish acoustic differences between a set of stimuli. In an identification task, on the other hand, they need to match the stimuli to their corresponding spelling. Considering the different natures of the two tasks, identification training might be necessary to learn phoneme-grapheme correspondences. Indeed, Grenon et

al. (2018) found that 25% of participants mislabelled the target phonemes after discrimination training, and they suggested that discrimination training can be useful when accompanied by instructions on a phoneme-grapheme association. In the current study, participants did not have a chance to learn the association of letters and sounds. Therefore, it might be the case that the participants did improve the perception of the target sounds but did not improve their identification skills due to mislabelling issues.

Another reason for the discrepancy between the two tasks is task familiarity. Since the training was only conducted in a discrimination task format, the participants were more familiar with the discrimination task and might have completed the discrimination task at the post-test more confidently. In contrast, they were presented with the identification task only at the pre- and post-tests and the novel task may have caused the identification task to be more cognitively demanding.

Alternatively, the overall learning from the training may not have been significant as discussed above, and this could explain the lack of improvement in the identification task. In any cases, it remains unclear why learning did not transfer to identification skills. Thus, more research is needed to determine whether participants failed to become able to label each stimulus correctly due to being trained solely in a discrimination format, whether task familiarity influenced the results, or whether other factors, such as the modest improvement rate, contributed to these findings.

(2) Does production during perceptual training disrupt L2 phonetic perceptual learning?

The major findings to emerge were that the PROD-FIRST training led to significant improvement in the discrimination task. This provides evidence that disruption to perceptual learning due to the production of target phonemes was attenuated. This result is incompatible with studies conducted by Baese-Berk (Baese-Berk, 2010, 2019; Baese-Berk & Samuel, 2016), where their participants who had PROD-FIRST training did not show learning. This incompatibility might be attributed to a difference in participants' initial familiarity with the target phonemes. In Baese-Berk (2010, 2019), L1 American English speakers learned synthesised pre-voiced /d/ and short-lag /t/, and in Baese-Berk and Samuel (2016), participants' L1 was American English, and their target contrast was the Basque /s/–/ʃ/. In these studies, participants were not familiar with the target sounds before training except for Baese-Berk and Samuel's (2016) Experiment 2.

In the present experiment, however, participants were Japanese learners of English, and the target contrast was the English /l/–/r/. They had studied English for at least six years in school before training. Thus, the training sessions were not the first encounter with the target phonemes for the participants, although they had not been exposed to the target phonemes enough to have robust representations because the English /l/–/r/ is an extremely challenging contrast for L1 Japanese speakers (Aoyama et al., 2004; Strange & Dittmann, 1984) and English is a foreign language in Japan. Baese-Berk and Samuel (2016) found that learners’ prior exposure to the target language might alleviate the disruption to perceptual learning; consequently, for the participants in the current experiment, the prior exposure to the target phonemes might have facilitated their perceptual learning better.

A second important finding of this experiment is that the disruption to perceptual learning from production practice was detected for PROD-LAST training in the discrimination task. This suggests that disruption occurs when learners repeat the stimulus aloud after answering the discriminating question. This finding is contrary to that of Thorin et al. (2018), who found that disruption was not caused by repeating stimuli when learners did it after answering the questions. This discrepancy could be due to a difference in participants’ proficiency in the target language and the training designs.

As Baese-Berk and Samuel (2016) found that prior exposure to the target language might alleviate disruption to perceptual learning, the difference between participants’ proficiency in the target language might have played a role. More specifically, it can be presumed that the higher the learner’s proficiency in the target language, the less likely disruption to perceptual learning will occur. The participants in Thorin et al. (2018) were upper-intermediate EFL learners, while the present experiment focused on beginner to lower-intermediate learners. Therefore, disruption might have been alleviated in Thorin et al. (2018) but not in this experiment.

Another possible explanation is the different degrees of phonetic difficulty. In Thorin et al. (2018), L1 Dutch speakers learned the English /æ/–/ɜ/ vowels in training. Considering the difficulty of the contrast, Dutch has the /ɜ/ sound, which is a good exemplar of the English /ɜ/. In contrast, there is no /æ/ in Dutch, so /æ/ is perceived as /ɜ/. Therefore, this pattern is considered the Category-Goodness assimilation according to PAM-L2 (Best, 1995; Best & Tyler, 2007). In contrast, L1 Japanese speakers learned English /l/ and /r/ in this study. These sounds are perceived as Japanese flap /r/, which is not a good exemplar for either of the sounds. Therefore, it is the Single-Category assimilation. Based on PAM-L2, the

Single-Category assimilation is more difficult than the Category-Goodness assimilation, so it is assumed that the participants in this experiment learned more challenging phonetic contrast.

Another possibility is limited exposure. In Thorin et al. (2018), participants received four sessions of 200 trials, while in this experiment, participants received four sessions of 100 trials. That is, the number of trials in this experiment was half that of Thorin et al. (2018). Increasing the number of training trials may be necessary to further investigate the production effect on perceptual learning.

An additional factor to consider is the differences in the training designs between the two experiments. In Thorin et al. (2018), participants listened to a series of words and indicated which word they heard last from the two visually presented words on the screen. In this experiment, however, they listened to three words and then answered whether the word they heard last was the same as the first or the last word. That is, training in Thorin et al. (2018) contained an identification element, whereas this experiment employed a discrimination task regimen. In fact, some researchers argued that identification training is superior to discrimination training (Jamieson & Morosan, 1986; Logan et al., 1991; Strange, 1992; Strange & Dittmann, 1984). Because of the advantage of using an identification element, Thorin et al. (2018) might not have shown disruption to perceptual learning, but disruption occurred in the present experiment.

Although the findings of the current study contradict previous studies regarding the conditions causing disruption to perceptual learning, what is consistent among the studies is that production practice in perceptual training did not bring about an extra benefit on perceptual learning. It is a counter-intuitive outcome, given the assumption that perception and production skills are tightly connected. Indeed, some previous studies showed that production training leads to perceptual learning (Akahane-Yamada et al., 1997; Aliaga-Garcia & Mora, 2009; Herd et al., 2013; Kartushina et al., 2015; Lacabex et al., 2008; Sakai, 2016; Thorin, 2020). Hence, the question rises as to why producing target sounds in perceptual training did not bring about extra benefits in perceptual learning.

Thorin et al. (2018) raised a hypothesis that this null-effect of production may be due to a lack of immediate and informative feedback on production. Perception-and-production-mixed training in previous studies, including the present experiment, did not provide feedback on participants' production. In contrast, studies that have shown positive outcomes

of production training often included some sort of immediate feedback, such as corrective feedback (Saito & Lyster, 2012), spectrograms (Akahane-Yamada et al., 1997; Herd et al., 2013), written feedback from Automatic Speech Recognition systems (e.g., Inceoglu et al., 2023; McCrocklin, 2019), and visual displays from ultrasound imaging (Antolík et al., 2019; Inceoglu & Gnevshcheva, 2020; Kühnert & Pillot-Loiseau, 2022; Noguchi et al., 2015). For example, Saito and Lyster (2012) showed that when teaching the English /r/ through form-focus instruction (FFI), only the group that received corrective feedback significantly improved their pronunciation of the sound but not the group without corrective feedback. Therefore, the absence of immediate feedback might have brought this null effect of production on perceptual learning.

It is important to note, however, that the small sample size in this experiment may have limited the statistical power of the findings, potentially contributing to the null effect of production on perceptual learning. Each group consisted of about 15 participants, which may not have been sufficient to detect subtle training effects or interactions between production practice and perceptual learning outcomes. Therefore, future research with larger sample sizes is needed to verify the findings.

(3a) Is perceptual accuracy prior to phonetic training related to learners' PSTM?

PSTM was not a significant predictor of either their discrimination or identification scores before training. These findings are in line with those of Safronova and Mora (2012) and Ghaffarvand Mokari and Werner (2019), but contradicted many previous studies, which showed that PSTM was a significant predictor (Cerviño-Povedano et al., 2011; Inceoglu, 2019; Kondo, 2019; MacKay et al., 2001).

This disconnection might be attributed to an inconsistency between mora-based items for PSTM tasks and syllable-based training items. In the studies which found a relation between PSTM and L2 perceptual accuracy, syllable-based nonwords were used, and perceptual test items were also syllable-based (Cerviño-Povedano et al., 2011; Inceoglu, 2019; MacKay et al., 2001). In the present experiment, nonwords were created based on mora according to the Japanese phonological system because PSTM is usually measured using participants' L1. However, since the target language was English, the perception test items were syllable-based words. This discrepancy made it complex to interpret the data.

It might be possible that PSTM is more taxed for some L1 Japanese speakers who are not familiar with syllables when they hear syllable-based words. If that is the case, there might have been some participants whose PSTM was taxed heavily when processing English when they were learning English and failed to construct accurate phonetic categories, although they got a higher score on the PSTM task with mora-based items. PSTM tasks are for measuring one's PSTM capacity and are often represented as one index per person; however, when analysing whether PSTM can predict L2 perceptual ability, the difference in phonotactics between participants' L1 and the target language needs to be taken into account carefully.

The PSTM task of the present experiment was created following Kondo's (2019) nonword memory test. She carried out the test with L1 Japanese speakers and found that the PSTM test score was a significant predictor of English listening skills. In her study, participants' listening skill was assessed with a sentence-writing task, which means that there was a possibility that her participants managed to answer better by using their grammatical knowledge rather than the test solely assessing whether they could store English sounds in their PSTM. Furthermore, the effect size was small. Thus, more research on the relation between PSTM and L2 perception is needed in a situation where the phonotactics of learners' L1 is greatly deviant from a target language (e.g., syllable and mora).

(3b) Is the degree of perceptual learning by phonetic training related to learners' PSTM?

PSTM capacity was not a significant predictor of the post-test scores of the discrimination task. These findings contradict some previous studies that reported a relationship between PSTM and perceptual accuracy (Cerviño-Povedano et al., 2011; Inceoglu, 2019; Kondo, 2019; MacKay et al., 2001). In contrast, the results of the identification task revealed a negative relationship between PSTM capacity and the degree of perceptual learning, suggesting that the higher their PSTM score was, the lower the improvement in identification accuracy after training. A possible explanation for this negative correlation is that there was a participant who received an extremely high PSTM score but a low identification score at post-test. This data point might have skewed the result to a negative correlation. In fact, when removing this outlier, results showed that the PSTM score was not a significant predictor of the gain score. However, there were also a few students

who scored low on the PSTM task but received a high identification score after training. To unveil the reason for the negative correlation, more research needs to be conducted.

What is more important here is a positive relationship between the PSTM score and perceptual accuracy was not observed. These findings are in line with Ghaffarvand Mokari and Werner (2019), but not with Aliaga-García et al. (2011). However, PSTM was not a significant predictor of the degree of perceptual learning probably because of the same reason for the lack of relationship between the pre-test scores and the PSTM score, which is that participants' PSTM was taxed heavily when processing syllable-based words during perceptual training. Aliaga-García et al. (2011) claimed that PSTM is related to the degree of L2 learning because L2 learners who have a larger PSTM can retain phonetic information longer, which, in turn, facilitates the storing of information in the long-term memory, eventually leading to the formation of robust phonetic categories. In the present study, however, some participants might have consumed PSTM capacity more when they were processing syllable-based words in training than when processing mora-based nonwords in the PSTM task. Thus, the relationship between PSTM and perceptual learning may not be apparent because of the nature of the experiment.

Another possible reason is the participants' proficiency in a target language. In this experiment, participants were at the lower-intermediate level and had an experience of studying English for at least six years, a profile similar to that of the participants in Ghaffarvand Mokari and Werner (2019). In their study, PSTM was neither related to discrimination accuracy after training nor to the degree of perceptual learning. It might be the case that participants' experience of perceptual learning greatly varied; some participants might have been better at discriminating the sounds than others at the beginning. Therefore, the degree of perceptual learning by those participants might have been smaller. Thus, if the target language is one that participants have never been exposed to prior to training, PSTM might demonstrate a more tightly related relationship with perceptual learning.

It is also important to note that the number of times that participants completed the PSTM task varied because some of them completed the first PSTM task a few times, as mentioned in the Method. Although another set of stimuli was created and participants were asked to conduct the second task at post-test to make sure all participants did the second task once, task familiarity might have influenced the results, potentially causing them to be skewed. Given that there is still little research on individual differences, particularly on

PSTM, in L2 perception (Daidone & Darcy, 2021; Nagle, 2022), more research needs to be conducted on this topic.

3.5 Summary of Experiment 1

Experiment 1 had two main goals: a) to examine the effects of production during perceptual training on L2 perception; and b) to examine the relationship between L2 perception and PSTM. Findings revealed that all groups improved their perceptual scores numerically, but the degree of improvement varied. PERC-ONLY training and PROD-FIRST training significantly improved participants' discrimination ability. PROD-LAST training also led to improvement but not significantly. With regard to identification accuracy, all groups significantly improved without a significant Time and Group interaction. However, none of the groups showed a significant improvement.

The findings shed light on the effects of producing target phonemes during perceptual training and suggest that repeating stimuli after answering a discrimination question leads to disruption in perceptual learning. Yet, although repeating a stimulus before a discrimination question did not cause learning disruption, production practice during perceptual training did not enhance perceptual learning. These results align with Thorin et al. (2018), which found no facilitative effect of production during perceptual training, but differ from Baese-Berk and Samuel (2016), which observed consistent disruption.

Regarding PSTM, the PSTM score was not a significant predictor of perceptual accuracy before training, and none of the groups showed a positive correlation between PSTM and the degree of learning. These findings contradict some previous studies, yet the apparent contradiction might be attributed to differences in phonological systems of the learners' L1 and the target language. In this experiment, the PSTM task used mora-based nonwords, while perceptual training and testing involved syllable-based stimuli, potentially affecting participants' ability to process and retain the target phonetic information.

Chapter 4 Experiment 2: The effects of production during identification training and the relationship between L2 perception and PSTM

4.1 Introduction

Experiment 1 investigated the effects of producing target sounds during perceptual training by adapting the training methods of Baese-Berk and Samuel (2016) and Thorin et al. (2018) and addressing cross-linguistic differences between the two studies with a single group of L1 Japanese learners of English. Results showed that PERC-ONLY training and PROD-FIRST training, where participants repeated the stimulus right after hearing it, significantly improved accuracy scores in the discrimination tasks but not in the identification tasks. One account for why learning in discrimination accuracy did not improve identification accuracy is a mislabelling issue. The discrimination training study conducted by Grenon et al. (2018) found that 25% of the participants mislabelled target sounds after discrimination training, and Wee et al. (2019) reported similar findings. That is, even if learners become able to distinguish target sounds, they might need to learn phoneme-grapheme correspondences; otherwise, they might not be able to identify which phoneme they heard correctly.

In addition, PROD-LAST training, where participants produced the word after answering a perceptual question, did not show as much improvement as the other two training types. This finding is inconsistent with Thorin et al.'s (2018) study that reported that PROD-LAST training was as effective as PERC-ONLY training. These findings suggested that production during perceptual training did not disrupt perceptual learning, unlike some studies showing a disruption effect of production (Baese-Berk, 2010, 2019; Baese-Berk & Samuel, 2016; Wright, 2021). This discrepancy may be attributed to the difference in training task formats. Thorin et al. (2018) employed an identification task format rather than a discrimination task format. Some researchers claimed that identification training is superior to discrimination training because it helps learners tap into their phonetic representations, which leads to phonetic category construction, while discrimination training has learners only compare the acoustic differences (Jamieson & Morosan, 1986). Thus, the superiority of the identification training format might attenuate the disruption to perceptual learning. Thus, Experiment 2 investigates whether production disrupts perceptual learning in identification training.

Following Experiment 1, Experiment 2 also investigates the relationship between PSTM and L2 perceptual learning. In Experiment 1, PSTM was not a significant predictor of

the degree of improvement. As for the accuracy of discriminating target sounds, only the PROD-FIRST group showed a significant association between their trained stimuli score and PSTM score before training. Regarding identification accuracy, PSTM was a significant predictor before and after training, but only for the PROD-FIRST group. However, when considering all groups combined, PSTM was not a significant predictor of either discrimination or identification scores. These findings contradicted some previous studies (Cerviño-Povedano & Mora, 2011; Inceoglu, 2019; Kondo, 2019; MacKay et al., 2001) but in line with (Ghaffarvand Mokari & Werner, 2019; Safronova & Mora, 2012). It is still unclear why PSTM was a significant predictor of the accuracy only before and after the PROD-FIRST training, and more research needs to be conducted.

To further explore the nature of the effects of producing target sounds during perceptual training and the possible advantage of identification training, Experiment 2 used a phonetic training similar to Experiment 1 but implemented as an identification task. The research questions are as follows:

1. Does perceptual training using identification tasks improve L2 phonetic perception?
2. Does production during perceptual training in an identification task format disrupt L2 phonetic perceptual learning?
- 3a. Is perceptual accuracy prior to phonetic training related to learners' PSTM?
- b. Is the degree of perceptual learning by phonetic training related to learners' PSTM?

4.2 Method

4.2.1 Participants

Data collection for this study was conducted at the same university where Experiment 1 was conducted. In total, 110 Japanese learners of English participated in this phonetic training study. Of the participants, those who did not complete the pre- and post-tests of the discrimination and identification tasks and all the training sessions properly were excluded from the analysis. Thus, the data of 63 participants were initially used for analysis. However, the number further decreased to 61 after analysing the pre-test data because two participants received scores over 90 out of 100 on the discrimination and identification tasks. In addition, four participants who did not complete the PSTM task properly (e.g., recording errors) were excluded from analysis for the relationship between PSTM and perceptual learning.

Participants were 18- to 21-year-old undergrad students. Their majors were School Nursing, Art and Calligraphy Education, Mathematics Education, Home Economics Education, and Music Education. All participants were taking an English course covering the same topics and using the same learning materials facilitated by the same instructor. Their English proficiency was A2 to B1 in the CEFR, which is a beginner to lower-intermediate level. None of them reported that they had lived in English-speaking countries for more than three months. The participants who completed all the training sessions and pre- and post-training tasks received extra credits for the English course.

4.2.2 Materials

4.2.2.1 English stimuli for training sessions and tests

The same stimuli from Experiment 1 were used.

4.2.2.2 Stimuli for the PSTM tasks

The same set of stimuli applied in the post-test session in Experiment 1 was used.

4.2.2.3 Language background questionnaire

The same questionnaire from Experiment 1 was used.

4.2.3 Procedure

The experiment was conducted online from June 2021 to July 2021 (during the COVID-19 pandemic) and included a pre-test, four training sessions, and a post-test. The pre- and post-training tasks and perception and production training tasks were created using IBEX farm, a web-based experimental presentation platform. For communicating with participants, online tools such as email, Zoom, and Google Forms were used. Communication between L1 Japanese participants about the research, including instructions for each task, was conducted in their first language, Japanese. This procedure was reviewed and approved by ANU Human Ethics Committee, and participants provided consent.

The timeline of the data collection is summarised in Figure 4.1. Participants were divided into three groups: the Production-first group (PROD-FIRST), the Production-last group (PROD-LAST), and the Perception-only group (PERC-ONLY) based on which English class group they belonged. First, all L1 Japanese participants completed pre-training tasks: a language background questionnaire, a discrimination task, and an identification task. Then, participants received four training sessions based on their group over a month. Once they had completed all the training sessions, the participants completed post-training tasks, a

discrimination task and an identification task. Pre- and post-training tasks were the identical to those in Experiment 1, except participants completed the PSTM task only once individually with the researcher on Zoom during the data collection period to ensure that participants did it only once.

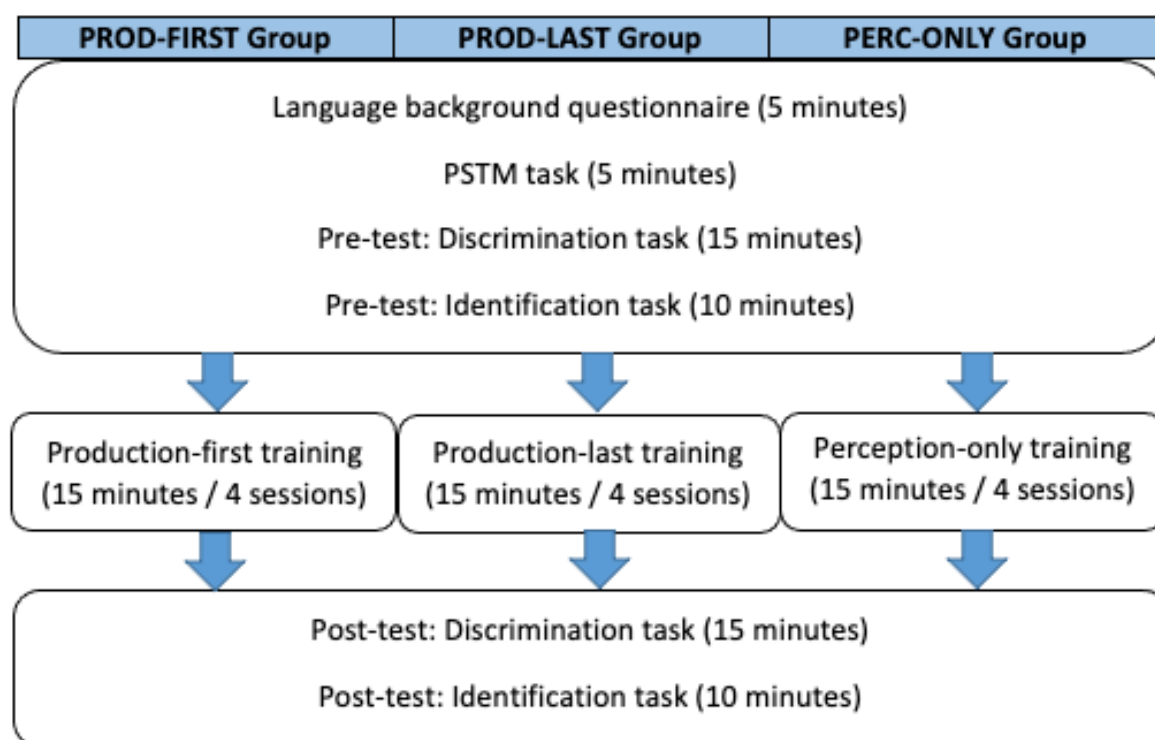


Figure 4.1 Experiment 2: Timeline of Data Collection

4.2.3.1 Phonetic training

The training was designed in an identification task format. In the PROD-FISRT group, participants listened to a word, repeated the last word aloud and answered which word they thought they had heard from the two minimally-paired words visually presented on the screen. Those in the PROD-LAST group listened to a word and chose which word they thought they had heard from the visually presented words before repeating the aurally presented word aloud. Those in the PERC-ONLT group listened to a word and answered which word they thought they had heard from visually presented words; this group served as a control group. Participants in all groups received immediate feedback on their responses during these training sessions. The number of sessions and trials per session was the same as in Experiment 1 (i.e., four sessions of 100 trials).

4.2.3.2 PSTM task

The second PSTM task that was conducted in the post-test session in Experiment 1 was used.

4.2.3.3 Discrimination task

The same discrimination task from Experiment 1 was used.

4.2.3.4 Identification task

The same identification task from Experiment 1 was used.

4.2.4 Statistical analysis

As mentioned in the participants' section, two participants who scored higher than 90% on the pre-test in discrimination and identification tasks were excluded from data analysis entirely to avoid a ceiling effect (Cramer & Howitt, 2004). Thus, the number of participants in each group was 14 in the PROD-FIRST group, 25 in the PROD-LAST group, and 22 in the PERC-ONLY group. Regarding the PSTM task, four participants who did not complete the task properly were excluded from the analysis. The following are the numbers of participants in each group: PROD-FIRST group = 13, PROD-LAST group = 24, and PERC-ONLY group = 20. Statistical analyses were performed in the same procedure as Experiment 1.

4.3 Results

4.3.1 Discrimination task

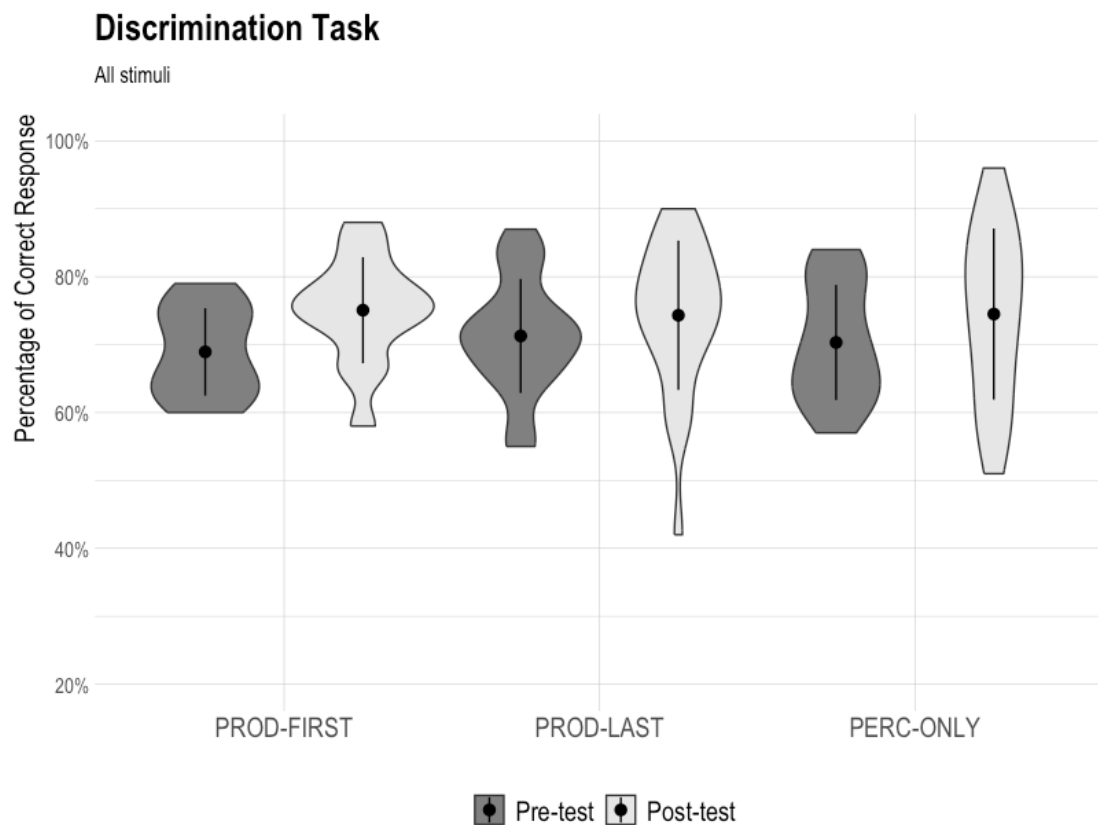
4.3.1.1 Descriptive statistics

Table 4.1 shows the mean number of correct responses, standard deviations, and percentage of correct responses in discrimination tasks per training group and familiarity type of stimuli. The maximum score of all the stimuli was 100, which combined the scores of four stimulus types. The TWTS (trained words spoken by a trained speaker) stimuli and TWUS (trained words spoken by the untrained speaker) stimuli had a maximum of 40 points each, and the UWTS (untrained words spoken by a trained speaker) stimuli and UWUS (untrained words spoken by the untrained speaker) stimuli had a maximum of 10 points each. First, the results of all stimuli combined are presented, followed by each stimulus type one by one below.

Table 4.1 Experiment 2: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-Tests of the Discrimination Task

		PROD-FIRST		PROD-LAST		PERC-ONLY	
		Mean (<i>S.D.</i>)	Percentage	Mean (<i>S.D.</i>)	Percentage	Mean (<i>S.D.</i>)	Percentage
All stimuli	Pre-test	68.93 (6.44)	68.9%	71.28 (8.40)	71.3%	70.32 (8.47)	70.3%
	Post-test	75.07 (7.81)	75.1%	74.32 (10.99)	74.3%	74.50 (12.60)	74.5%
Trained words x Trained speaker	Pre-test	25.36 (3.61)	63.4%	25.40 (3.59)	63.5%	26.36 (3.85)	65.9%
	Post-test	28.86 (2.54)	72.1%	27.44 (4.79)	68.6%	28.05 (4.27)	70.1%
Trained words x Untrained speaker	Pre-test	29.07 (3.41)	72.7%	31.16 (4.50)	77.9%	29.73 (4.13)	74.3%
	Post-test	30.29 (4.16)	75.7%	31.76 (4.98)	79.4%	30.68 (6.13)	76.7%
Untrained words x Trained speaker	Pre-test	6.43 (1.60)	64.3%	7.00 (1.55)	70.0%	6.91 (1.63)	69.1%
	Post-test	8.14 (1.51)	81.4%	7.32 (1.63)	73.2%	7.82 (1.65)	78.2%
Untrained words x Untrained speaker	Pre-test	8.07 (1.21)	80.7%	7.72 (1.37)	77.2%	7.32 (1.55)	73.2%
	Post-test	7.79 (1.25)	77.9%	7.80 (1.85)	78.0%	7.95 (1.84)	79.5%

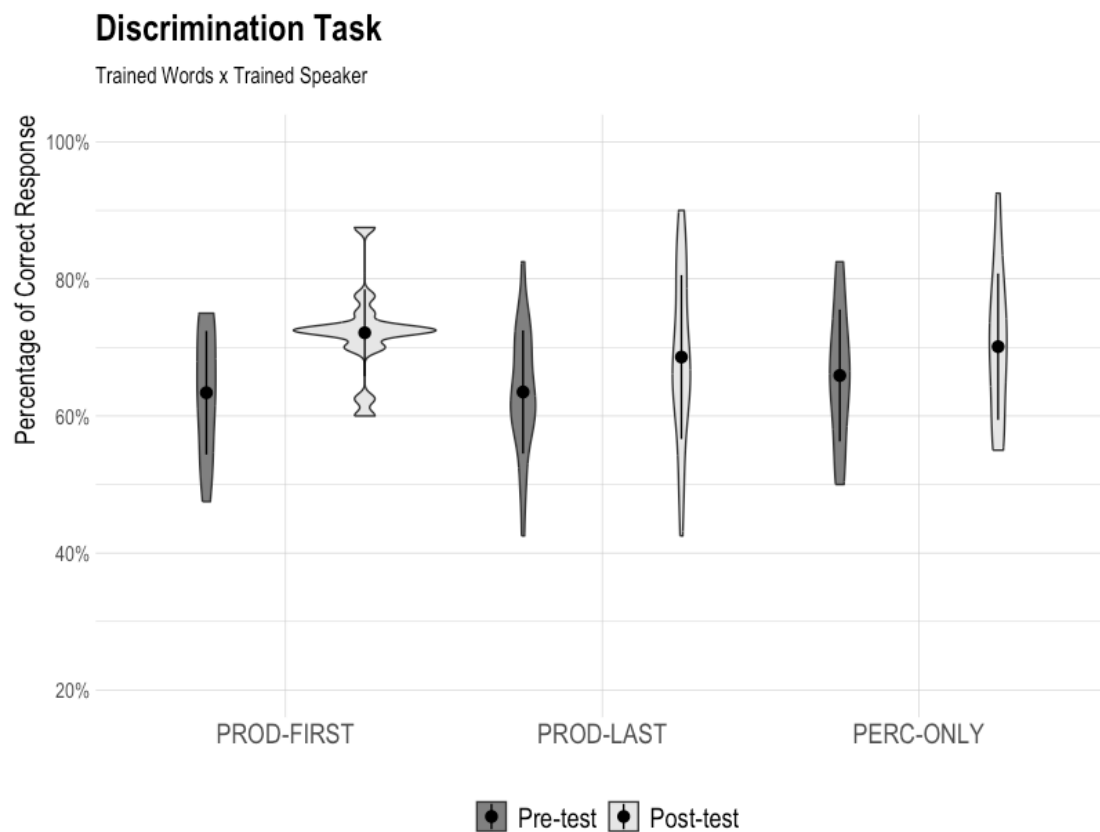
Figure 4.2 illustrates the improvement of discrimination accuracy for all stimuli. The PROD-FIRST group received the lowest score (68.9%) at pre-test, while the other two groups scored over 70% accuracy. After training, the PROD-FIRST group showed the largest improvement (6.2%) and reached 75% accuracy. The PERC-ONLY group increased their accuracy by 4.2%, followed by the PROD-LAST group's 3% increase.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 4.2 Experiment 2: Discrimination Task Results with All Stimuli

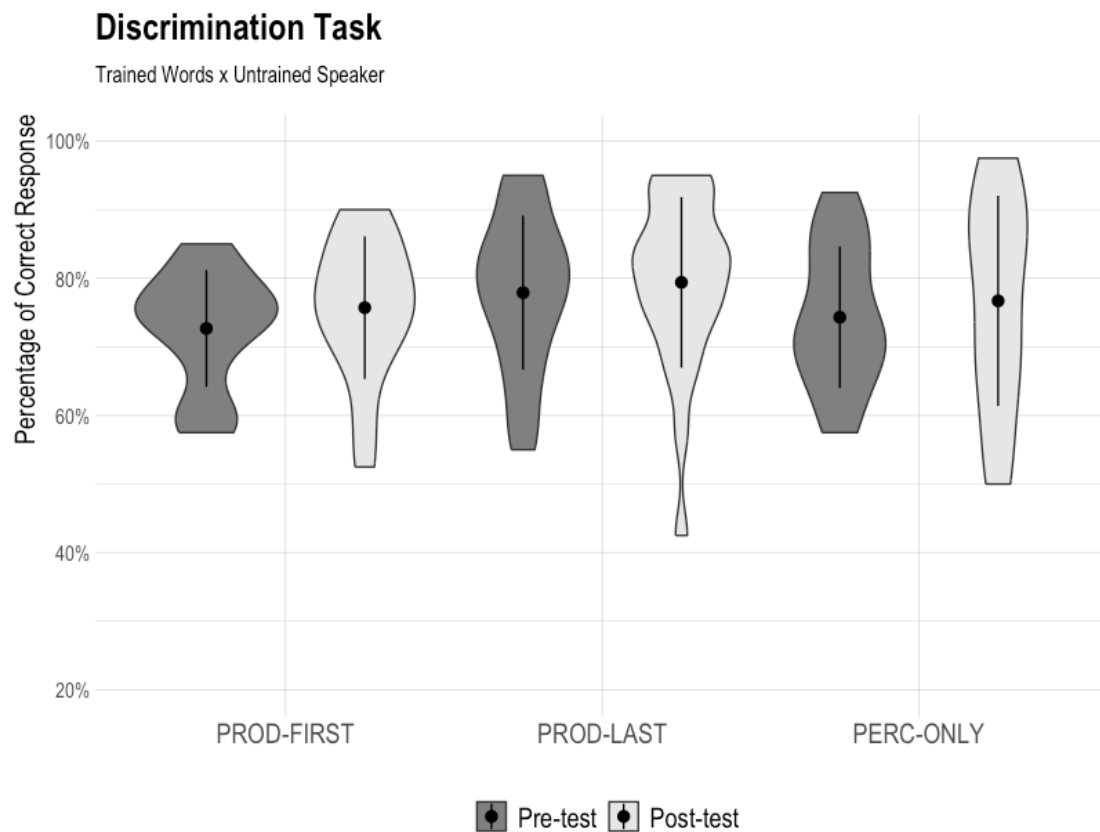
Figure 4.3 shows the improvement of the TWTS stimuli, which were used in training. All groups' pre-test scores (PROD-FIRST, PROD-LAST, and PERC-ONLY groups) were very close, respectively 25.36, 25.40, and 26.36 out of 40. After training, all groups increased their mean scores; the PROD-FIRST group increased by 8.7%, the PROD-LAST group by 5.1%, and the PERC-ONLY group by 4.2%.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 4.3 Experiment 2: Discrimination Task Results with the TWTS Stimuli

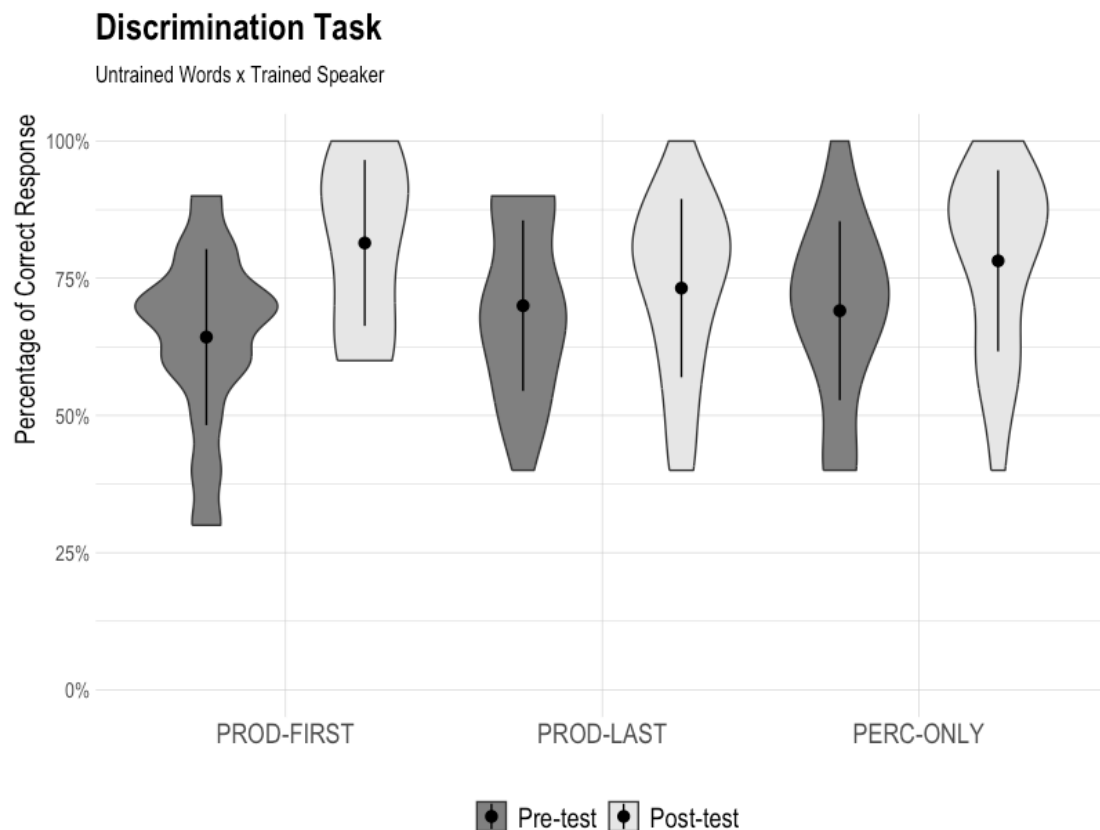
Figure 4.4 depicts the discrimination test results of the TWUS stimuli. At pre-test, the PROD-FIRST group scored 72.7%, the lowest of the three groups, followed by the PERC-ONLY group's 74.3%. The PROD-LAST group received the highest score (77.9%). After training, the PROD-FIRST group increased their discrimination accuracy by 3%, and the PERC-ONLY group by 2.4%. The PROD-LAST group improved only 1.5% from pre-test to post-test.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 4.4 Experiment 2: Discrimination Task Results with the TWUS Stimuli

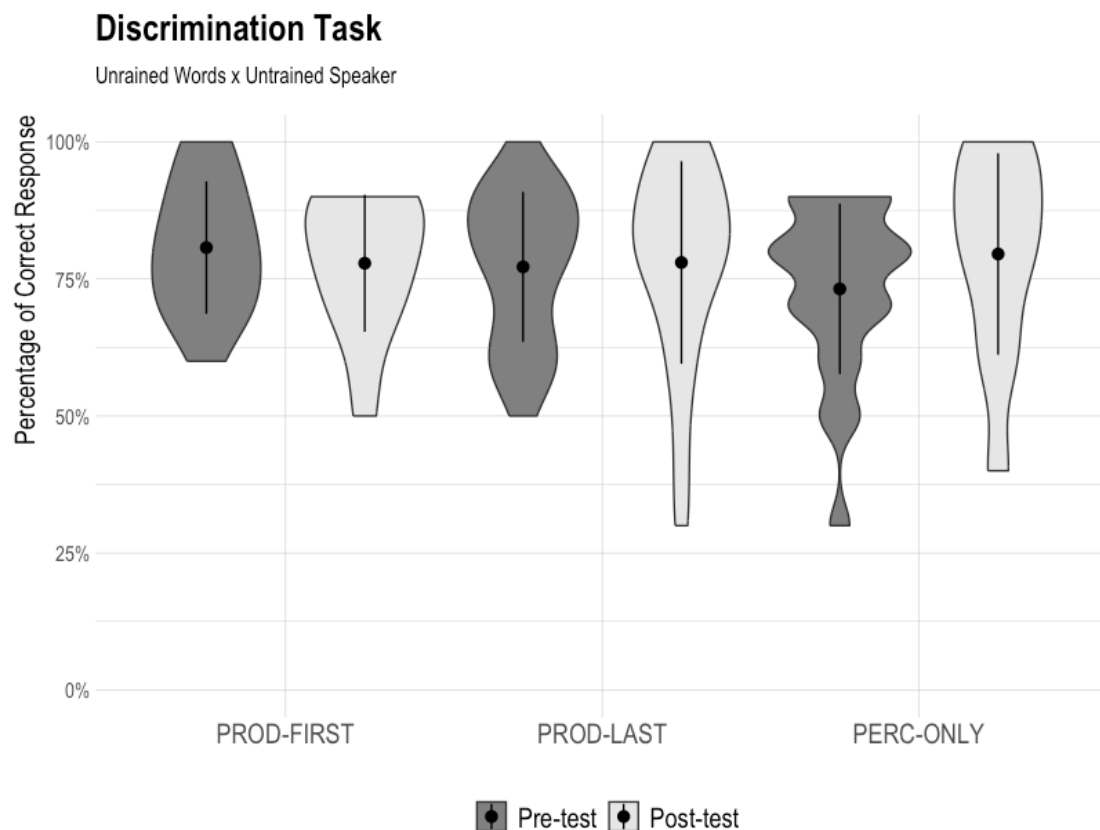
Figure 4.5 shows the results of the UWTS stimuli. The PROD-FIRST group achieved 64.3% accuracy at pre-test, the lowest of the groups. The PROF-LAST group and the PERC-ONLY group were marked with 70% and 69.1% accuracy at pre-test respectively. After training, the PROD-FIRST improved 17.1% and reached 81.4% accuracy. The PERC-ONLY group increased by 9.1% and attained 78% accuracy. The PROD-LAST group increased only by 3.2% from pre-test to post-test.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 4.5 Experiment 2: Discrimination Task Results with UWTS Stimuli

Figure 4.6 illustrates the results of the UWUS stimuli. The PERC-ONLY group received the lowest score, 73.2% accuracy, followed by the PROD-LAST group's 77.2%. The PROD-FIRST group achieved 80.7% accuracy before training. After training, the PERC-ONLY group's accuracy score improved by 6.3%, and the PROD-LAST group 1.2%. The PROD-FIRST group decreased their score by 2.8%.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 4.6 Experiment 2: Discrimination Task Results with the UWUS Stimuli

4.3.1.2 Inferential statistics

Table 4.2 shows the output of a logistic mixed effects model with the PERC-ONLY group as the reference level for all stimuli. The three-way interaction was not significant and was removed from the final model. Neither the Time $\times$ Group interaction nor the Time $\times$ Stimuli interaction was significant. However, there was a significant Group $\times$ Stimuli interaction ($p = 0.009$), indicating that participants in the PRDO-LAST group answered the TWUS stimuli significantly less accurately than the TWTS stimuli compared to participants

in the PERC-ONLY group. In addition, Stimuli (TWUS: $p < 0.001$; UWUS: $p = 0.011$) was a significant predictor, suggesting that participants in the PERC-ONLY group performed significantly better on the TWUS and UWUS stimuli than on the TWTS stimuli. Group was not a significant predictor, indicating that there were no significant differences in discrimination accuracy for the TWTS stimuli between the PROD-FIRST or PROD-LAST groups when compared to the PERC-ONLY group. Lastly, Time was a significant effect, reflecting an improvement in discrimination accuracy for the TWTS stimuli in the PERC-ONLY group from pre-test to post-test.

Table 4.2 Experiment 2: Model Output of a Logistic Mixed Effects Model for the Discrimination Task

<i>Predictors</i>	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	2.02 ***	1.60 – 2.56	< 0.001
group [PROD-FIRST]	0.97	0.70 – 1.34	0.839
group [PROD-LAST]	0.90	0.68 – 1.20	0.482
stimuli [TWUS]	1.48 ***	1.27 – 1.73	< 0.001
stimuli [UWTS]	1.35	0.96 – 1.92	0.089
stimuli [UWUS]	1.58 *	1.11 – 2.25	0.011
time [Post-test]	1.25 ***	1.15 – 1.36	< 0.001
group [PROD-FIRST] × stimuli [TWUS]	0.93	0.73 – 1.19	0.579
group [PROD-LAST] × stimuli [TWUS]	1.33 **	1.07 – 1.64	0.009
group [PROD-FIRST] × stimuli [UWTS]	0.97	0.66 – 1.42	0.858
group [PROD-LAST] × stimuli [UWTS]	0.98	0.71 – 1.37	0.927
group [PROD-FIRST] × stimuli [UWUS]	1.20	0.80 – 1.80	0.382
group [PROD-LAST] × stimuli [UWUS]	1.18	0.84 – 1.67	0.336
Random Effects			
σ^2	3.29		
τ_{00} participant	0.18		
τ_{00} word	0.13		
ICC	0.09		
N _{participant}	61		
N _{word}	50		
Observations	12200		
Marginal R ² / Conditional R ²	0.021 / 0.105		
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$			

Because Time was a significant predictor, pairwise comparisons between Time and Group were performed and revealed that the PROD-FIRST group ($\beta = -0.319$, $SE = 0.086$, $z = -3.696$, $p = 0.003$) and PERC-ONLY group ($\beta = -0.222$, $SE = 0.070$, $z = -3.199$, $p = 0.017$)

significantly improved their discrimination accuracy, while the PROD-LAST group ($\beta = -0.163$, $SE = 0.065$, $z = -2.489$, $p = 0.127$) did not.

Table 4.3 summarises only significant findings from pairwise comparisons between Group and Stimuli for the discrimination task. In the PROD-FIRST group, discrimination accuracy was significantly higher for the TWUS stimuli compared to the TWTS stimuli. A similar pattern was observed when comparing the PROD-FIRST and PROD-LAST groups for this stimulus type. In the PROD-LAST group, the difference between TWTS and TWUS was even larger. Comparisons with the PERC-ONLY group revealed that participants in the PROD-LAST group also performed significantly better on the TWUS stimuli than those in the PERC-ONLY group. In addition, there was a significant difference between the TWTS stimuli and the UWUS stimuli. In the PERC-ONLY group, discrimination accuracy was significantly higher for the TWUS stimuli compared to the TWTS stimuli, but the PERC-ONLY group's TWTS score was significantly lower than the PROD-LAST group's TWUS score.

Table 4.3 Experiment 2: Summary of Pairwise Comparisons between Group and Stimuli for the Discrimination Task

<i>TWTS</i>	<i>TWUS</i>	β	<i>SE</i>	<i>z</i>	<i>p</i>
PROD-FIRST	PROD-FIRST	-0.326	0.095	-3.421	0.031
	PROD-LAST	-0.609	0.165	-3.695	0.012
PROD-LAST	PROD-LAST	-0.675	0.074	-9.108	<.0001
	PERC-ONLY	-0.494	0.144	-3.430	0.030
PERC-ONLY	PROD-LAST	-0.575	0.145	-3.960	0.004
	PERC-ONLY	-0.394	0.078	-5.085	<.0001
<i>TWTS</i>	<i>UWUS</i>				
PROD-LAST	PROD-LAST	-0.624	0.175	-3.563	0.019

The results of pairwise comparisons between Time and Stimuli revealed a significant difference in pre-test scores between the TWTS and TWUS stimuli ($p < 0.001$), the TWTS and UWUS stimuli ($p = 0.003$), and the UWTS and UWUS stimuli ($p = 0.022$), indicating that participants answered more correctly to stimuli spoken by an untrained speaker compared to those spoken by a trained speaker before training. Additionally, there was a significant improvement from pre-test to post-test only for the TWTS stimuli ($p = 0.001$) and the UWTS stimuli ($p = 0.014$). These findings suggest that participants significantly improved their discrimination accuracy for the stimuli spoken by a trained speaker. However,

the accuracy score of the TWTS stimuli remained significantly lower than stimuli spoken by an untrained speaker.

Table 4.4 Experiment 2: Summary of Pairwise Comparisons between Time and Stimuli for the Discrimination Task

<i>Pre-test</i>	<i>Pre-test</i>	β	<i>SE</i>	<i>z</i>	<i>p</i>
TWTS	TWUS	-0.560	0.065	-8.634	<.0001
	UWTS	-0.201	0.162	-1.240	0.920
	UWUS	-0.640	0.167	-3.839	0.003
TWUS	UWTS	0.360	0.163	2.203	0.350
	UWUS	-0.080	0.168	-0.475	1.000
UWTS	UWUS	-0.439	0.133	-3.302	0.022
<i>Post-test</i>	<i>Post-test</i>				
TWTS	TWUS	-0.417	0.067	-6.204	<.0001
	UWTS	-0.389	0.168	-2.324	0.280
	UWUS	-0.489	0.169	-2.895	0.074
TWUS	UWTS	0.028	0.169	0.164	1.000
	UWUS	-0.072	0.170	-0.425	1.000
UWTS	UWUS	-0.100	0.141	-0.708	0.997
<i>Pre-test</i>	<i>Post-test</i>				
TWTS	TWTS	-0.270	0.063	-4.301	0.001
	TWUS	-0.687	0.066	-10.407	<.0001
	UWTS	-0.659	0.167	-3.949	0.002
	UWUS	-0.759	0.169	-4.507	0.000
TWUS	TWTS	0.290	0.066	4.389	0.000
	TWUS	-0.127	0.069	-1.836	0.595
	UWTS	-0.099	0.168	-0.589	0.999
	UWUS	-0.199	0.170	-1.173	0.940
UWTS	TWTS	-0.070	0.162	-0.428	1.000
	TWUS	-0.486	0.164	-2.972	0.059
	UWTS	-0.459	0.133	-3.439	0.014
	UWUS	-0.559	0.135	-4.131	0.001
UWUS	TWTS	0.370	0.167	2.212	0.344
	TWUS	-0.047	0.168	-0.279	1.000
	UWTS	-0.019	0.139	-0.140	1.000
	UWUS	-0.119	0.141	-0.848	0.990

4.3.2 Identification task

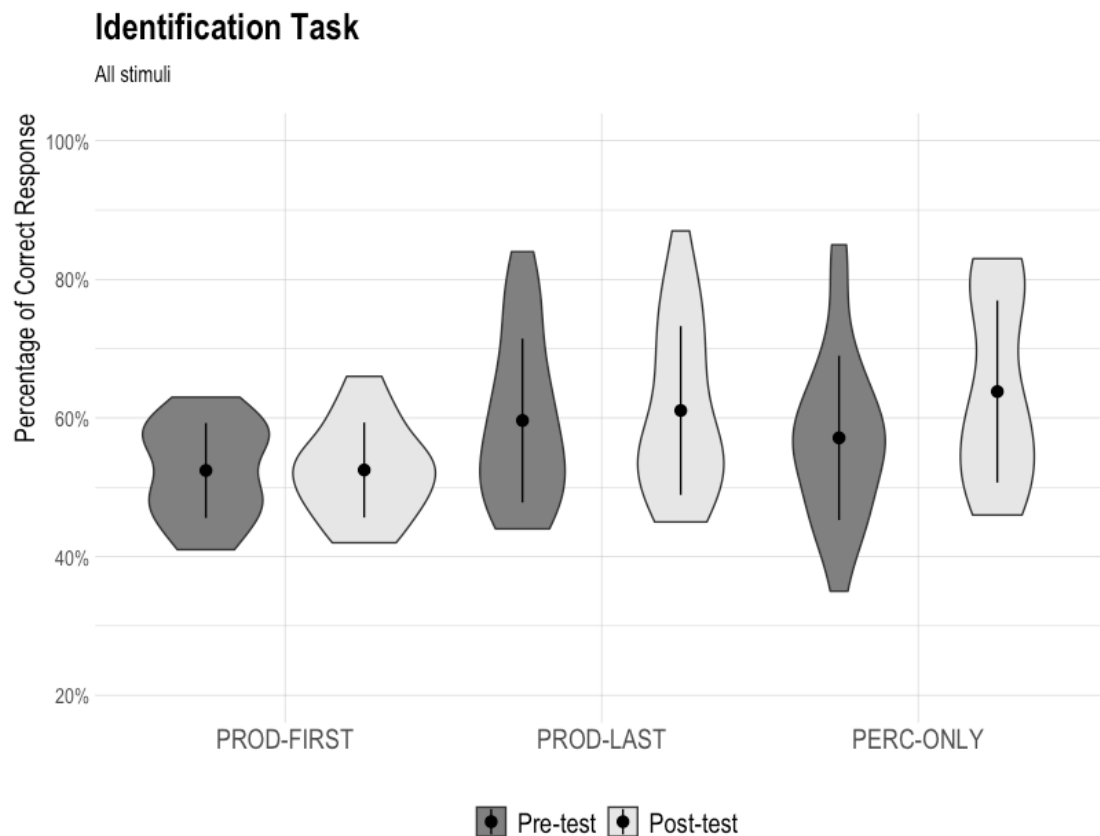
4.3.2.1 Descriptive statistics

Table 4.5 displays the mean number of correct responses, standard deviations and percentage of correct responses in identification tasks per training group and familiarity type of stimuli. The identification score was out of 100 and combined the scores of four stimulus types. The TWTS stimuli and TWUS stimuli had both a maximum of 40 points, and the UWTS stimuli and UWUS stimuli both had a maximum of 10 points. First, the results of all stimuli combined are presented, followed by each stimulus type one by one below.

Table 4.5 Experiment 2: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-Tests of the Identification Task

		PROD-FIRST		PROD-LAST		PERC-ONLY	
		Mean (<i>S.D.</i>)	Percentage	Mean (<i>S.D.</i>)	Percentage	Mean (<i>S.D.</i>)	Percentage
All stimuli	Pre-test	52.43 (6.86)	52.4%	59.64 (11.82)	59.6%	57.14 (11.86)	57.1%
	Post-test	52.50 (6.86)	52.5%	61.08 (12.18)	61.1%	64.14 (12.77)	64.1%
Trained words x Trained speaker	Pre-test	20.29 (3.81)	50.7%	23.36 (4.52)	58.4%	22.27 (5.24)	55.7%
	Post-test	20.86 (3.06)	52.1%	24.08 (4.37)	60.2%	25.18 (6.22)	63.0%
Trained words x Untrained speaker	Pre-test	21.14 (3.39)	52.9%	24.48 (5.75)	61.2%	22.91 (5.37)	57.3%
	Post-test	21.29 (4.97)	53.2%	24.44 (6.14)	61.1%	25.77 (5.79)	64.4%
Untrained words x Trained speaker	Pre-test	5.36 (1.69)	53.6%	5.64 (2.23)	56.4%	5.86 (2.05)	58.6%
	Post-test	5.07 (1.14)	50.7%	6.36 (1.58)	63.6%	6.18 (1.79)	61.8%
Untrained words x Untrained speaker	Pre-test	5.64 (1.60)	56.4%	6.16 (1.75)	61.6%	6.09 (1.87)	60.9%
	Post-test	5.29 (1.20)	52.9%	6.20 (1.80)	62.0%	7.00 (1.57)	70.0%

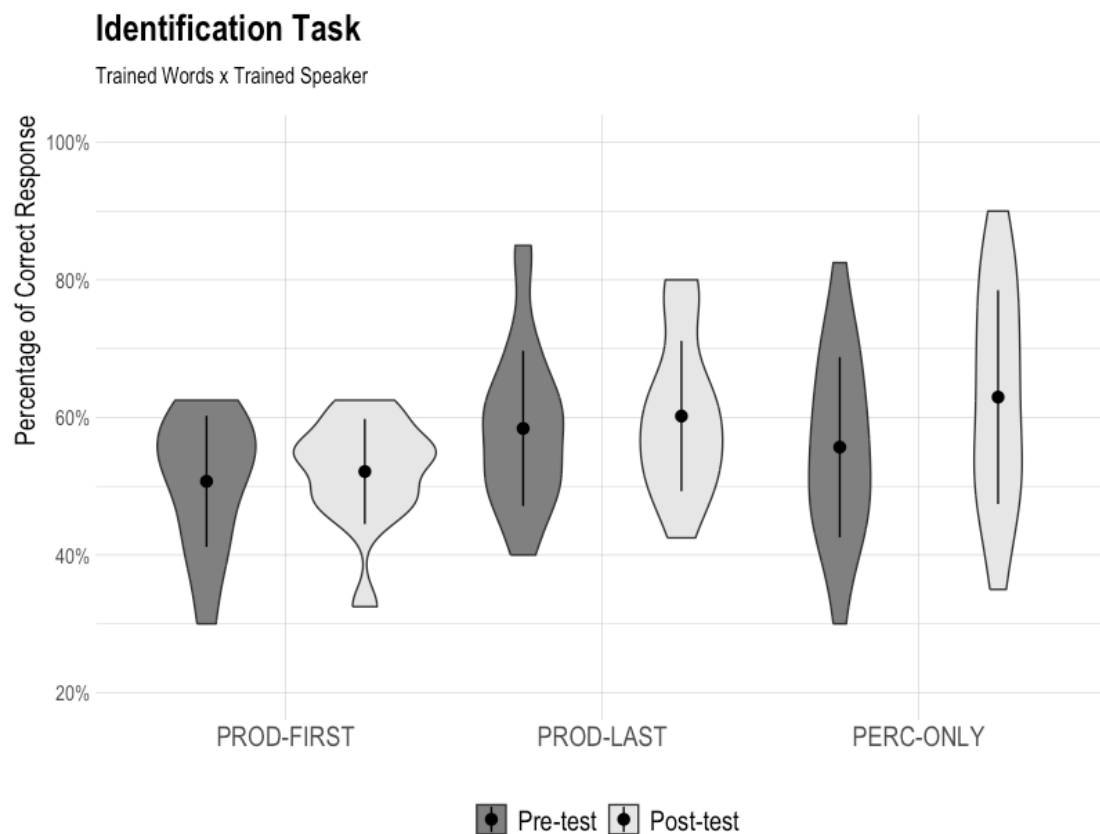
Figure 4.7 shows the identification task results with all stimuli. The PROD-FIRST group received the lowest pre-test identification score, 52.4%. The PERC-ONLY group scored 57.1% and the PROD-LAST group 59.6%. The PERC-ONLY group showed the largest improvement, a 7% increase, after training. The PROD-LAST group increased by only 1.5% from pre-test to post-test, and the PROD-FIRST group kept almost the same accuracy rate (0.1% increase).



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 4.7 Experiment 2: Identification Task Results with All Stimuli

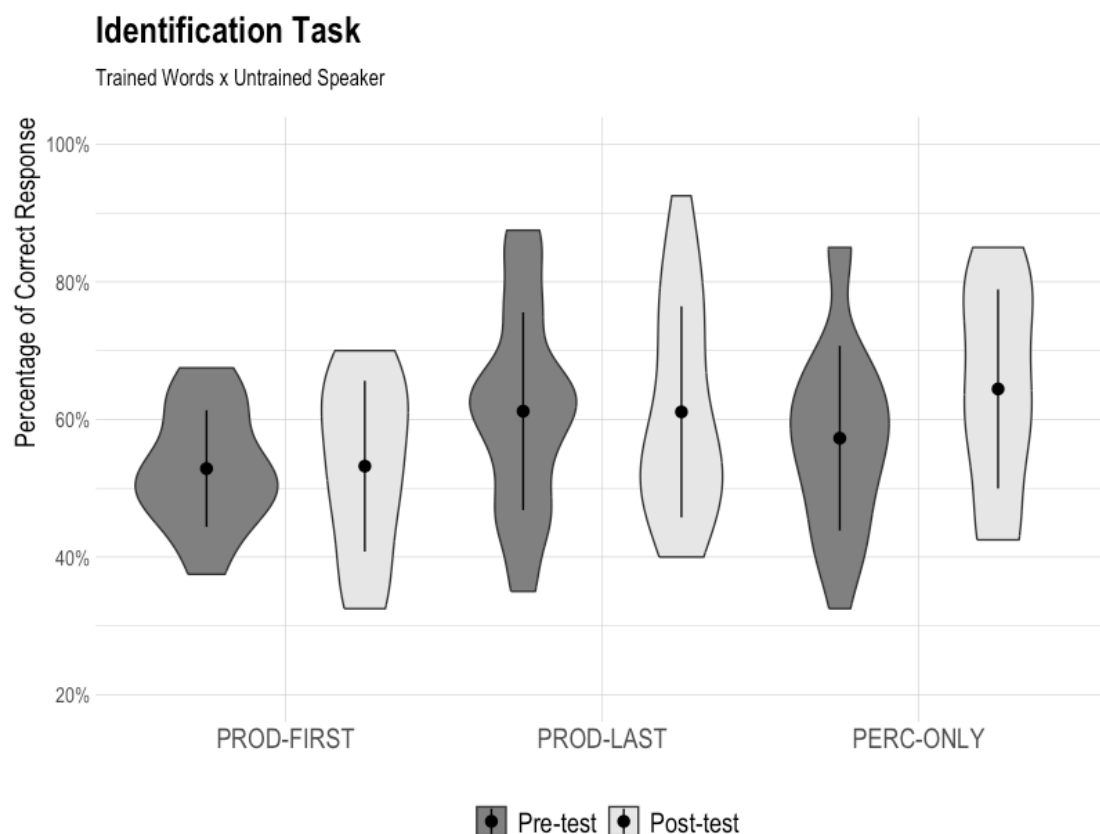
Figure 4.8 depicts the identification test results of the TWTS stimuli. The PROD-FIRST group's accuracy was 50.7%, the lowest of the three groups. The PROD-LAST group attained 58.4%, and the PERC-ONLY group reached 55.7% accuracy. After training, the PERC-ONLY group increased their accuracy score by 7.3%. The PROD-FIRST group improved by 1.4%, and the PROD-LAST group, 1.8%.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 4.8 Experiment 2: Identification Task Results with the TWTS stimuli

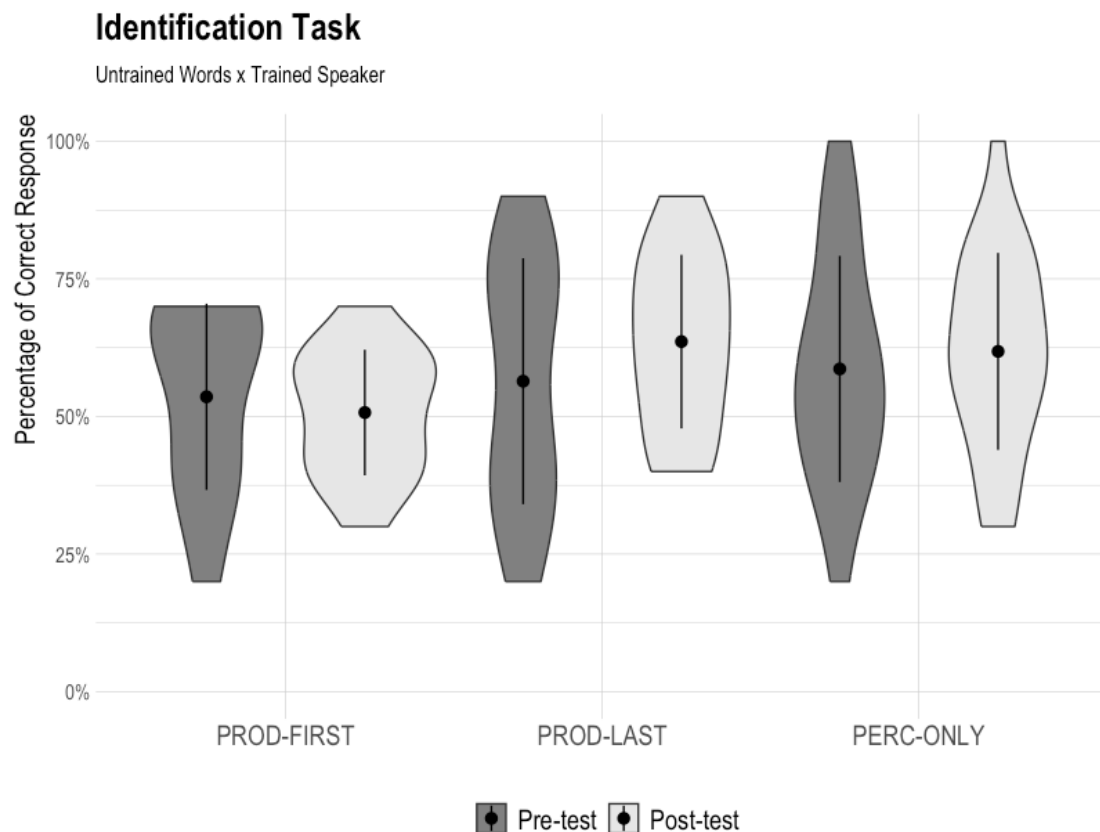
Figure 4.9 displays the results of the TWUS stimuli. Similar to the results of the TWTS stimuli, the PROD-FIRST group's pre-test score was the lowest, 52.9%, followed by the PERC-ONLY group, 57.3%. The PROD-LAST group received the highest score at pre-test (61.2%). As for the post-test results, only the PERC-ONLY group improved from pre-test to post-test, which was a 7.1% increase. The accuracy scores of the PROD-LAST group and the PROD-FIRST group remained almost the same as their pre-test scores (0.3% increase and -0.1% decrease) respectively.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 4.9 Experiment 2: Identification Task Results with the TWUS Stimuli

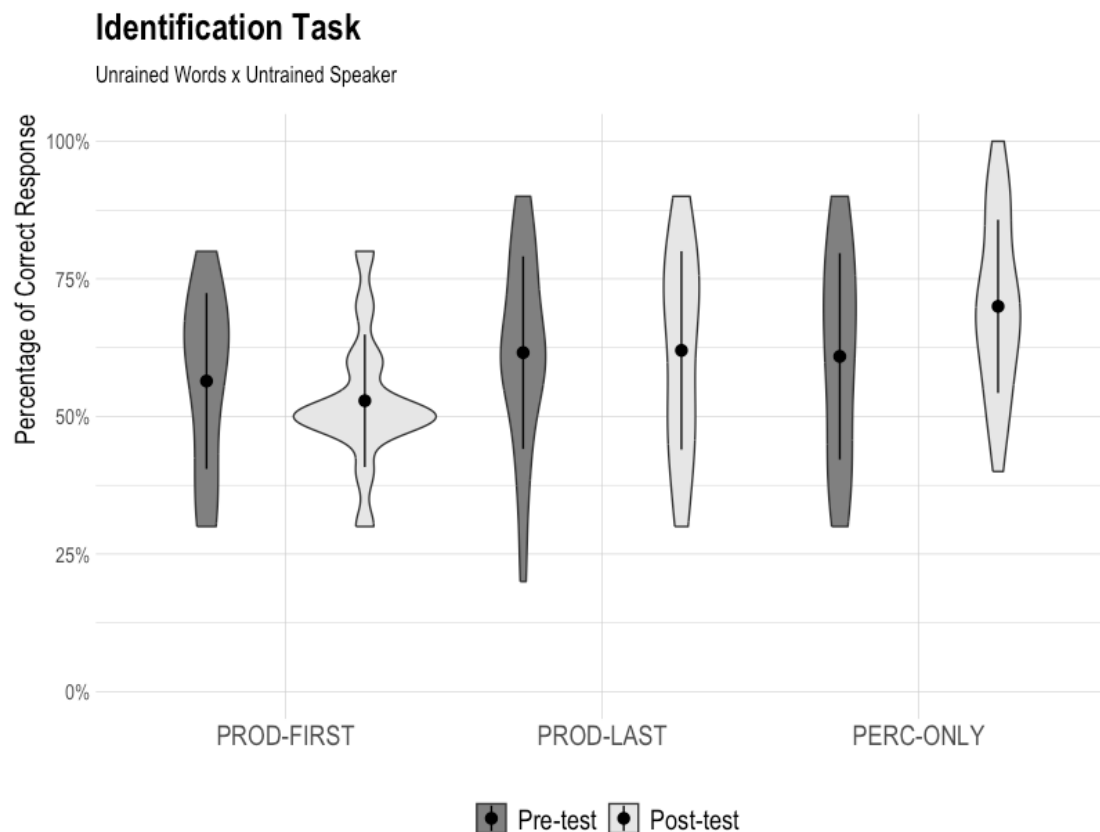
Figure 4.10 depicts the identification test results of the UWTS stimuli. At pre-test, the PERC-ONLY group scored 58.6% accuracy, the highest among the three groups. The PROD-LAST group received 56.4% accuracy, and the PROD-FIRST group 53.6%. The PROD-LAST group increased by 7.2%, followed by the PERC-ONLY group's 3.2% increase, and both groups reached over 60% accuracy. The PROD-FIRST group's accuracy score decreased 2.9%.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 4.10 Experiment 2: Identification Task Results with the UWTS Stimuli

Figure 4.11 shows the results for the UWUS stimuli. The PROD-LAST group and the PERC-ONLY group received close scores at pre-test, 61.6% and 60.9% respectively. The PROD-FIRST group scored 56.4% accuracy, which was lower than the other two groups. The PROD-FIRST group even decreased the score at post-test by 3.5%. The PROD-LAST group remained at almost the same accuracy rate (0.4% increase). Only the PERC-ONLY group improved, a 9.1% increase.



Note. The black dot of the point range shows the mean, and the length of the bar shows the range of one standard deviation.

Figure 4.11 Experiment 2: Identification Task Results with the UWUS Stimuli

4.3.2.2 Inferential statistics

Table 4.6 shows the model output of a logistic mixed-effects model with all stimuli when the reference level was the PREC-ONLY group's pre-test with the TWTS stimuli. The three-way interaction was not significant and was excluded from the final model. In addition, Stimuli as a main effect and its interactions with Group and Stimuli were not significant and were pruned from the model, indicating that stimulus type did not have a significant influence on the learning outcomes.

Table 4.6 Experiment 2: Model Output of a Logistic Mixed Effects Model for the Discrimination Task

<i>Predictors</i>	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	1.37 **	1.09 – 1.73	0.007
time [Post-test]	1.37 ***	1.21 – 1.56	<0.001
group [PROD-FIRST]	0.81	0.58 – 1.11	0.193
group [PROD-LAST]	1.12	0.85 – 1.48	0.416
time [Post-test] × group [PROD-FIRST]	0.73 **	0.60 – 0.89	0.002
time [Post-test] × group [PROD-LAST]	0.78 **	0.65 – 0.92	0.004
Random Effects			
σ^2	3.29		
τ_{00} participant	0.19		
τ_{00} word	0.17		
ICC	0.10		
$N_{\text{participant}}$	61		
N_{word}	50		
Observations	12200		
Marginal R^2 / Conditional R^2	0.009 / 0.108		
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$			

The final model revealed a significant interaction between Time and Group (PROD-FIRST: $p = 0.002$; PROD-LAST: $p = 0.004$), and Time was also a significant predictor ($p < 0.001$), suggesting that the PREC-ONLY group significantly improved their identification accuracy, and the improvement was significantly larger than that of the PROD-FIRST and PROD-LAST groups. There was no significant difference between groups at pre-test.

Table 4.7 shows the results of pairwise comparison analyses between Time and Group. The analyses revealed that only the PERC-ONLY group ($p < 0.001$) significantly increased their accuracy score after training, and the PROD-FIRST group ($p = 1.000$) and the PROD-LAST group ($p = 0.887$) did not show significant differences between their pre-test and post-test results. Additionally, the PERC-ONLY group's post-test score was significantly higher than that of the PROD-FIRST group ($p = 0.017$) and the PROD-FIRST group's pre-test score ($p = 0.016$).

Table 4.7 Experiment 2: Summary of Pairwise Comparisons between Time and Group for the Discrimination Task

<i>Pre-test</i>	<i>Pre-test</i>	β	<i>SE</i>	<i>z</i>	<i>p</i>
PROD-FIRST	PROD-LAST	-0.329	0.161	-2.047	0.316
	PERC-ONLY	-0.214	0.165	-1.301	0.785
PROD-LAST	PERC-ONLY	0.115	0.141	0.814	0.965
<i>Post-test</i>	<i>Post-test</i>				
PROD-FIRST	PROD-LAST	-0.392	0.161	-2.434	0.144
	PERC-ONLY	-0.529	0.165	-3.205	0.017
PROD-LAST	PERC-ONLY	-0.137	0.142	-0.967	0.928
<i>Pre-test</i>	<i>Post-test</i>				
PROD-FIRST	PROD-FIRST	-0.003	0.078	-0.039	1.000
	PROD-LAST	-0.395	0.161	-2.453	0.138
	PERC-ONLY	-0.532	0.165	-3.224	0.016
PROD-LAST	PROD-FIRST	0.326	0.161	2.028	0.326
	PROD-LAST	-0.066	0.060	-1.087	0.887
	PERC-ONLY	-0.203	0.142	-1.430	0.709
PERC-ONLY	PROD-FIRST	0.211	0.165	1.283	0.795
	PROD-LAST	-0.181	0.141	-1.278	0.797
	PERC-ONLY	-0.318	0.064	-4.939	<.0001

4.3.3 Summary of results on perceptual learning

The results of the discrimination task revealed a significant Time effect, indicating an improvement in the PERC-ONLY group from pre-test to post-test. However, the Group $\times$ Time interaction was not significant. The model showed a significant Group $\times$ Stimuli interaction, showing that participants in the PROD-LAST group performed significantly better on the TWUS stimuli compared to the TWTS stimuli, relative to the PERC-ONLY group. The PERC-ONLY group also performed better on the TWUS and UWUS stimuli than on the TWTS stimuli.

Pairwise comparisons revealed that the PERC-ONLY and PROD-FIRST group significantly improved, while the PROD-LAST group did not. However, since there was no significant interaction between Time and Group, the differences in improvement across groups were not statistically significant. In addition, participants in all groups had higher accuracy for the TWUS stimuli

compared to the TWTS stimuli, and those in the PROD-LAST group answered more correctly to the UWUS stimuli over the TWTS stimuli. Moreover, participants significantly improved discrimination accuracy on the TWTS and UWTS stimuli after training.

The results of the identification task showed a significant interaction between Time and Group, indicating that the PERC-ONLY group improved their identification accuracy more than the PROD-FIRST and PROD-LAST groups. Stimulus types did not influence learning outcomes. Pairwise comparisons confirmed that only the PERC-ONLY group significantly improved after training, while the PROD-FIRST and PROD-LAST groups showed no significant changes between pre-test and post-test.

4.3.4 Relationship between PSTM and perceptual learning

First, an overview of participants' performances in the PSTM task is provided. Next, the results of a linear regression model examining the relationship between PSTM and discrimination accuracy prior to training are presented. Following this, the results of the second linear regression model with the gain score from the discrimination task as the dependent variable are discussed. Finally, the results of the identification task data are provided in the same order.

4.3.4.1 PSTM task scores

Table 4.8 displays the mean scores, standard deviations, and minimum and maximum scores of the PSTM task, and Figure 4.12 depicts the PSTM score of each group. The mean score of the PROD-FIRST was 85.2, the PROD-LAST group 90.2, and the PERC-ONLY group 85.4, out of a maximum of 170 points. A linear regression model with the PSTM score as the dependent variable and Group as the independent variable revealed that there was no significant difference among the three groups.

Table 4.8 Experiment 2: Mean Scores, Standard Deviations, and Minimum and Maximum Scores of the PSTM Task

	Mean	<i>S.D.</i>	Min.	Max.
All groups	87.4	17.7	49.5	121.5
PROD-FIRST	85.2	17.1	56.2	108.7
PROD-LAST	90.2	20.0	49.5	118.2
PERC-ONLY	85.4	15.4	55.8	121.5

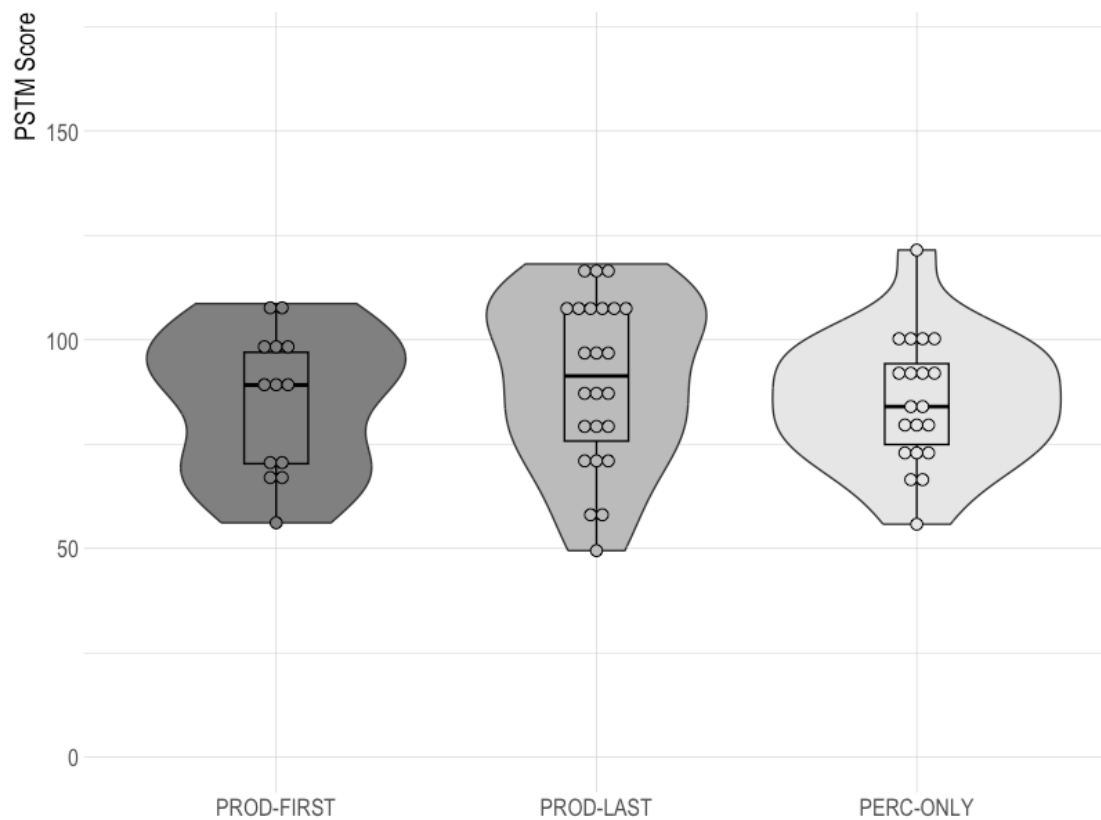


Figure 4.12 Experiment 2: PSTM Scores per Training Group

4.3.4.2 PSTM and discrimination task score

Figure 4.13 shows the correlation between the PSTM score and the discrimination task score at pre-test. A linear regression model was fit to pre-test data with the discrimination task score as the dependent variable and the PSTM score as the independent variable. The model revealed that PSTM was not a significant predictor of the discrimination accuracy before training ($\beta = 0.049$, $SE = 0.058$, $t = 0.834$, $p = 0.408$).

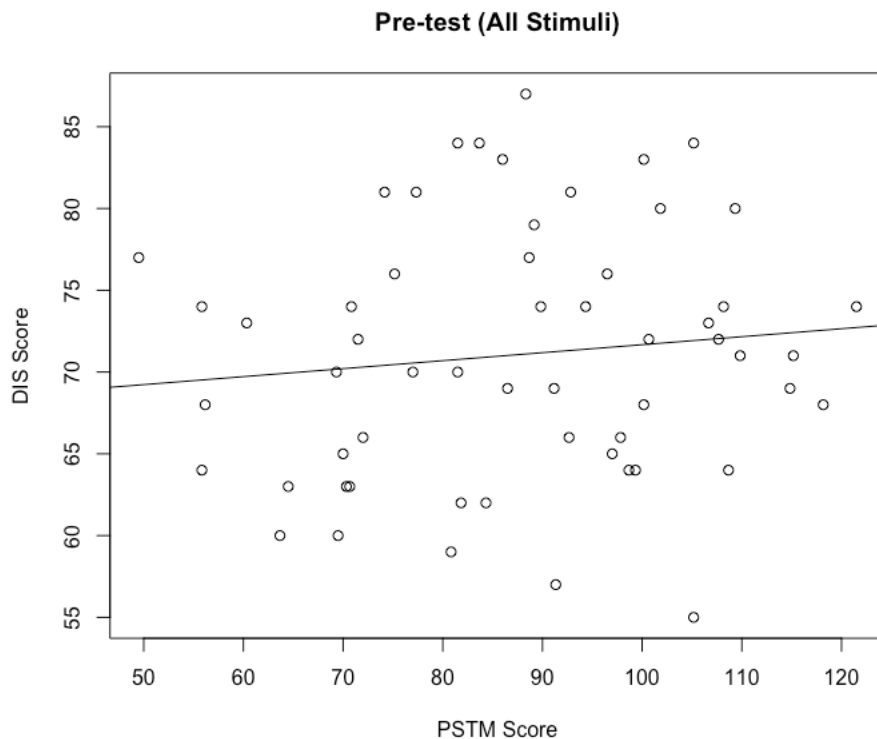


Figure 4.13 Experiment 2: Correlation between the PSTM Score and the Discrimination Task Pre-Test Score

Next, Table 4.9 presents the model output of a linear regression model with the gain score as the dependent variable and the PSTM score, Pre-test score, and Group as predictors. The model revealed a significant three-way interaction, suggesting that the relationship between PSTM and the gain from training differed by both the pre-test score and the training type. Specifically, in the PROD-LAST group, it was indicated that higher PSTM scores led to higher gain scores, and this effect became stronger as pre-test scores increased. However, no significant interaction was found in the PROD-FIRST group.

Table 4.9 Experiment 2: Model Output of a Linear Regression Model with the PSTM score for the Discrimination Task

<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	-135.77	-354.29 – 82.75	0.217
PSTM	1.95	-0.62 – 4.52	0.134
Pre-test	1.70	-1.42 – 4.82	0.278
group [PROD-FIRST]	164.70	-278.93 – 608.33	0.458
group [PROD-LAST]	428.22 **	116.93 – 739.50	0.008
PSTM × Pre-test	-0.02	-0.06 – 0.01	0.189
PSTM × group [PROD-FIRST]	-1.71	-6.92 – 3.49	0.511
PSTM × group [PROD-LAST]	-5.12 **	-8.62 – -1.62	0.005
Pre-test × group [PROD-FIRST]	-2.17	-8.74 – 4.39	0.508
Pre-test × group [PROD-LAST]	-5.79 *	-10.15 – -1.43	0.010
(PSTM × Pre-test) × group [PROD-FIRST]	0.02	-0.05 – 0.10	0.558
(PSTM × Pre-test) × group [PROD-LAST]	0.07 **	0.02 – 0.12	0.007
Observations	57		
R ² / R ² adjusted	0.307 / 0.137		

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In the PERC-ONLY group, PSTM was not a significant predictor. The two-way interaction between PSTM and Group was significant, suggesting that PSTM had a different effect on gain scores in the PROD-LAST group compared to the PERC-ONLY group. The Pre-test × Group interaction was also significant, and it was a negative relationship. These findings suggest that the pre-test score was a significant predictor of the gain score in the PROD-LAST group, and participants with higher pre-test scores showed less improvement.

Figure 4.14 visualises the interaction between PSTM, Pre-test score, and Group on the gain score. The plot depicts the relationship between predicted gain scores and PSTM scores at three different pre-test scores (65, 70, and 75) for all groups. At pre-test score of 65, a clear positive relationship between the PSTM and gain scores is observed for the PERC-ONLY group, suggesting that participants with the larger PSTM capacity benefit more from PERC-ONLY training. In contrast, the PROD-LAST group shows a negative relationship between the PSTM and gain scores,

indicating that participants with the larger PSTM capacity improved less than those with the smaller PSTM capacity. The PROD-FIRST group exhibits a relatively flat relationship, implying that PSTM has little effect on improvement.

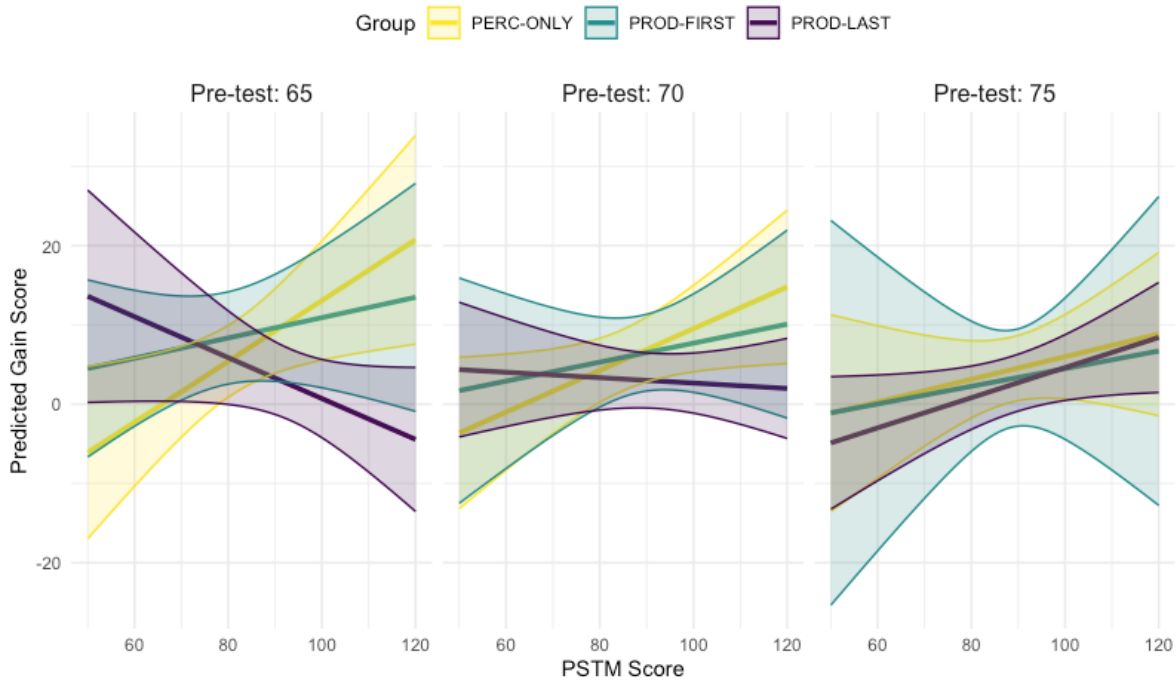


Figure 4.14 Experiment 2: Interaction between the PSTM and Pre-test Scores on the Gain Score in the Discrimination Task across Groups

At the pre-test score of 70, the general pattern remains consistent. The PERC-ONLY group continues to exhibit a positive relationship between the PSTM and gain scores, although the slope is less steep compared to the pre-test score of 65, suggesting a slightly reduced effect of PSTM at this level. The PROD-LAST group still shows a negative relationship, similar to the pattern seen at the pre-test score of 65. The PROD-FIRST group maintains a flat relationship between the PSTM and gain scores, indicating that PSTM remains a weak predictor for the gain score at this pre-test level.

At the pre-test score of 75, the PERC-ONLY group continues to show a positive relationship between the PSTM and gain scores, though the effect becomes slightly weaker compared to the lower pre-test levels. The PROD-LAST group exhibited a positive relationship between PSTM and gain scores, meaning that higher PSTM scores lead to larger improvement. This suggests that for participants in the PROD-LAST group, having the larger PSTM capacity facilitate perceptual learning from phonetic training, particularly at higher pre-test scores. In contrast, the PROD-FIRST group shows minimal change in the relationship.

4.3.4.3 PSTM and identification task score

Figure 4.15 shows the correlation between the PSTM score and the identification task score at pre-test. A linear regression model was fit to pre-test data with the identification task score as the dependent variable and the PSTM score as the independent variable. The model revealed that PSTM was not a significant predictor of the identification accuracy before training ($\beta = 0.156$, $SE = 0.084$, $t = 1.865$, $p = 0.068$), although the p value was approaching significance.

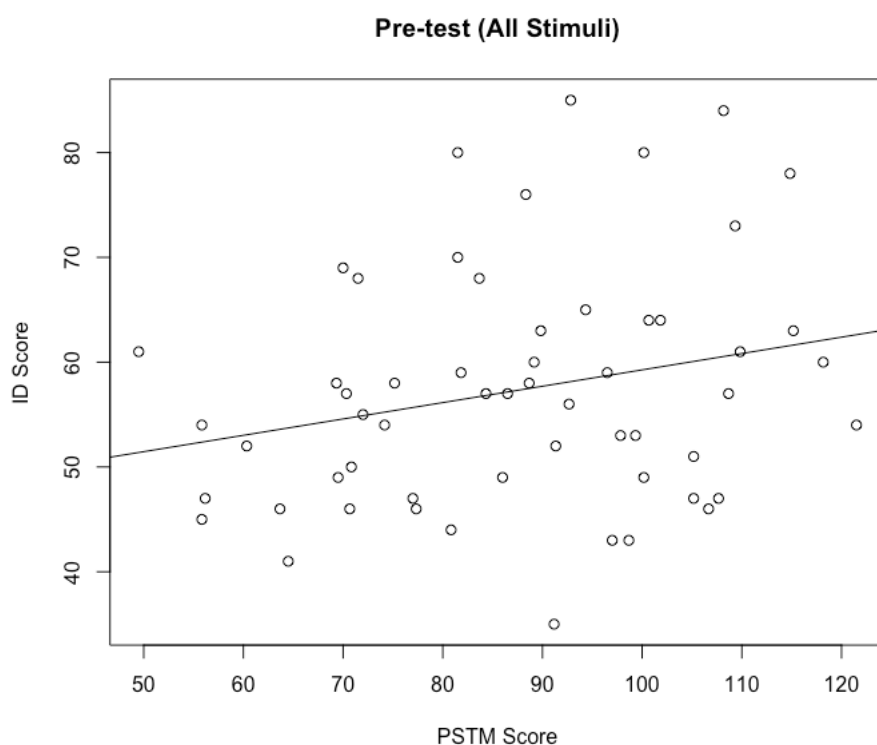


Figure 4.15 Experiment 2: Correlation between the PSTM Score and the Identification Task Pre-Test Score

Next, Table 4.10 presents the correlation between the PSTM score and the gain score of all stimuli across groups. A linear regression model was fit with the gain score as the dependent variable and the PSTM score, Pre-test score, and Group as predictors. The model revealed that a three-way interaction between PSTM, Pre-test, and Group and a two-way interaction between PSTM and Pre-test were not significant and were pruned from the final model. The model revealed an interaction between PSTM and Group, suggesting that learning outcomes of participants in the PROD-FIRST and PROD-LAST groups were less influenced by their PSTM capacity compared to those in the PERC-ONLY group. In contrast, PSTM was a significant predictor of the gain score. These findings

indicate that participants with the larger PSTM capacity showed greater improvement in identification accuracy only in the PERC-ONLY group.

Table 4.10 Experiment 2: Model Output of a Linear Regression Model with the PSTM score for the Identification Task

<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	8.95	-14.26 – 32.16	0.442
PSTM	0.45 ***	0.20 – 0.69	0.001
group [PROD-FIRST]	14.78	-28.65 – 58.20	0.497
group [PROD-LAST]	6.36	-25.09 – 37.81	0.686
Pre-test	-0.68 ***	-0.99 – -0.38	<0.001
PSTM × group [PROD-FIRST]	-0.39 *	-0.76 – -0.02	0.040
PSTM × group [PROD-LAST]	-0.52 ***	-0.82 – -0.22	0.001
group [PROD-FIRST] × Pre-test	0.13	-0.60 – 0.87	0.713
group [PROD-LAST] × Pre-test	0.56 **	0.14 – 0.98	0.010
Observations	57		
R ² / R ² adjusted	0.469 / 0.381		

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

A significant interaction between the pre-test score and Group was found only for the PROD-LAST group, indicating that participants in the PROD-LAST group with higher pre-test scores showed greater improvement compared to the PERC-ONLY group. In addition, Pre-test score was a significant predictor with a negative estimate, suggesting that participants who achieved higher pre-test scores showed lower improvement in identification accuracy, except in the PROD-LAST group.

Figure 4.16 illustrates the interaction between the PSTM score and Group. The plot demonstrates that the PERC-ONLY group exhibited a strong positive relationship between the PSTM and gain scores, suggesting that higher PSTM scores resulted in greater improvement. In contrast, the PROD-LAST group and PROD-FIRST groups showed flatter or slightly negative relationship.

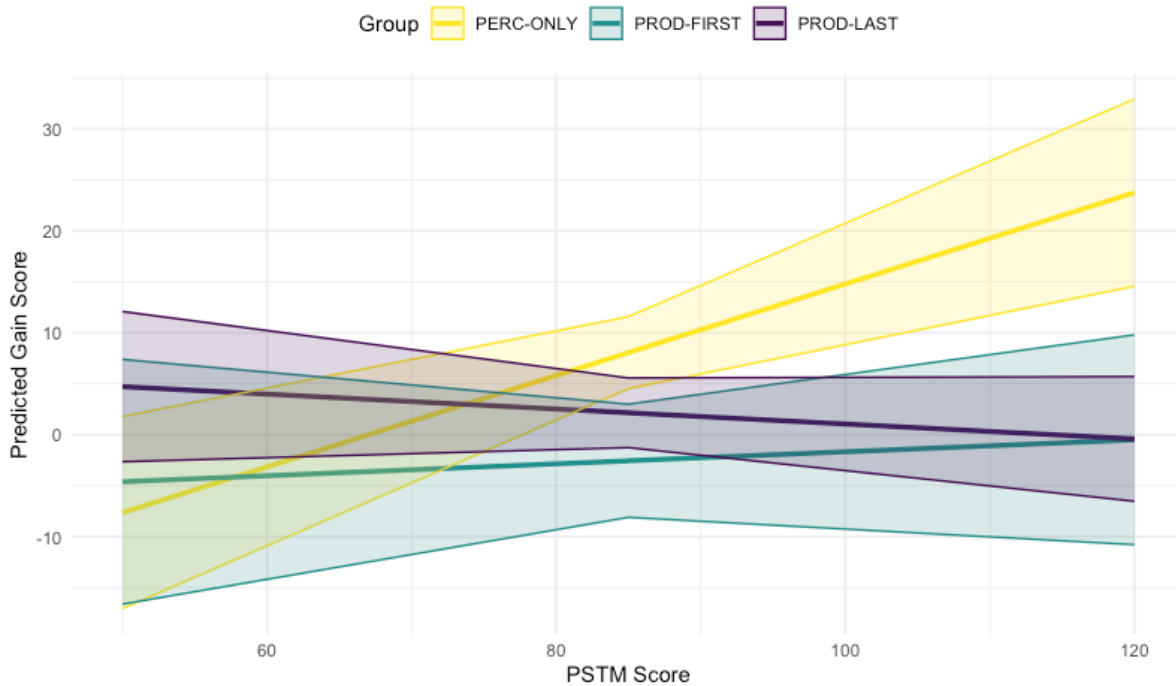


Figure 4.16 Experiment 2: Interaction between the PSTM and Gain Scores in the Identification Task across Groups

4.3.5 Summary of results on PSTM

The results from the discrimination task suggest that the relationship between the PSTM and gain scores is highly dependent on the training type and the pre-test score. In the PERC-ONLY group, PSTM consistently predicts higher gain scores across all pre-test levels, suggesting that participants with stronger PSTM benefit more from perceptual training. However, PSTM was not a significant predictor of the gain score. In contrast, in the PROD-LAST group, a positive relationship was observed only at higher pre-test scores. This indicates that participants with the larger PSTM capacity improved more from PROD-LAST training, only when starting with higher pre-test scores. In the PROD-FIRST group, PSTM shows a little effect on the gain score, indicating that PSTM may not play a significant role in improving discrimination accuracy for PROD-FIRST group. Regarding learning in identification accuracy, the results indicate that while PSTM is a strong positive predictor of the gain score only for the PERC-ONLY group but not for the PROD-FIRST and PROD-LAST groups.

4.4 Discussion

This section begins with a summary of the results, followed by a discussion of the findings presented in the previous section and a comparison with those of previous studies on L2 speech learning. The discussion addresses each research question in turn.

The results revealed that in the discrimination task, the PERC-ONLY and PROD-FIRST groups significantly improved from pre-test to post-test, but the PROD-LAST group did not. However, the differences in improvements were not statistically different across groups. The Group $\times$ Stimuli interaction revealed that participants in all groups performed significantly better on the TWUS stimuli compared to the TWTS stimuli.

In the identification task, a significant Time $\times$ Group interaction indicated that the PERC-ONLY group improved more than the other groups, with only this group showing significant gains from pre-test to post-test. Stimulus type did not significantly influence the identification learning.

Regarding PSTM, the results showed that the relationship between the PSTM and gain scores is influenced by both training type and pre-test scores. In the PERC-ONLY group, higher PSTM consistently predicted larger improvement in discrimination accuracy, suggesting that PSTM capacity plays an important role in the success of PERC-ONLY training. However, in the PROD-LAST group, this relationship was only positive at higher pre-test scores, indicating that participants with higher PSTM benefit more from PROD-LAST training when they have higher initial perceptual accuracy. In contrast, PSTM had little effect on gains in the PROD-FIRST group. As for identification accuracy, while PSTM was a significant predictor of the gain score in the PERC-ONLY group, this effect was not observed in the PROD-FIRST and PROD-LAST groups.

(1) Does perceptual training using identification tasks improve L2 phonetic perception?

The results of the identification task showed that PERC-ONLY training in an identification task format improved L2 learners' accuracy in identification of the English /l/-/r/ contrast. This finding aligns with previous studies on perceptual training (e.g., Bradlow et al., 1997; Shinohara & Iverson, 2018; Thomson, 2012b) and supported that the identification training is effective on improving identification accuracy of L2 sounds. The results of the discrimination task also showed that PERC-ONLY training in an identification task format significantly improved L2 learners' accuracy in discriminating the English /l/ and /r/ sounds. It means that the participants improved both discrimination and identification skills through identification training. Together with the findings from Experiment 1, Experiments 1 and 2 posit that identification training can improve both

identification and discrimination accuracy, while discrimination training can improve only discrimination accuracy but not identification accuracy. These findings are in line with previous studies (Carlet & Cebrian, 2019; Grenon et al., 2018; Jamieson & Morosan, 1986; Law et al., 2019; Logan et al., 1991; Strange, 1992).

In the discussion of Experiment 1, three possible factors that were raised that might explain why discrimination training did not improve identification accuracy: no opportunity to learn phoneme-grapheme associations, task familiarity, and the lack of robust improvement despite its significance. Since identification accuracy improved discrimination accuracy in Experiment 2, it is less likely that task familiarity played a role. Moreover, the improvement rate of identification accuracy is around 7%, which was lower than that of discrimination accuracy in Experiment 1. This indicates that identification training facilitates learning in discrimination accuracy, independent of the improvement rate.

Therefore, it is most likely that participants in Experiment 1 did not learn phoneme-grapheme associations during discrimination training, and that led to lack of improvement in identification accuracy, as suggested by Grenon et al. (2018). However, although training in these two experiments involved the same number of trials, the amount of exposure to the target sounds differed (participants in Experiment 1 received three times more exposure). In addition, the improvement rate was relatively small compared to previous studies (e.g., Bradlow et al., 1997; Shinohara & Iverson, 2018). Thus, it is important to conduct more research by taking into account the amount of exposure to target sounds.

(2) Does production during perceptual training disrupt L2 phonetic perceptual learning?

The results of identification tasks showed that neither the PROD-FIRST group nor the PROD-LAST group significantly improved their identification accuracy after training. Since PERC-ONLY training showed a significant improvement, this finding suggests that production during perceptual training disrupted perceptual learning, which supports Baese-Berk (2010, 2019) and Baese-Berk and Samuel (2016). However, it contradicts other studies showing no disruption to perceptual learning after perception-and-production mixed training (Lu et al., 2015; Thorin et al., 2018).

However, both PROD-FIRST group improved as much as the PERC-ONLY group in the discrimination task, indicating that production during identification training did not negatively impacted discrimination learning. Combining the results from both the identification and

discrimination tasks, similar results to those of Experiment 1 emerged, which is that production practice during perceptual training did not enhance perceptual learning; neither production group outperformed the PERC-ONLY group in both Experiments 1 and 2.

A potential reason for the null-effect of production practice in perceptual training is a lack of feedback on participants' production, as mentioned in the discussion in Experiment 1 (see Chapter 3.4). Another possibility could be that simply production does not enhance perceptual learning, like other studies have suggested (Baese-Berk, 2010, 2019; Baese-Berk & Samuel, 2016; Lu et al., 2015; Thorin et al., 2018). As for transfer from production learning to perceptual learning, there are mixed findings. Some studies found that production training leads to significant improvement in perception (Herd et al., 2013; Hirata, 2004; Kartushina et al., 2015; Kissling, 2014; Sakai, 2016; Thorin, 2020), while other studies showed no improvement in perception after production training (Hattori, 2009; Liakin et al., 2013, 2015; Tateishi & Winters, 2015).

Considering L2 speech models' hypotheses, Best's PAM predicts that production practice could enhance perceptual learning due to shared phonetic representations. In contrast, SLM (Flege, 1995a) and the NLM model (Kuhl & Iverson, 1995) posit that perceptual learning precedes production learning. Therefore, production practice may not enhance perceptual learning. The results from this experiment support SLM and the NLM over PAM. However, the disruption effect of production was also observed, indicating a negative relationship between perception and production. Therefore, more research needs to be conducted to untangle this relationship between the perception and production. About the relationship between the two modalities is discussed more in Chapter 6.2.

This experiment showed that production practice during identification training disrupted identification learning, but the disruption effect was not observed in discrimination learning. The cause of the inconsistency between discrimination learning and identification learning remains unclear, and more research with more extended training is needed to further explore the disruption effects on perceptual learning. Potential causes of disruption to perceptual learning caused by production during perceptual training are discussed in Chapter 6.1.

(3a) Is perceptual accuracy related to learners' PSTM?

The PSTM score was not a significant predictor of the pre-test scores of either the discrimination task or the identification task. This finding further supported the results of Experiment 1, which found that PSTM was not a significant predictor of perceptual accuracy before training. As mentioned in the Discussion of Experiment 1, these findings are similar to those of Safronova and

Mora (2012) and Ghaffarvand Mokari and Werner (2019) and different from other studies (Cerviño-Povedano et al., 2011; Inceoglu, 2019; Kondo, 2019; MacKay et al., 2001). The contradiction might be explained by the inconsistency between mora-based items in the PSTM task and syllable-based words in training (see Chapter 3.4).

(3b) Is the degree of perceptual learning by phonetic training related to learners' PSTM?

The results revealed that PSTM was a significant predictor of the gain score of the identification accuracy in the PERC-ONLY group but not in the PROD-FIRST and PROD-LAST groups. This suggested a positive association between PSTM and perceptual accuracy, while it was limited to the PERC-ONLY group. Given that PSTM was a significant predictor of the perceptual accuracy of both the discrimination task and identification task, it supports previous studies showing PSTM can predict L2 perceptual ability (Cerviño-Povedano et al., 2011; Inceoglu, 2019; Kondo, 2019; MacKay et al., 2001) and learning (Aliaga-Garcia et al., 2011).

The possible reason for the inconsistency in the results among groups might be the different degrees of perceptual learning after training. The PERC-ONLY group is the only group that significantly improved in the identification task, where the PSTM score predicted the gain in the group. Since the other two groups did not significantly increase their scores, the relationship between PSTM and perceptual accuracy remained dissociated after training. In contrast, the participants in the PERC-ONLY group improved their perceptual accuracy, whose degree aligned with their PSTM capacity.

Another major finding is regarding discrimination accuracy, the PSTM effect varied over time only in the PROD-LAST group. A positive PSTM effect on the gain score of the discrimination task was observed in the PROD-LAST group when participants possessed higher discrimination accuracy prior to training, while a negative relationship was observed with lower pre-test scores. Participants with lower pre-test scores likely had more room to improve, and their learning may have occurred regardless of their PSTM capacity because initial gains are often easier when starting from a lower baseline. However, as participants reached a higher level of accuracy, further improvement became more challenging, which could explain why PSTM started to play a more significant role. In this case, those with stronger PSTM were better equipped to process sound, leading to better learning. Thus, it is possible that the role of PSTM becomes more evident when participants are already at a relatively advanced stage of perceptual learning.

One potential explanation for the negative relationship could be the learners' overreliance on L1 phonetic categories. Studies have shown that the phonological loop stores phonologically familiar items more than unfamiliar ones (Gathercole & Baddeley, 1990; Storkel, 2001). This suggests that the larger PSTM capacity allows faster and more automatic retrieval of L1 phonetic representations, which may cause learners to rely on L1 categories even when processing L2 sounds. While this automaticity may contribute to a high PSTM test score in L1-based tasks, it can reduce an ability to carefully process auditory input. In contrast, learners with weaker PSTM may rely less on memory retrieval and more on perceptual processing, helping them notice subtle differences in target sounds during learning, leading to better learning outcomes.

These findings suggest the importance of considering the initial perceptual ability when investigating the relationship between PSTM and the degree of perceptual learning. To further investigate the potential change in how the PSTM incapacity influences L2 perceptual learning, more research needs to be conducted.

4.5 Summary of Experiment 2

Experiment 2 investigated the effects of production on L2 perceptual learning through identification training. The PERC-ONLY group improved both discrimination and identification accuracy. In contrast, the PROD-FIRST group improved only in discrimination accuracy, while the PROD-LAST group did not. Despite these mixed results, the findings suggest that producing target phonemes during perceptual training either disrupted perceptual learning or did not contribute to additional learning, potentially due to the lack of feedback on participants' production.

Experiment 2 also explored the relationship between PSTM and perceptual learning and found that PSTM capacity predicted gains in identification accuracy only in the PERC-ONLY group. In the PROD-FIRST group, PSTM was not a significant predictor of perceptual learning. However, in the PROD-LAST group, the relationship between PSTM and perceptual learning appeared to shift over the course of learning.

Chapter 5 Experiment 3: The effects of perception and production training on phonetic learning and its relationship with PSTM

5.1 Introduction

Experiments 1 and 2 investigated how producing target sounds during discrimination training and identification training, respectively, affected perceptual learning. These experiments built upon the methods of Baese-Berk and Samuel (2016) and Thorin et al. (2018), adapting them for a group of Japanese learners of English to address differences in the timings of oral repetition and cross-linguistic issues. The results uncovered that discrimination-based PERC-ONLY training (without oral repetition) and PROD-FIRST training (repeating a stimulus aloud immediately right after hearing it) significantly improved discrimination accuracy, though with minimal impact on identification accuracy. This discrepancy might stem from a mislabelling issue, as noted by Grenon et al. (2019), who found that many learners mislabelled target sounds after discrimination training. In contrast, identification-based phonetic training indicated that while both PERC-ONLY training and PROD-FIRST training significantly improved discrimination accuracy, only PERC-ONLY training significantly enhanced identification accuracy.

Interestingly, PROD-LAST training (repeating a stimulus aloud after answering a perceptual question) did not improve discrimination and identification accuracy as effectively as the other two training methods in both discrimination and identification training conditions. This contradicts Thorin et al.'s (2018) findings, which suggested that oral repetition does not disrupt perceptual learning. To further explore whether producing stimuli disrupts perceptual learning, Experiment 3 employed a PROD-LAST training format, where participants repeated the sounds after completing a perceptual task, instead of a PROD-FIRST format, where they repeated the sounds immediately after hearing them. Additionally, Experiment 3 increased the number of trials to address the limited perceptual learning observed in Experiments 1 and 2.

In addition to examining the effects of the two training types, Experiment 3 introduced production-only training using an ASR system to explore the production effect on perceptual learning. Some studies have shown that production-only training enhances perceptual accuracy (e.g., Herd et al., 2013; Kartushina et al., 2015; Sakai, 2016; Thorin, 2020; Wong, 2014), while others report no transfer effect from production to perception (Hattori, 2009; Liakin et al., 2013, 2015). Moreover, this experiment investigated whether prior exposure to target sounds could alleviate the disruptive effects of production, as Baese-Berk and Samuel (2016) suggested. To this end, a training condition was introduced that began with perception-only training and concluded with production-

only training. Therefore, Experiment 3 examined the interactive effects of perceptual and production training on both modalities.

Following Experiments 1 and 2, Experiment 3 also explored the relationship between PSTM and L2 perceptual learning. As Experiments 1 and 2 yielded mixed findings regarding whether PSTM predicts perceptual learning, further research is necessary to clarify this relationship. In addition, although several studies have examined the connection between PSTM and production learning (Moorman, 2017; Mora & Darcy, 2016; Nagle, 2013; Slevc & Miyake, 2006; Venkatagiri & Levis, 2007; Zahler & Lord, 2022), to the best of my knowledge, no research has examined this relationship before and after phonetic training. To address this gap in the literature, Experiment 3 explored whether PSTM predicts both perceptual and production learning.

In summary, this experiment further investigated the effects of different types of phonetic training—focusing on either perception, production, or the combination of the two—on both perceptual and production learning. Additionally, it examined whether learners' PSTM capacity predicts their phonetic learning outcomes. The research questions are as follows:

1. Does perceptual training in an identification task format improve L2 phonetic perception and production?
2. Does production training using the ASR system improve L2 phonetic perception and production?
3. Does production during perceptual training in an identification task format disrupt L2 phonetic perceptual and production learning?
4. Does prior exposure to target phonemes alleviate the disruption of production to L2 perceptual and production learning?
5. Is phonetic learning related to learners' PSTM?
6. How do learners perceive each phonetic training?

5.2 Method

5.2.1 Participants

5.2.1.1 Japanese learners of English

Data collection for this study took place at a university in Japan. In total, 89 Japanese learners of English participated in this phonetic training study. Of these, 66 participants who completed all tasks during pre- and post-tests, as well as all training sessions correctly, were initially included in the analysis. However, six participants were excluded because they scored over 90 out of 100 on both discrimination and identification tasks at pre-test. Consequently, the final analysis for

perception and production learning was conducted with 60 participants. With regard to the PSTM task, one participant who did not complete the task properly was excluded from the analyses involving the PSTM data.

Participants were undergraduate students aged 18 to 19, majoring in School Nursing, Art and Calligraphy Education, Mathematics Education, Home Economics Education, and Music Education. They were all enrolled in the same English course covering the same topics and using the same learning materials facilitated by the same instructor. Their English proficiency ranged from A2 to B1 in the CEFR, corresponding to a beginner to lower-intermediate level. None of them reported having lived in English-speaking countries for more than three months. Those who completed all the training sessions and pre- and post-training tasks received extra credits for the English course.

5.2.1.2 L1 English speakers for an L1 rating task

In addition to the Japanese participants, three L1 English speakers (mean age = 35.3, two male) were recruited for L1 rating tasks for the recordings of L1 Japanese speakers' speeches. Two of them were L1 Australian English speakers, and one was an L1 New Zealand English speaker with extended exposure to Australian English and American English. The raters' linguistic backgrounds align with the training and testing stimuli, which were recorded by L1 Australian speakers. All raters had experience interacting with English spoken by L2 speakers including L1 Japanese speakers. All of them self-reported normal hearing.

5.2.2 Materials

5.2.2.1 English stimuli for training sessions and tests

The same stimuli as Experiments 1 and 2 were used. However, the numbers of trained stimuli and untrained stimuli was adjusted to include more untrained items to test learning generalisation. The new stimulus types are summarised in Table 5.1. In Experiment 3, trained stimuli comprised 15 minimal pairs, while untrained stimuli comprised 10 minimal pairs. The number of speakers remained consistent with Experiments 1 and 2. However, for testing, one of the female trained speakers replaced the male trained speaker who appeared in pre-test and pos-test in Experiments 1 and 2 because participants answered significantly less accurately with the stimuli spoken by the male-trained speaker. Therefore, both speakers who appeared in the perceptual tests in Experiment 3 were female.

Table 5.1 Experiment 3: Stimulus Types of Different Word and Speaker Familiarity

	Word familiarity	Speaker familiarity	Minimal pairs
TWTS	Trained	Trained	15
TWUS	Trained	Untrained	15
UWTS	Untrained	Trained	10
UWUS	Untrained	Untrained	10

5.2.2.2 Stimuli for phonological short-term memory tasks

The same set of stimuli was used in the post-test session in Experiments 1 and 2.

5.2.2.3 Language background questionnaire

The same questionnaire from Experiments 1 and 2 was used.

5.2.3 Procedure

Data collection took place online from June to July 2022. The pre- and post-training tasks and the perception and production training tasks were developed using PCIBex farm, a web-based experimental presentation platform. A different platform was used after Experiment 2 due to the closure of Ibex farm in September 2021. Communication with participants was facilitated through online tools such as email, Zoom, and Google Forms. All communication between L1 Japanese participants and the researcher, including task instructions, was conducted in their first language, Japanese.

The timeline of the data collection is summarised in Figure 5.1. Participants were divided into four groups based on their English class: perception-only (PERC-ONLY) group, perception-and-production-mixed group (MIX), production-only group (PROD-ONLY), and perception-to-production group (PERC-PROD). First, all L1 Japanese participants completed pre-training tasks, which included a language background questionnaire, a picture discrimination task, a sentence reading task, a word-list reading task, a discrimination task, and an identification task. Then, they underwent four training sessions over the course of a month, with one session occurring per week. One week after completing all training sessions, participants completed post-training tasks, consisting of a picture discrimination task, a sentence reading task, a discrimination task, an identification task and an exit questionnaire. Additionally, they also performed a PSTM task with the researcher individually on Zoom during the data collection period.

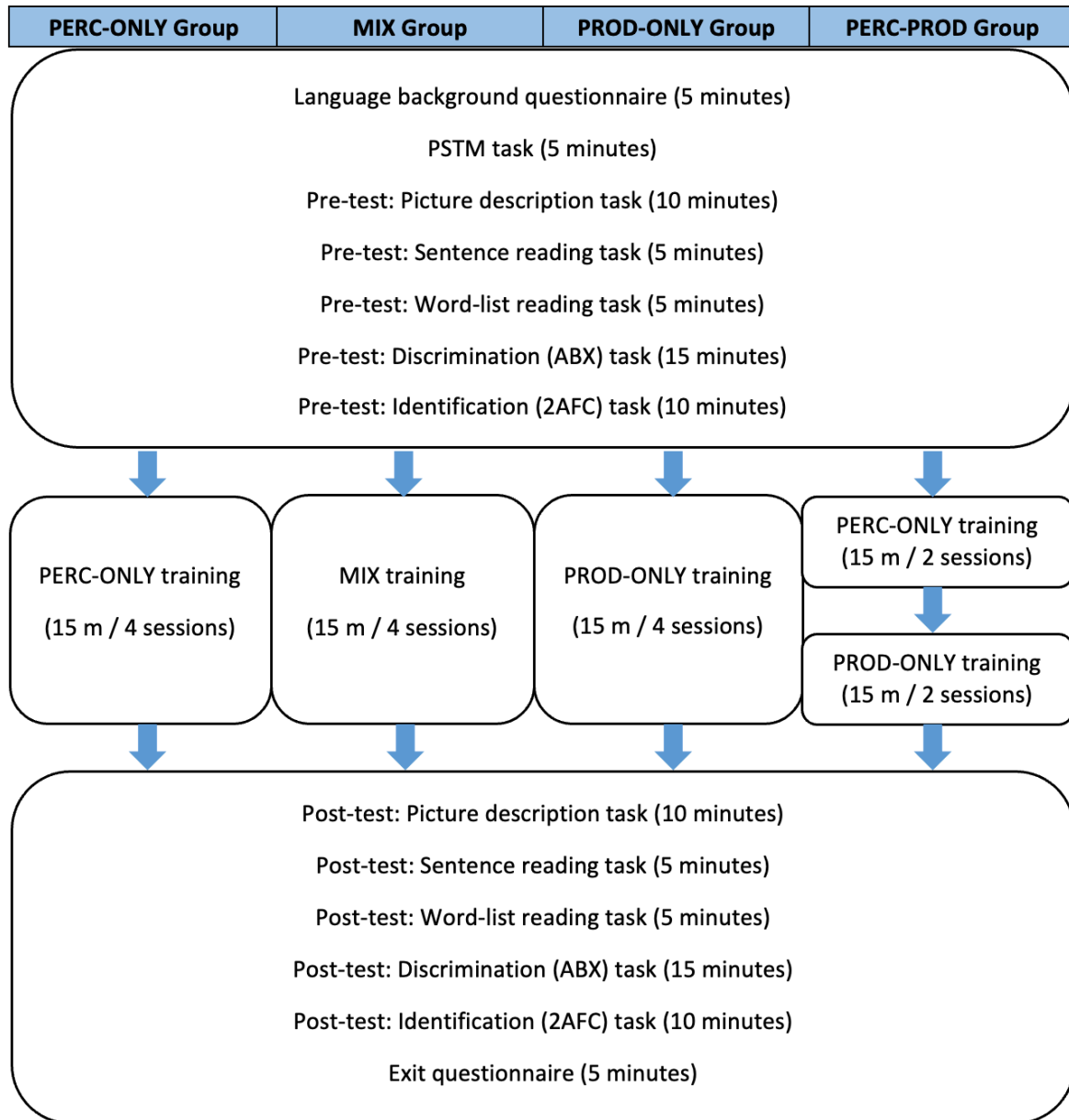


Figure 5.1 Experiment 3: Timeline of Data Collection

5.2.3.1 Phonetic training

The training was designed in an identification task format. The PERC-ONLY group underwent the same training as the PERC-ONLY group from Experiment 2 (listening to a word and identifying which word it was from orthographically presented words). The MIX group received similar training as the PROD-LAST training from Experiment 2 (listening to a word and choosing which word they thought they heard from the orthographically presented words and repeating the word aloud). In the PROD-ONLY group, participants practised pronouncing English /l/ and /r/ sounds using the ASR system on Google Docs. Participants were instructed to check if ASR

correctly recognised their intended words. In the PERC-PROD group, participants first received two sessions of PERC-ONLY training followed by two sessions of PROD-ONLY training. During training, all participants received immediate feedback or ASR's immediate response on their perceptual decisions or productions.

The number of trials in each training session increased from 100 in Experiments 1 and 2 to 150 with the aim of achieving better learning outcomes. Each training session included 30 words spoken by five speakers, with each token encountered four times throughout the training period. The order of presenting the stimuli was randomised, and each session started with practice trials using words that did not contain /l/ and /r/ sounds.

5.2.3.2 PSTM task

The same PSTM test as in Experiment 2 was used.

5.2.3.3 Picture description task

The picture description task consisted of ten pictures. Two to three keywords or phrases were given to each picture to elicit the English /l/ and /r/ sounds, following Saito and Munro (2014). Japanese translations of some keywords were provided and participants were encouraged to use all the keywords during the task. Appendix D lists the expected descriptions for each picture along with the keywords. Participants had thirty seconds to describe each picture, and the presentation order was randomised. Before the main task, participants practised with three pictures containing keywords that did not involve the English /l/ and /r/ sounds. This task aimed to elicit target sounds in spontaneous speech rather than in controlled speech.

5.2.3.4 Sentence reading task

In the sentence reading task, participants read aloud a list of 26 sentences, comprising five untrained minimal pairs and eight trained minimal pairs (see Appendix E). The sentence order was randomised. In addition to these sentences, participants read ten practice sentences aloud before the main task. As a semi-controlled production task, the sentence reading task was designed to encourage more attention to pronunciation than the more spontaneous picture description task (Derwing & Munro, 2015).

5.2.3.5 Word-list reading task

In the word-list reading task, participants read aloud the same list of words used in the discrimination and identification tasks, with the word order randomised. As a controlled production task, it allowed participants to focus closely on pronunciation.

5.2.3.6 Discrimination task

The same discrimination task as in Experiments 1 and 2 was used.

5.2.3.7 Identification task

The same identification task as in Experiments 1 and 2 was used.

5.2.3.8 Exit questionnaire

A questionnaire asked participants for their perspectives on the training they underwent (see Appendix F). Two questions asked about their desire to improve their English skills and pronunciation on a 10-point Likert scale, with 10 being the highest (‘そう思う’—I believe so). The next questions focused on the phonetic training participants experienced, including the effectiveness of the training and their willingness to continue the training, rated on a scale of 1 to 10, along with the reasons for their ratings. Lastly, participants provided additional thoughts and comments on the training.

5.2.4 Rating tasks

Three L1 English speakers conducted two rating tasks to evaluate the Japanese participants’ English /l/–/r/ sounds. The first task was conducted in a 2AFC identification task format, following the same procedure as the identification test completed by Japanese participants. One point was awarded if two or more raters answered correctly; otherwise, no point was given.

The second task was a goodness rating task, where an intended word was presented orthographically and a Japanese participant’s production was presented auditorily. Raters evaluated the quality of English /l/ or /r/ on a Likert scale from 1 (poor) to 9 (excellent). The interrater reliability, calculated by Cronbach’s alpha, for the 2AFC identification task was $\alpha = 0.769$, indicating acceptable consistency, and for the goodness rating task, $\alpha = 0.847$, indicating good consistency (Plonsky & Derrick, 2016).

The stimuli for these tasks included three minimal pairs of trained words (lock-rock, lent-rent, late-rate) and three pairs of untrained words (low-row, light-right, lace-race). In total, there were 4,072 tokens for stimuli (176 tokens were removed due to low-quality recordings), which resulted in six sessions of six blocks (approximately 220 tokens per block). The stimuli order was randomised within each block, and the block and session orders were counterbalanced among the raters. The tasks were self-paced, with breaks between blocks, and each block took about 10 to 15 minutes. All sessions were completed over four weeks. The tasks were conducted using PCibex in a quiet room. In both tasks, raters were instructed to focus solely on the English /l/–/r/ sounds and ignore other sounds or mistakes, such as different vowel productions and vowel insertions (e.g., light‘o’).

5.2.5 Statistical analysis

As mentioned in the participants' section, six participants who scored higher than 90% on the pre-test in both discrimination and identification tasks were excluded from data analysis to avoid a ceiling effect (Cramer & Howitt, 2004). Thus, the number of participants in each group was 15 in the PERC-ONLY group, 16 in the MIX group, 15 in the PROD-ONLY group, and 14 in the PERC-PROD group. Furthermore, one participant in the PERC-ONLY group was removed from data analysis for production due to recording issues. With regard to the PSTM task, one participant in the MIX group who did not complete the task properly was excluded from the analyses involving the PSTM data.

Statistical analyses for perceptual data were performed using the same procedure as in Experiments 1 and 2. For production data, identification rating scores and goodness rating scores were analysed separately. An alpha level was set to 0.05 for all statistical tests, and models were pruned by excluding the most non-significant predictor in a step-wise fashion.

First, a logistic mixed-effects model was fitted to the 2AFC identification rating score as the dependent variable. The fixed effects were Group (PERC-ONLY, MIX, PROD-ONLY, and PERC-PROD), Time (pre-test and post-test), Word type (Trained words and Untrained words), and Task type (Word-list reading, Sentence reading, and Picture description), and random effects were Participant and Word. Next, for goodness rating scores, a linear mixed-effects model was fitted to the goodness rating scores as the dependent variable. Group, Time, Word type, and Task type were included as fixed effects, and Participant and Word were random effects. As post-hoc analyses, Tukey's pairwise comparison tests were performed using the *emmeans* function (Lenth, 2023) for further analyses.

The relationship between perception data and the PSTM score was analysed in the same procedure as in Experiments 1 and 2. Production data were analysed in a similar manner—a linear regression model was fitted to the identification rating scores and goodness rating scores separately. First, a linear regression model was fitted to the pre-test scores as the dependent variable and the PSTM scores as the independent variable. Next, to explore the relationship between PSTM and production improvement, a linear regression model was fitted to the identification rating scores and goodness rating scores separately with the gain score as the dependent variable and the PSTM score, Group, and the Pre-test score as independent variables.

5.3 Results

5.3.1 Discrimination test

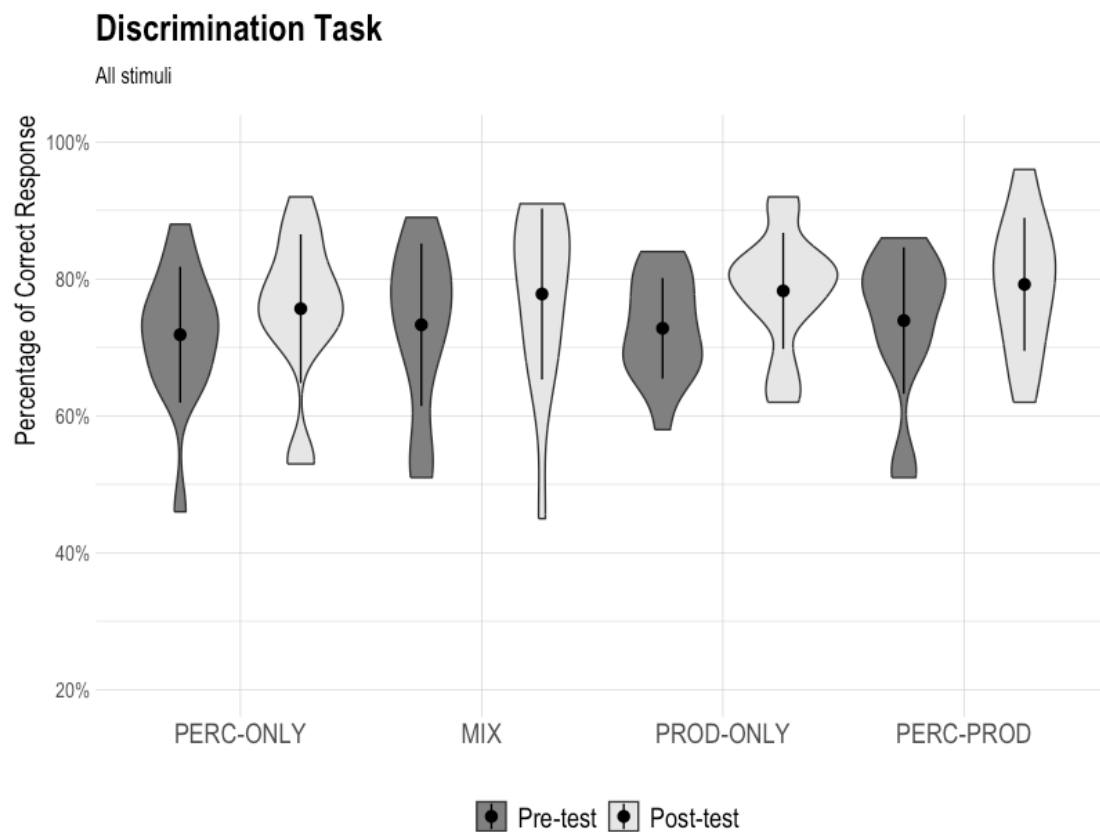
5.3.1.1 Descriptive Statistics

Table 5.2 shows the mean number of correct responses, standard deviations, and the percentage of correct responses in discrimination tasks per training group and familiarity type of stimuli. The maximum score for all stimuli was 100, which combined the scores of four stimulus types. The TWTS stimuli and TWUS stimuli had a maximum of 40 points each, and the UWTS stimuli and UWUS stimuli had a maximum of 10 points each.

Table 5.2 Experiment 3: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-tests of the Discrimination Task

		PERC-ONLY		MIX		PROD-ONLY		PERC-PROD	
		Mean (S.D.)	Percentage	Mean (S.D.)	Percentage	Mean (S.D.)	Percentage	Mean (S.D.)	Percentage
All stimuli	Pre-test	71.87 (9.93)	71.9%	73.31 (11.86)	73.3%	72.80 (7.35)	72.8%	73.93 (10.69)	73.9%
	Post-test	75.67 (10.87)	75.7%	77.81 (12.49)	77.8%	78.27 (8.47)	78.3%	79.21 (9.72)	79.2%
Trained words x Trained speaker	Pre-test	22.67 (3.85)	75.6%	21.88 (4.56)	72.9%	23.60 (2.53)	78.7%	23.57 (4.52)	78.6%
	Post-test	24.87 (4.44)	82.9%	25.38 (4.75)	84.6%	24.20 (3.69)	80.7%	25.57 (3.30)	85.2%
Trained words x Untrained speaker	Pre-test	21.93 (3.69)	73.1%	22.38 (3.44)	74.6%	21.40 (2.92)	71.3%	21.57 (3.08)	71.9%
	Post-test	22.27 (3.41)	74.2%	22.00 (4.21)	73.3%	23.73 (2.22)	79.1%	23.64 (2.95)	78.8%
Untrained words x Trained speaker	Pre-test	14.13 (2.03)	70.7%	15.00 (3.16)	75.0%	14.00 (2.27)	70.0%	15.64 (2.31)	78.2%
	Post-test	14.73 (3.13)	73.7%	15.94 (2.35)	79.7%	16.13 (2.67)	80.7%	15.71 (2.95)	78.6%
Untrained words x Untrained speaker	Pre-test	13.13 (2.36)	65.7%	14.06 (2.26)	70.3%	13.80 (1.52)	69.0%	13.14 (2.21)	65.7%
	Post-test	13.80 (2.31)	69.0%	14.50 (3.41)	72.5%	14.20 (2.34)	71.0%	14.29 (2.37)	71.4%

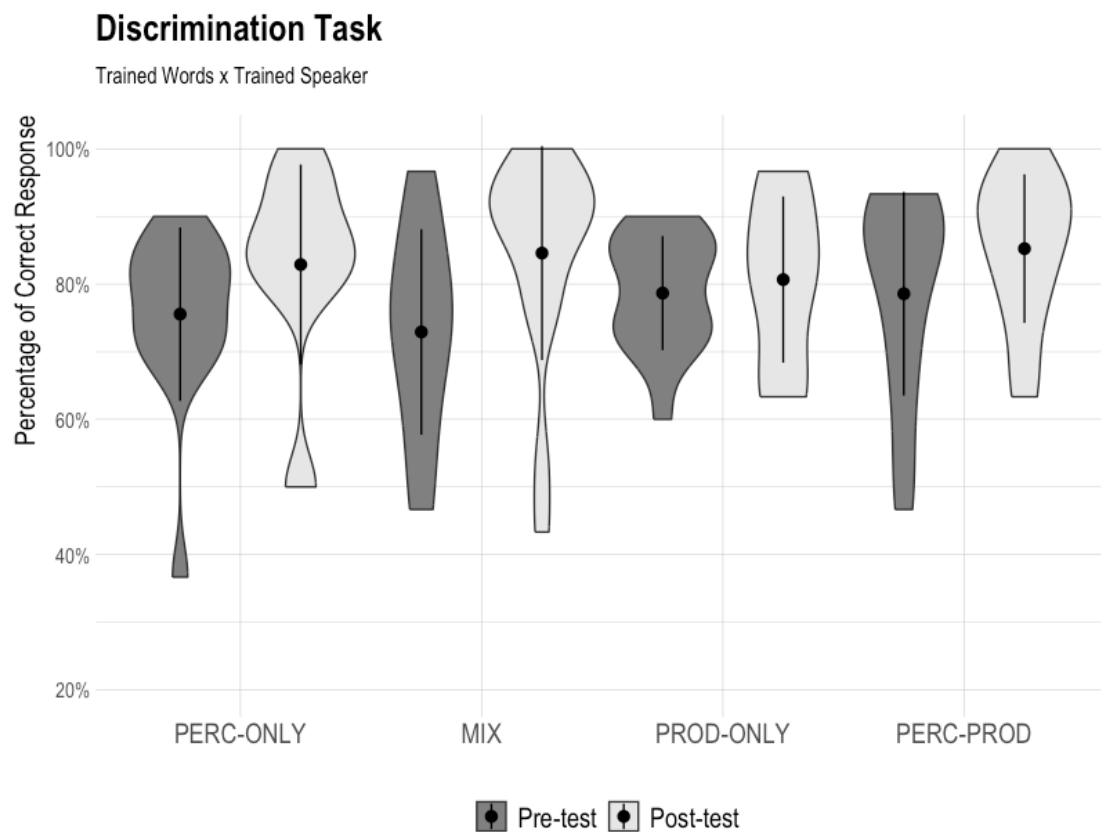
Figure 5.2 illustrates the improvement in discrimination accuracy for all stimuli. At pre-test, the PERC-PROD group achieved the highest score (73.9%), followed by the MIX group (73.3%). The PROD-ONLY group scored 72.8%, and the PERC-ONLY group received the lowest accuracy score (71.9%). After training, the PROD-ONLY group showed the largest improvement (5.5%) and reached 78.3% accuracy, followed by the PERC-PROD group with a 5.3% improvement. The MIX group improved by 4.5%. The PERC-ONLY group displayed the smallest improvement, with an increase of 3.8%.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.2 Experiment 3: Discrimination Task Results with All Stimuli

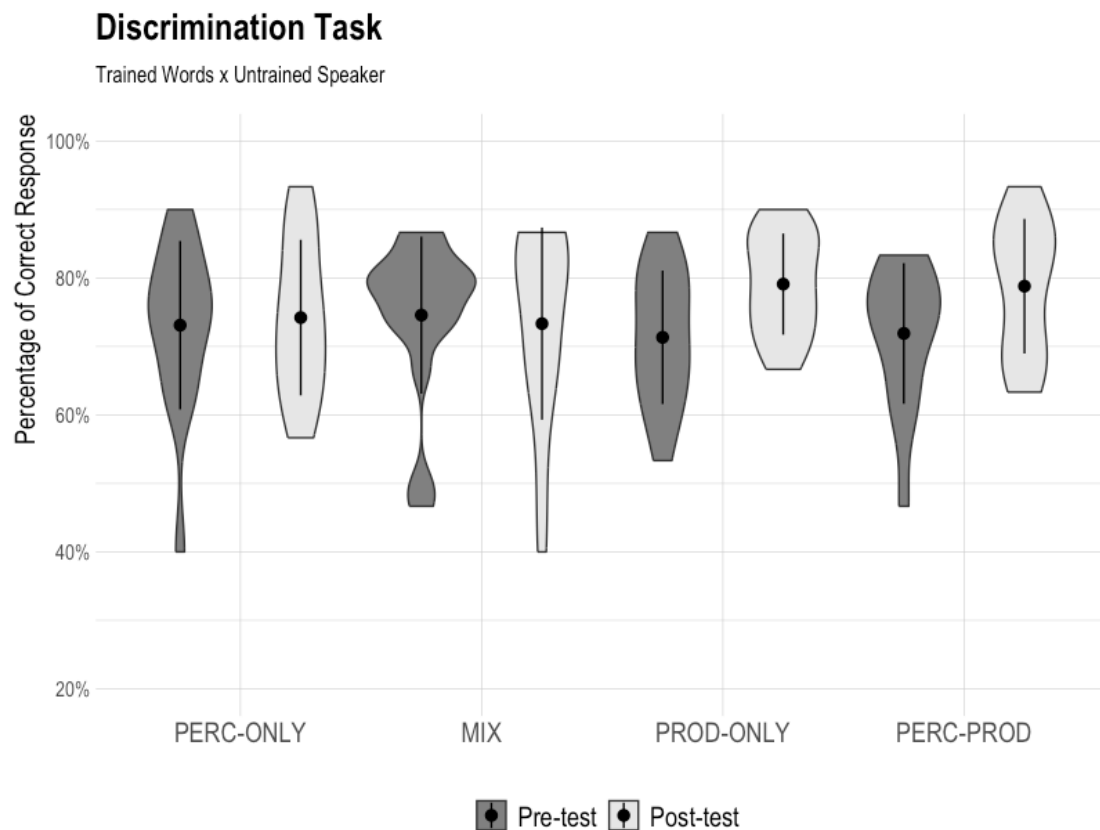
Figure 5.3 illustrates the improvement of the TWTS stimuli, which were used during training. At pre-test, the PROD-ONLY group and PERC-PROD group received similar accuracy scores, 78.7% and 78.6%, respectively. The PERC-ONLY group scored 75.6%, followed by the MIX group at 72.9%. Following the training, the MIX group increased their score by 11.7%, which was the largest improvement. The PERC-ONLY group increased by 7.3%, and the PERC-PROD group increased by 6.7%. The PROD-ONLY group exhibited the smallest improvement, increasing by only 2.0%.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.3 Experiment 3: Discrimination Task Results with the TWTS Stimuli

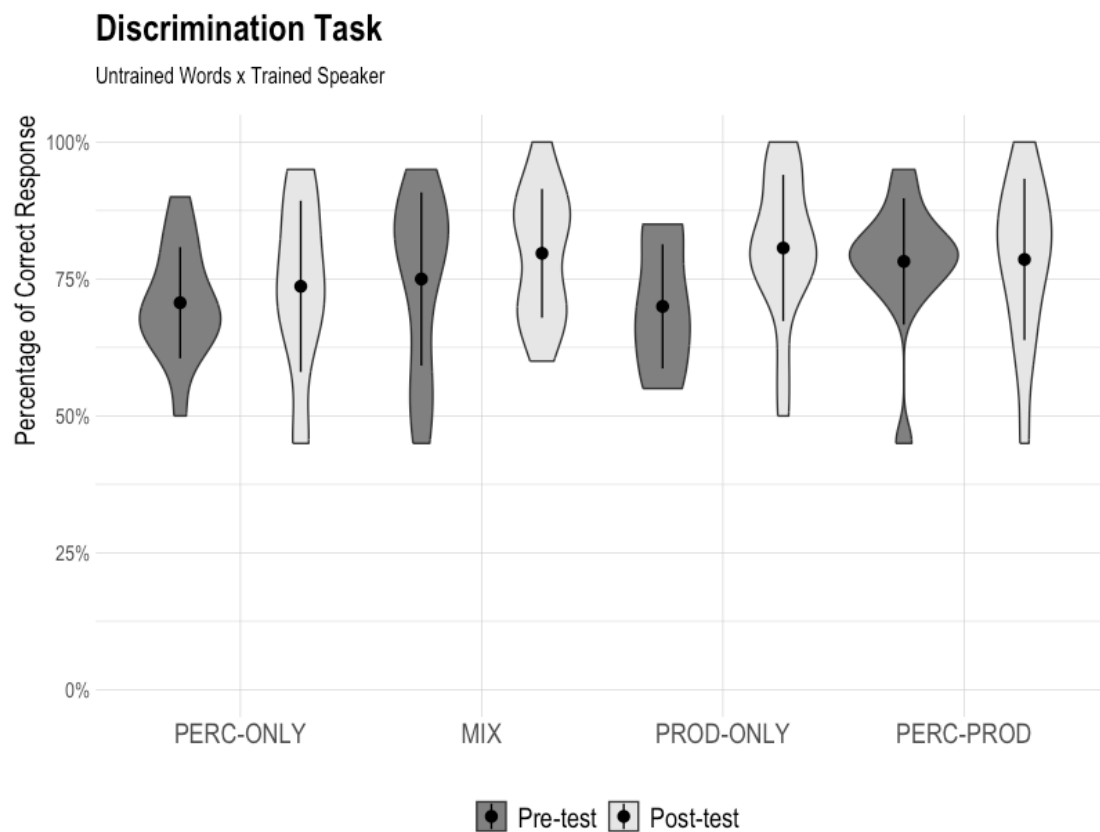
Figure 5.4 presents the discrimination test results for the TWUS stimuli. At pre-test, the MIX group scored 74.6%, the highest of the four groups, followed by the PERC-ONLY group's 73.1%. The PERC-PROD group achieved a score of 71.9%, while the PROD-ONLY group scored the lowest at 71.3%. After training, the PROD-ONLY group increased by 7.8%, followed by the PERC-PROD group with a 6.9% increase. The PERC-ONLY group improved by only 1.1% from pre-test, while the MIX group's accuracy score decreased by 1.3% after training.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.4 Experiment 3: Discrimination Task Results with the TWUS Stimuli

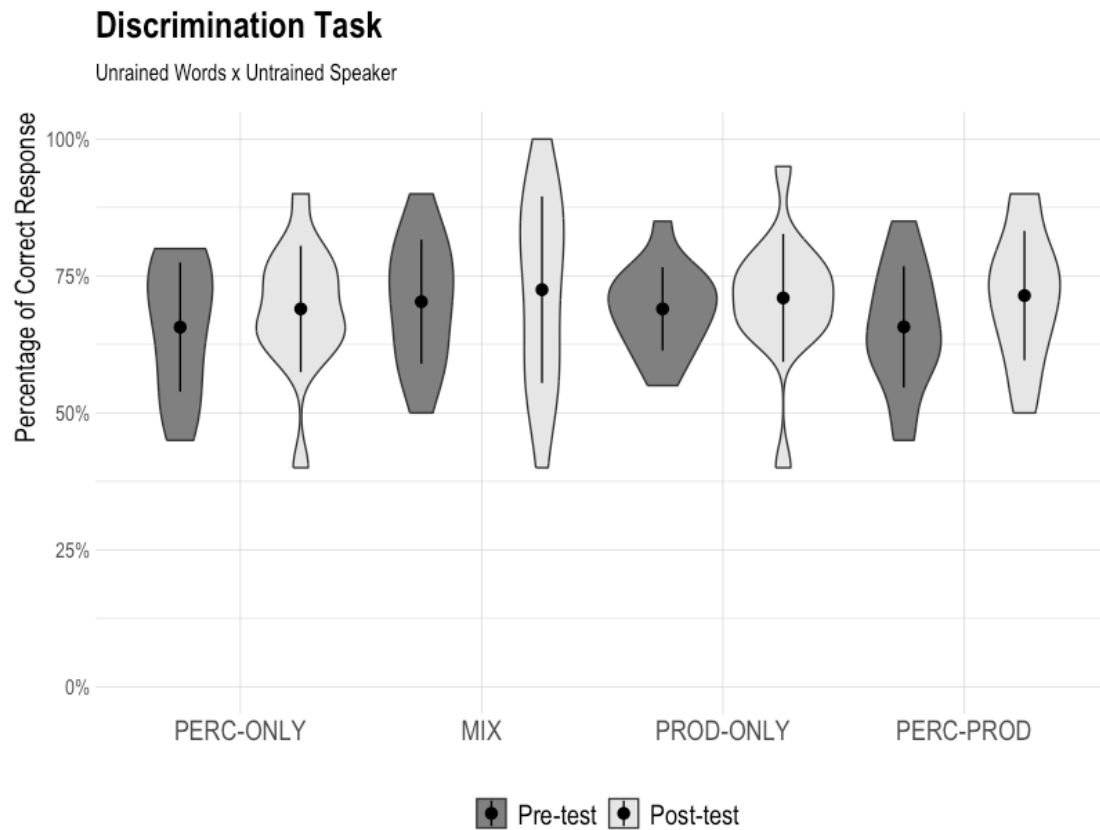
Figure 5.5 shows the results for the UWTS stimuli. The PERC-PROD group scored the highest accuracy score at pre-test with 78.2%, followed by the MIX group with 75.0% accuracy. The PROD-ONLY and PERC-ONLY groups received close accuracy scores, 70.0% and 70.7%, respectively. After training, the PROD-ONLY group increased by 10.7% and reached 80.7% accuracy. The MIX group increased by 4.7%, attaining 79.7% accuracy. The PERC-ONLY group improved by 3.0% from their pre-test score. The PERC-PROD group's score remained nearly unchanged with a 0.4% improvement.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.5 Experiment 3: Discrimination Task Results with the UWTS Stimuli

Figure 5.6 shows the results for the UWUS stimuli. The MIX group attained 70.3% accuracy at pre-test, the highest accuracy score among all groups, followed by the PROD-ONLY group with 69.0%. Both the PERC-ONLY group and PERC-PROD group received close accuracy scores of 65.7%. At post-test, the PERC-PROD group improved the most (5.7%), followed by the PERC-ONLY group with a 3.3% increase. The PROD-ONLY group and MIX group demonstrated similar improvement rates, with increases of 2.0% and 2.2%, respectively.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.6 Experiment 3: Discrimination Task Results with the UWUS Stimuli

5.3.1.2 Inferential Statistics

A logistic mixed-effects model was fitted to discrimination task data with the binary score as the dependent variable and Group, Time, and Stimulus type as independent variables, and the reference level was the PERC-ONLY group's pre-test with the TWTS stimuli (see the model output in Appendix G). The model revealed a significant three-way interaction between Time, Group, and Stimuli (Time $\times$ Group [PROD-ONLY] $\times$ Stimuli [UWTS]: $p = 0.027$; Time $\times$ Group [PROD-

ONLY] $\times$ Stimuli [TWUS]: $p = 0.024$), suggesting that the improvement patterns of the PROD-ONLY group on the TWUS and UWTS stimuli were significantly different compared to that of the PERC-ONLY group.

The Time $\times$ Group interaction was not significant, suggesting that the improvement from pre-test to post-test in the PERC-ONLY group for the TWTS stimuli was not significantly different from the improvements in the other groups. The interaction between Time and Stimuli did not reach significance, indicating no significant differences in performance between the stimulus types over time in the PERC-ONLY group. No significant interaction between Group and Stimuli was found either, suggesting that there were no significant differences in performance on different stimulus types between groups compared to the PERC-ONLY group.

A significant Time effect was found ($p = 0.005$), indicating that the PERC-ONLY group's score with the TWTS stimuli significantly increased from pre-test to post-test. No significant Group effect was observed, suggesting that there were no significant differences comparing groups with the PERC-ONLY group at pre-test. A significant Stimuli effect was found for the UWUS stimuli ($p = 0.015$), showing that this stimulus type was more difficult than the TWTS stimuli in the PERC-ONLY group at pre-test.

Pairwise comparisons between groups were conducted to examine whether there were any significant differences in discrimination accuracy for each stimulus type among the groups at pre-test. The results revealed no significant differences between the groups before training. Next, pairwise comparisons across groups and stimuli at pre-test were performed to investigate whether there were any significant differences in how participants responded to each stimulus type (Table 5.3).

Table 5.3 Experiment 3: Summary of Pairwise Comparisons across Groups and Stimuli at Pre-test for the Discrimination Task

<i>Groups</i>	<i>Comparisons</i>		<i>Odds Ratios</i>	<i>SE</i>	<i>z</i>	<i>p</i>
PERC-ONLY	TWTS	TWUS	1.152	0.185	0.886	0.812
		UWTS	1.370	0.320	1.347	0.533
		UWUS	1.754	0.405	2.433	0.071
	TWUS	UWTS	1.189	0.276	0.745	0.879
		UWUS	1.522	0.349	1.832	0.258
	UWTS	UWUS	1.280	0.232	1.362	0.523
MIX	TWTS	TWUS	0.906	0.142	-0.631	0.922
		UWTS	0.925	0.216	-0.332	0.987
		UWUS	1.201	0.276	0.799	0.855
	TWUS	UWTS	1.021	0.240	0.090	1.000
		UWUS	1.326	0.306	1.223	0.612
	UWTS	UWUS	1.298	0.242	1.400	0.499
PROD-ONLY	TWTS	TWUS	1.541	0.250	2.665	0.039
		UWTS	1.719	0.405	2.300	0.098
		UWUS	1.808	0.425	2.520	0.057
	TWUS	UWTS	1.115	0.257	0.474	0.965
		UWUS	1.173	0.270	0.695	0.899
	UWTS	UWUS	1.052	0.192	0.275	0.993
PERC-PROD	TWTS	TWUS	1.495	0.254	2.369	0.083
		UWTS	1.081	0.270	0.313	0.989
		UWUS	2.133	0.510	3.169	0.008
	TWUS	UWTS	0.723	0.177	-1.324	0.547
		UWUS	1.427	0.334	1.520	0.425
	UWTS	UWUS	1.973	0.393	3.414	0.004

In the PERC-ONLY group, there were no significant differences across most stimulus types. However, the comparison between the TWTS and UWUS stimuli approached significance ($p = 0.071$), suggesting that the UWUS stimuli may have been more challenging than the TWTS stimuli. In the MIX group, there were no significant differences between stimulus types, suggesting that participants found all stimulus types to be of similar difficulty before training. For the PROD-ONLY group, a significant difference was observed between the TWTS and TWUS stimuli ($p = 0.039$), indicating that participants found the TWUS stimuli significantly more challenging than the TWTS stimuli. In addition, there was a difference approaching significance between the TWTS and UWUS stimuli ($p = 0.057$). In the PERC-PROD group, a significant difference was found between the TWTS and UWUS stimuli ($p = 0.008$) and the UWTS and UWUS stimuli ($p = 0.004$). These

findings suggest that participants in the PROD-ONLY and PERC-PROD groups found stimuli spoken by an untrained speaker more difficult than stimuli spoken by a trained speaker.

Finally, pairwise comparisons between pre-test and post-test were conducted across groups and stimulus types for the discrimination task (Table 5.4). In the PERC-ONLY group, there was a significant improvement only with the TWTS stimuli ($p = 0.005$), indicating that participants' discrimination accuracy significantly improved after training, but the learning did not generalise to untrained words and voice. The MIX group also showed a significant improvement with the TWTS stimuli ($p < 0.001$). However, like the PERC-ONLY group, no significant changes were found for the stimuli that did not appear during training.

Table 5.4 Experiment 3: Summary of Pre-test and Post-test Pairwise Comparisons across Groups and Stimuli for the Discrimination Task

<i>Groups</i>	<i>Stimuli</i>	<i>Odds Ratios</i>	<i>SE</i>	<i>z</i>	<i>p</i>
PERC-ONLY	TWTS	0.613	0.106	-2.827	0.005
	TWUS	0.938	0.149	-4.663	0.689
	UWTS	0.853	0.160	-0.78	0.397
	UWUS	0.850	0.153	-2.629	0.367
MIX	TWTS	0.448	0.077	-0.4	<0.001
	TWUS	1.077	0.169	0.474	0.635
	UWTS	0.745	0.148	-2.834	0.138
	UWUS	0.888	0.163	-2.458	0.519
PROD-ONLY	TWTS	0.874	0.151	-0.846	0.435
	TWUS	0.630	0.103	-1.485	0.005
	UWTS	0.541	0.107	-3.106	0.002
	UWUS	0.904	0.166	-0.107	0.581
PERC-PROD	TWTS	0.607	0.115	-0.902	0.009
	TWUS	0.658	0.112	-0.645	0.014
	UWTS	0.978	0.207	-0.553	0.915
	UWUS	0.749	0.142	-1.521	0.128

In contrast, participants in the PROD-ONLY group significantly improved on both the TWUS ($p = 0.005$) and UWTS ($p = 0.002$) stimuli. These results suggest that those in the PROD-ONLY group benefitted from training. Lastly, the PERC-PROD group showed significant improvement on both the TWTS ($p = 0.009$) and TWUS ($p = 0.014$) stimuli, suggesting that participants in the PERC-PROD group benefitted from training, and their learning generalised to a new voice.

5.3.2 Identification task

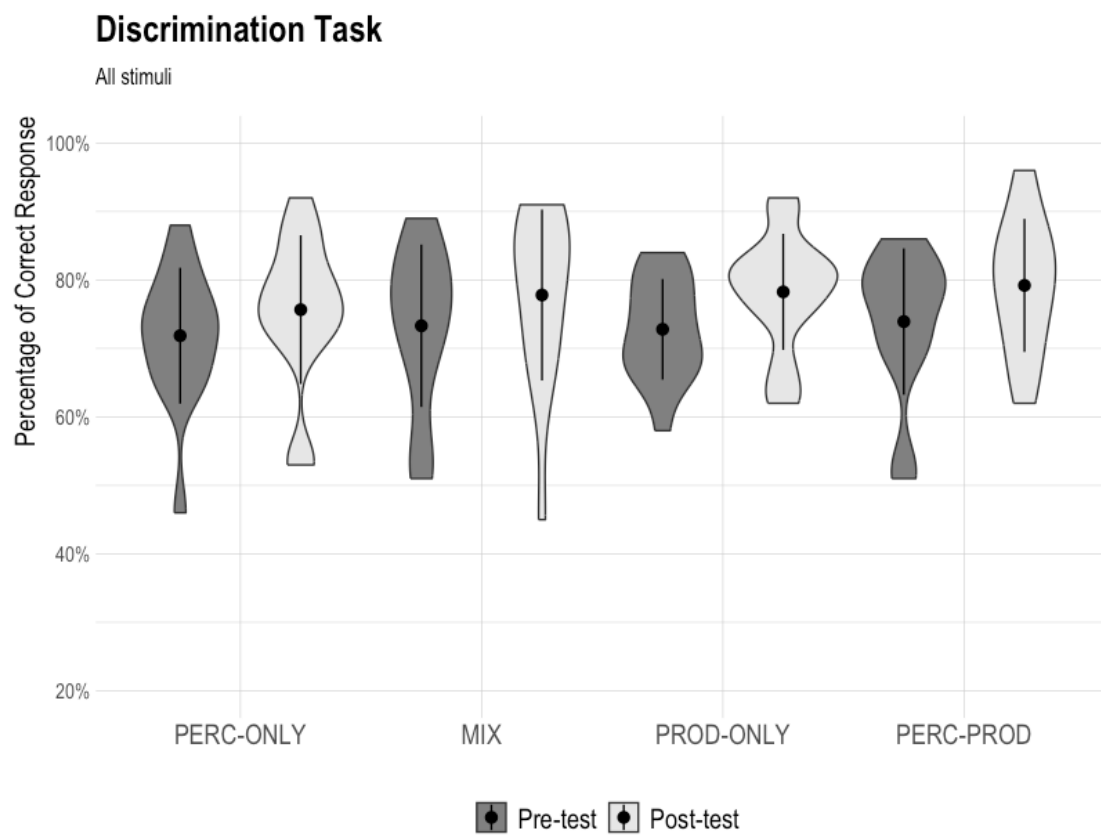
5.3.2.1 Descriptive Statistics

Table 5.5 presents the mean number of correct responses, standard deviations, and the percentage of correct responses in identification tasks per training group and familiarity type of stimuli. The maximum score of all stimuli was 100, combining the scores of four stimulus types. The TWTS stimuli and TWUS stimuli had a maximum of 40 points each, and the UWTS stimuli and UWUS stimuli had a maximum of 10 points each.

Table 5.5 Experiment 3: Mean, Standard Deviations, and Percentage of Correct Response at the Pre- and Post-tests of the Identification Task

		PERC-ONLY		MIX		PROD-ONLY		PERC-PROD	
		Mean (S.D.)	Percentage	Mean (S.D.)	Percentage	Mean (S.D.)	Percentage	Mean (S.D.)	Percentage
All stimuli	Pre-test	55.60 (9.12)	56.6%	58.31 (12.23)	58.3%	59.20 (12.10)	59.2%	60.50 (7.64)	60.5%
	Post-test	60.00 (11.22)	60.0%	64.44 (13.76)	64.4%	65.53 (12.76)	65.5%	63.29 (10.08)	63.3%
Trained words x Trained speaker	Pre-test	17.20 (3.97)	57.3%	17.56 (5.09)	58.5%	18.27 (5.51)	60.9%	17.79 (3.07)	59.3%
	Post-test	18.27 (4.37)	60.9%	20.63 (5.19)	68.8%	20.67 (4.81)	68.9%	19.71 (3.54)	65.7%
Trained words x Untrained speaker	Pre-test	16.07 (3.33)	53.6%	16.63 (3.52)	55.4%	17.47 (2.70)	58.2%	19.14 (2.68)	63.8%
	Post-test	17.93 (5.02)	59.8%	19.06 (4.52)	63.5%	19.27 (3.28)	64.2%	19.07 (3.08)	63.6%
Untrained words x Trained speaker	Pre-test	10.87 (2.90)	54.3%	12.75 (3.04)	63.8%	12.47 (3.52)	62.3%	11.57 (2.98)	57.9%
	Post-test	12.53 (2.42)	62.7%	12.75 (3.36)	63.8%	13.47 (2.95)	67.3%	12.71 (2.87)	63.6%
Untrained words x Untrained speaker	Pre-test	11.47 (2.13)	57.3%	11.38 (2.99)	56.9%	11.00 (2.33)	55.0%	12.00 (1.62)	60.0%
	Post-test	11.27 (1.83)	56.3%	12.00 (2.50)	60.0%	12.13 (3.11)	60.7%	11.79 (2.36)	58.9%

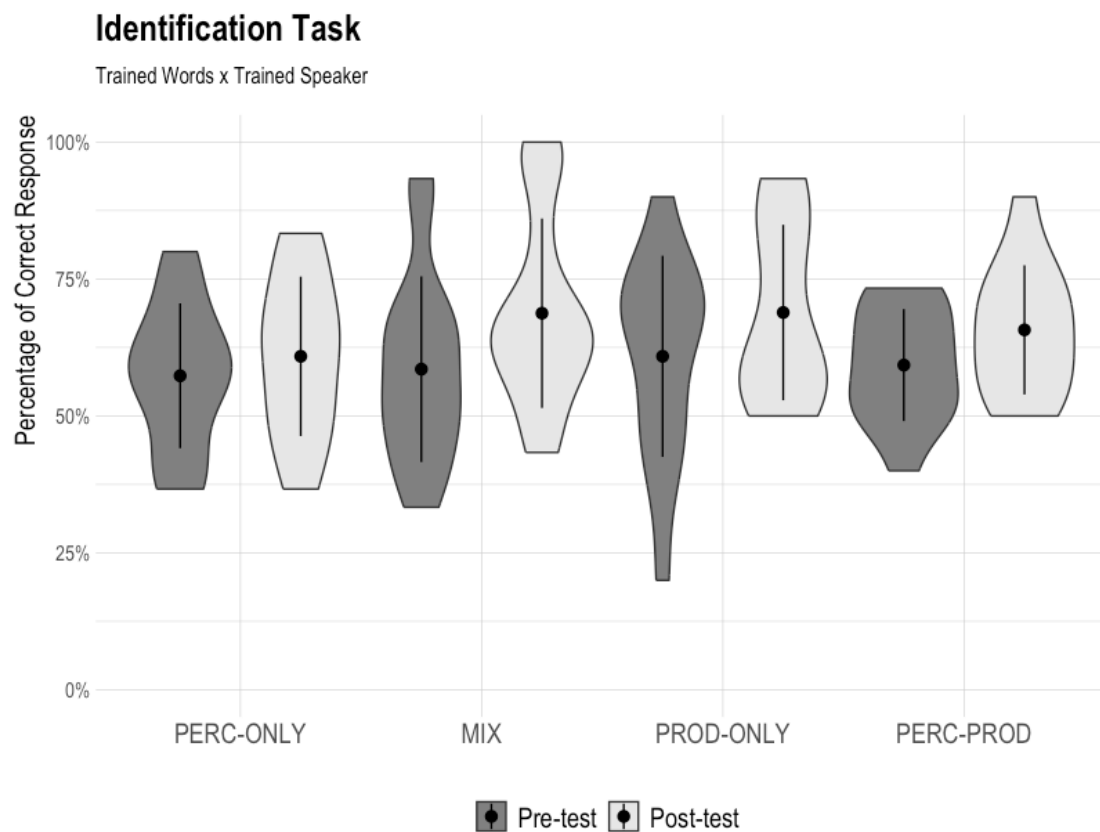
Figure 5.7 illustrates the identification test results with all stimuli. At pre-test, the groups scored in the range of 56.6% to 60.5%. After training, the PROD-ONLY group and MIX group increased their scores by 6.3% and 6.1%, respectively. The PERC-ONLY group and PERC-PROD group improved slightly less than the other two groups, with increases of 3.4% and 2.8%, respectively.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.7 Experiment 3: Identification Task Results with All Stimuli

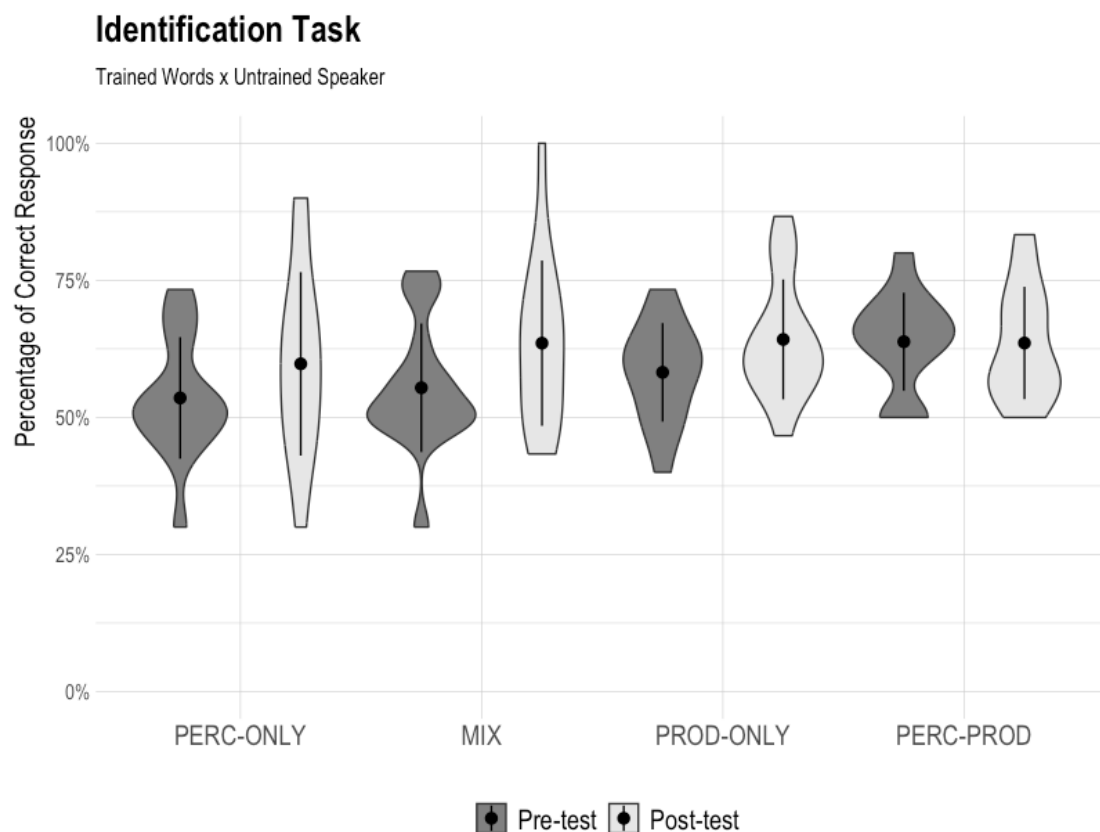
Figure 5.8 depicts the identification test results for the TWTS stimuli. At pre-test, all groups achieved similar accuracy scores ranging from 57.3% to 60.9%. After training, notable differences emerged: the MIX group increased their accuracy score by 10.2%, while the PERC-ONLY group did so by only 3.6%. The PROD-ONLY group and PERC-PROD group's improvement rates were 8.9% and 6.4%, respectively.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.8 Experiment 3: Identification Task Results with the TWTS stimuli

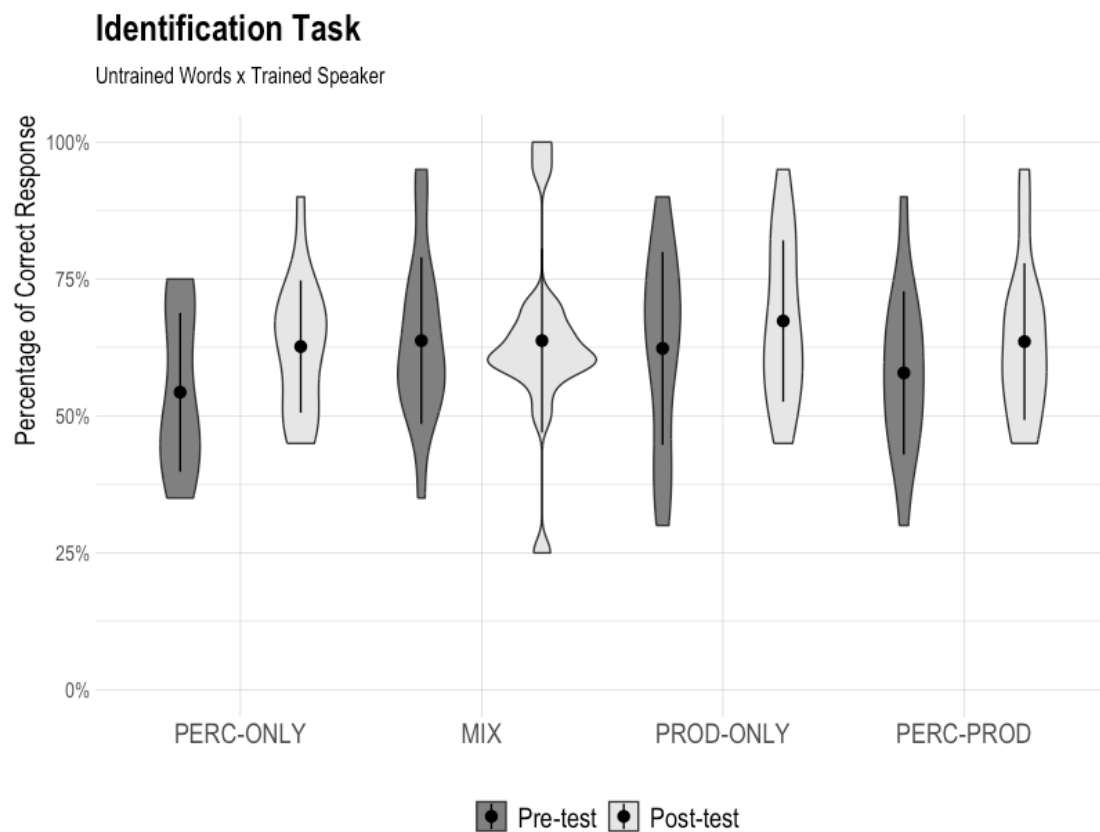
Figure 5.9 presents the test results for the TWUS stimuli. Prior to training, the PERC-PROD group achieved an accuracy of 63.8%, which was nearly 10% higher than the PERC-ONLY group's pre-test score. The PROD-ONLY group attained an accuracy of 58.2%, while the MIX group scored 55.4%. After training, the MIX group's accuracy score increased by 8.1%, representing the highest improvement among the four groups. The PERC-ONLY group and PROD-ONLY group increased by about 6%. The PERC-PROD group's accuracy score remained nearly unchanged (-0.2%); however, it is noteworthy that their post-test score was comparable to those of the PROD-ONLY group and MIX group and higher than that of the PERC-ONLY group.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.9 Experiment 3: Identification Task Results with the TWUS Stimuli

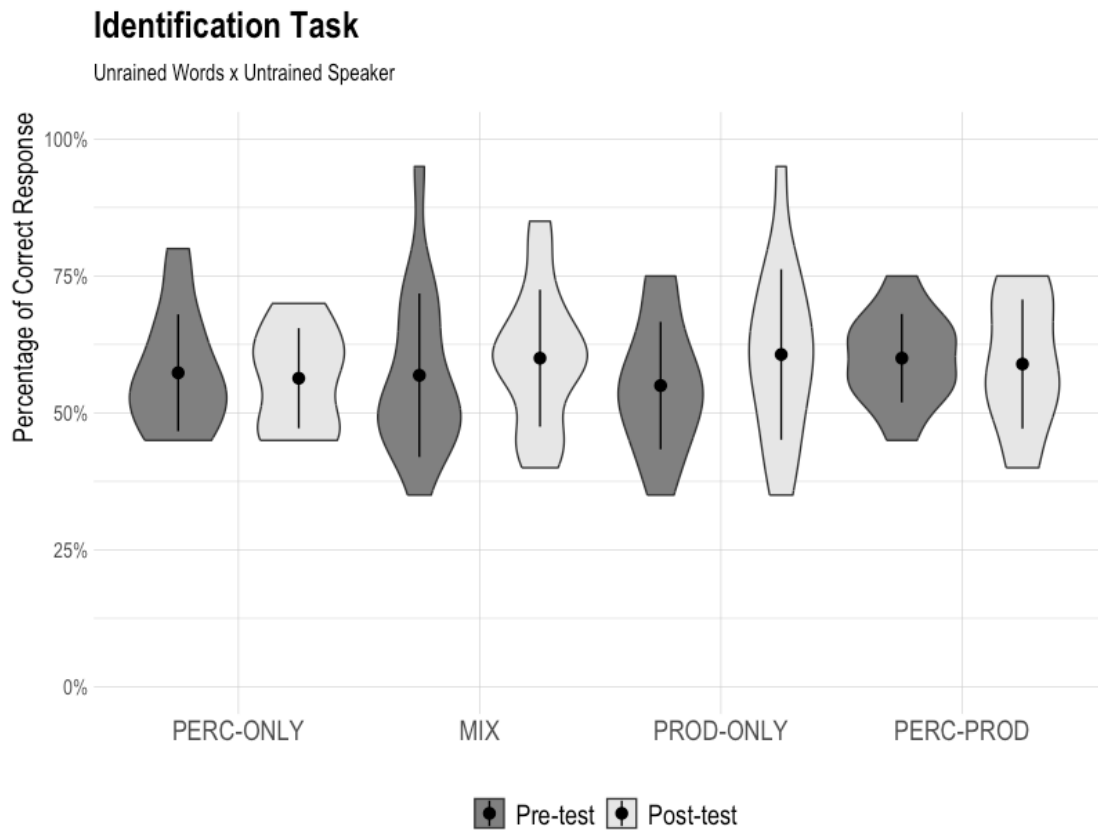
Figure 5.10 illustrates the identification test results for the UWTS stimuli. Before training, both the PROD-ONLY group and MIX group achieved over 60% accuracy (62.3% and 63.8%, respectively). The PERC-PROD group attained 54.3% accuracy and the PERC-ONLY group 57.9% accuracy. After training, the PERC-ONLY group improved by 8.3% in their identification accuracy, which was the largest improvement. The PERC-PROD group's accuracy score increased by 5.7%. The PROD-ONLY group improved by 5.0%, achieving a post-test accuracy of 67.3%, which was the highest among the four groups. The MIX group's accuracy rate remained unchanged.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.10 Experiment 3: Identification Task Results with the UWTS Stimuli

Lastly, Figure 5.11 displays the pre- and post-test results for the UWUS stimuli. At pre-test, the groups scored between 55.0% to 60.0%. The PROD-ONLY group received the lowest score (55.0%), while the PERC-PROD group received the highest score among the groups (60.0%). After training, The PROD-ONLY group improved their identification accuracy by 5.7%, followed by the MIX group with a 3.1% improvement. In contrast, both the PERC-ONLY group and PERC-PROD group decreased their score by approximately 1%.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.11 Experiment 3: Identification Task Results with the UWUS Stimuli

5.3.2.2 Inferential Statistics

A logistic mixed-effects model was fitted to the data with all stimuli with the reference level set to the PERC-ONLY group's pre-test with the TWTS stimuli (see Appendix H). The model revealed a significant three-way interaction between Time, Group, and Stimuli (Time $\times$ Group [MIX] $\times$ Stimuli [UWTS]: $p = 0.027$). This implies that the MIX group's improvement in discrimination accuracy on the UWTS stimuli was significantly smaller compared to the PERC-

ONLY groups' improvement on the TWTS stimuli. Besides this three-way interaction, none of the two-way interactions were significant, suggesting that improvements over time and across stimulus types were similar across groups compared to that of the PERC-ONLY group with the TWTS stimuli. The non-significant Time effect suggests that the PERC-ONLY group did not significantly improve their identification accuracy on the TWTS stimuli from pre-test to post-test. Group was not significant either, indicating that the PERC-ONLY group's identification accuracy on the TWTS stimuli at pre-test was similar to the other groups' accuracy at pre-test. Finally, the non-significant Stimuli effect suggests that identification accuracy for the TWTS stimuli was not significantly different from the other stimulus types at pre-test.

Pairwise comparisons between groups were conducted to examine whether there were any significant differences in identification accuracy for each stimulus type among the groups at pre-test. The results revealed no significant differences between the groups before training. Next, pairwise comparisons across groups and stimuli at pre-test were performed to investigate whether there were any significant differences in how participants responded to each stimulus type. The results showed no significant differences across stimulus types in each group either.

Table 5.6 presents a summary of pairwise comparisons between pre-test and post-test across groups and stimulus types for the identification task. In the PERC-ONLY group, there was a significant improvement only on the TWUS stimuli ($p = 0.049$) and UWTS stimuli ($p = 0.033$), indicating that participants' identification accuracy improved significantly after training for these stimuli. The MIX group showed a significant improvement on the TWTS stimuli ($p = 0.001$) and TWUS stimuli ($p = 0.007$). For the PROD-ONLY and PERC-PROD groups, there was a significant improvement only on the TWTS stimuli (PROD-ONLY: $p = 0.009$; PERC-PROD: $p = 0.045$), while the TWUS score of the PROD-ONLY group was approaching significance ($p = 0.045$).

Table 5.6 Experiment 3: Summary of Pre-test and Post-test Pairwise Comparisons across Groups and Stimuli for the Identification Task

<i>Groups</i>	<i>Stimuli</i>	<i>Odds Ratios</i>	<i>SE</i>	<i>z</i>	<i>p</i>
PERC-ONLY	TWTS	0.852	0.120	-1.135	0.256
	TWUS	0.759	0.107	-1.965	0.049
	UWTS	0.693	0.119	-2.136	0.033
	UWUS	1.044	0.178	0.255	0.799
MIX	TWTS	0.612	0.087	-3.453	0.001
	TWUS	0.687	0.096	-2.703	0.007
	UWTS	1.000	0.171	0.000	1.000
	UWUS	0.869	0.146	-0.839	0.401
PROD-ONLY	TWTS	0.681	0.100	-2.625	0.009
	TWUS	0.758	0.109	-1.934	0.053
	UWTS	0.791	0.140	-1.326	0.185
	UWUS	0.778	0.134	-1.458	0.145
PERC-PROD	TWTS	0.743	0.110	-2.001	0.045
	TWUS	1.011	0.151	0.073	0.942
	UWTS	0.775	0.138	-1.427	0.154
	UWUS	1.048	0.186	0.266	0.790

5.3.3 Summary of results on perceptual learning

Table 5.7 summarises the p values from the pairwise comparisons in both discrimination and identification tasks. In the discrimination task, all groups except for the PROD-ONLY group significantly improved discrimination accuracy on the TWTS stimuli, while the PROD-ONLY group improved on the TWUS and UWTS stimuli. Learning generalised to a new voice in the PERC-PROD group but not in the PERC-ONLY and MIX groups. In the identification task, all groups except for the PERC-ONLY group significantly improved identification accuracy on the TWTS stimuli, while the PERC-ONLY group improved on the TWUS and UWTS stimuli. Learning generalised to a new voice in the MIX group but not in the PROD-ONLY and PERC-PROD groups. However, the p value for the improvement on the TWUS in the PROD-ONLY group was approaching significance.

Table 5.7 Experiment 3: Summary of Pairwise Comparison Results (p values) for Perceptual Learning

Discrimination task	PERC-ONLY	MIX	PROD-ONLY	PERC-PROD
Trained words × Trained speaker	0.005	<0.001	×	0.009
Trained words × Untrained speaker	×	×	0.005	0.014
Untrained words × Trained speaker	×	×	0.002	×
Untrained words × Untrained speaker	×	×	×	×
Identification task				
Trained words × Trained speaker	×	0.001	0.009	0.045
Trained words × Untrained speaker	0.049	0.007	(0.053)	×
Untrained words × Trained speaker	0.033	×	×	×
Untrained words × Untrained speaker	×	×	×	×

5.3.4 Identification rating task

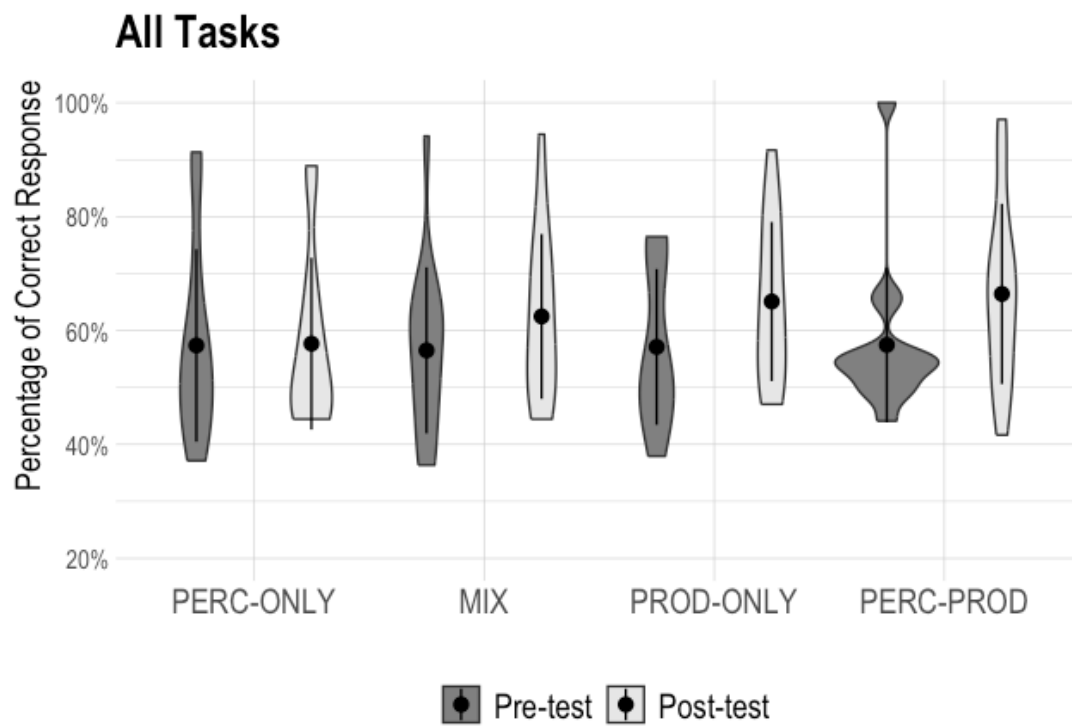
5.3.4.1 Descriptive statistics

Table 5.8 displays the percentage of correct responses in the identification rating task per training group, speaking task, and word type.

Table 5.8 Experiment 3: Mean Identification Rating Scores

			PERC-ONLY	MIX	PROD-ONLY	PERC-PROD
All Tasks	Combined	Pre-test	56.2%	56.6%	57.3%	57.5%
		Post-test	57.5%	62.5%	64.8%	66.3%
	Trained	Pre-test	57.7%	56.1%	56.8%	57.3%
		Post-test	53.9%	65.5%	64.6%	66.5%
	Untrained	Pre-test	54.7%	57.1%	57.8%	57.7%
		Post-test	61.1%	59.6%	64.9%	66.0%
Word	Combined	Pre-test	55.0%	64.0%	67.9%	62.5%
		Post-test	59.5%	72.1%	76.7%	79.2%
	Trained	Pre-test	58.4%	64.9%	66.7%	64.1%
		Post-test	53.6%	75.8%	75.9%	82.1%
	Untrained	Pre-test	51.4%	63.0%	69.0%	61.0%
		Post-test	65.5%	68.4%	77.5%	76.2%
Sentence	Combined	Pre-test	59.7%	53.8%	52.4%	56.0%
		Post-test	56.4%	58.7%	56.6%	60.2%
	Trained	Pre-test	58.5%	50.0%	51.2%	52.4%
		Post-test	55.6%	61.7%	55.7%	57.8%
	Untrained	Pre-test	61.0%	58.0%	53.8%	59.7%
		Post-test	57.1%	55.8%	57.5%	62.7%
Picture	Combined	Pre-test	53.7%	51.7%	51.7%	54.0%
		Post-test	56.6%	56.7%	60.9%	59.1%
	Trained	Pre-test	56.0%	53.3%	52.8%	55.6%
		Post-test	52.5%	58.7%	62.5%	59.5%
	Untrained	Pre-test	51.4%	50.0%	50.6%	52.5%
		Post-test	60.8%	54.7%	59.3%	58.8%

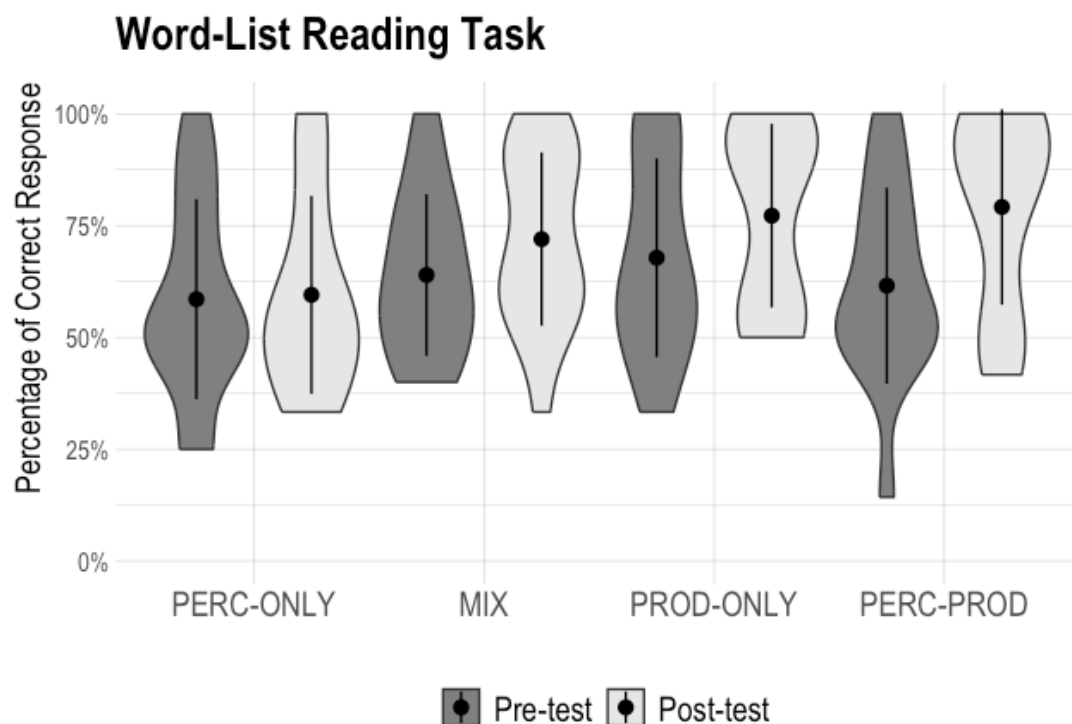
Figure 5.12 illustrates the changes in the identification rating scores of all tasks combined including trained and untrained words from pre-test to post-test. Initially, all groups started with similar scores ranging from 56% to 57%. After training, the PERC-PROD group exhibited the highest improvement (8.8%), followed by the PROD-ONLY group (7.5%) and MIX group (5.9%). The PERC-ONLY group improved by only 1.3%.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.12 Experiment 3: Identification Rating Task for All Tasks

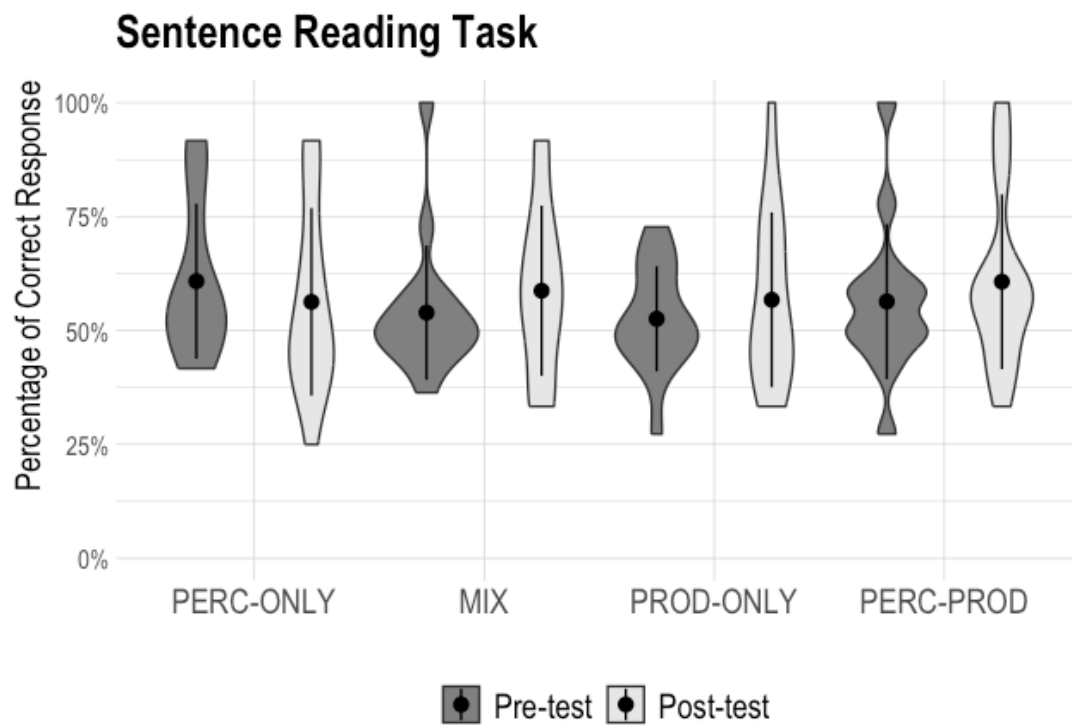
The scores of the word-list reading task followed a similar trend to the overall improvement observed across all groups. Figure 5.13 displays the pre- and post-test scores for the word-list reading task. The PERC-ONLY group began with the lowest accuracy at 55.0%, while the other groups scored over 60% at pre-test (MIX: 64.0%, PROD-ONLY: 76.9%, PERC-PROD: 62.5%). After training, the PERC-PROD group showed the greatest improvement (16.7%), followed by the PROD-ONLY group (8.8%) and the MIX group (8.1%). The PERC-ONLY group showed the least improvement (4.5%), despite starting with the lowest pre-test mean score.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.13 Experiment 3: Identification Rating Task for the Word-List Reading Task

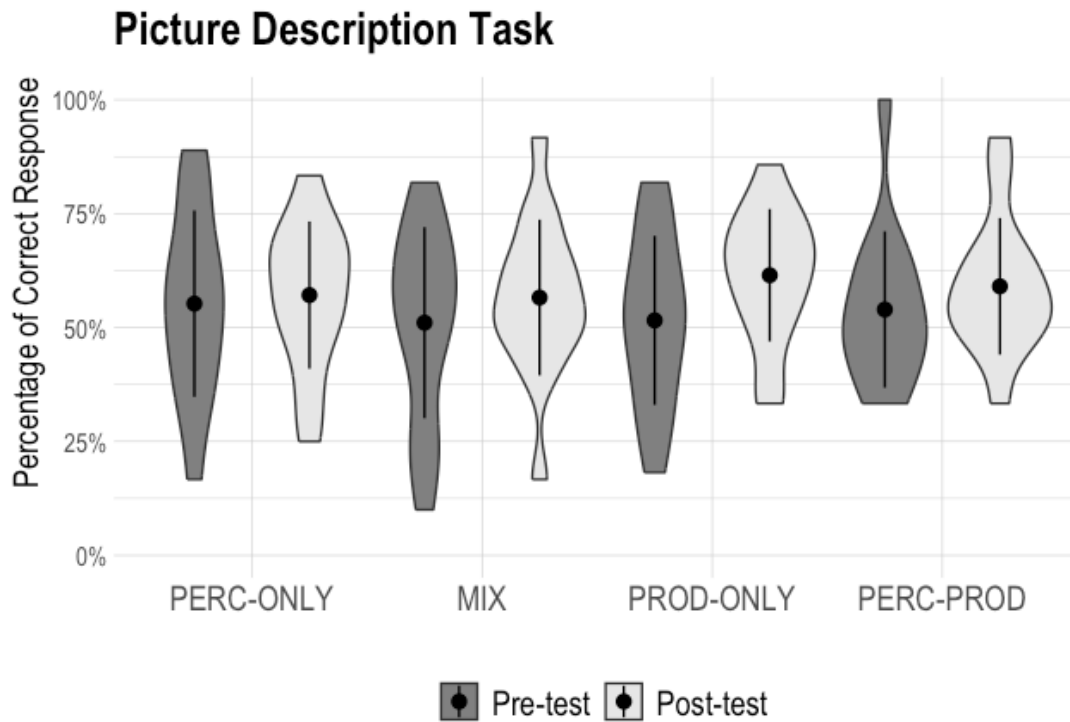
The identification rating scores of the sentence reading task showed a different trend compared to the word-list reading task scores. Figure 5.14 presents the pre- and post-test scores of the sentence reading task. All groups started with scores ranging between 52.4% and 59.7%. After training, the MIX group (4.9%), PERC-PROD group (4.2%), and PROD-ONLY group (4.2%) improved their scores. In contrast, the PERC-ONLY group's score decreased by 3.3%.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.14 Experiment 3: Identification Rating Task for the Sentence Reading Task

Figure 5.15 presents the pre- and post-test scores of the picture description task. All groups had pre-test scores around 53%. After training, the PROD-ONLY group showed the greatest improvement, including their score by 9.2%. The PERC-PROD group and MIX group improved by 5.1% and 5%, respectively. The PERC-ONLY group showed the least improvement, with only a 2.9% increase.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.15 Experiment 3: Identification Rating Task for the Picture Description Task

5.3.4.2 Inferential statistics

A logistic mixed-effects model was fitted to the identification rating score, with the PERC-ONLY group's pre-test score for trained words in the word-list reading task as the reference level (see Appendix I). The four-way interaction (Group, Time, Task, and Type), three-way interactions between Group, Time, and Task, between Group, Task, and Type, and between Time, Task, and Type were not significant, and these interactions were removed from the model. The final model revealed a significant three-way interaction between Group, Time, and Type (MIX: $p = 0.040$). This suggests that the MIX group had a different improvement pattern for trained and untrained words

compared to the PERC-ONLY group. Specifically, participants in the MIX group showed less improvement in the production accuracy of untrained words compared to trained words.

Regarding two-way interactions, the Group and Time interaction was significant (MIX: $p = 0.018$; PROD-ONLY: $p = 0.031$; and PERC-PROD: $p = 0.021$), suggesting that participants in the PERC-ONLY group improved significantly less on trained words compared to the other groups. An interaction between Group and Task was also significant, indicating that participants in the MIX, PROD-ONLY, and PERC-PROD groups answered significantly less correctly to trained words in the sentence reading and picture description tasks compared to the word-list reading task, relative to the PERC-ONLY group. The Time and Task interaction was also significant ($p = 0.028$), suggesting that the PERC-ONLY group's improvement was significantly smaller in the sentence reading task compared to the word-list reading task. The remaining two-way interactions, including the Group and Type interaction and the Time and Type interaction were not significant. These indicate that the difference between trained and untrained words was consistent across groups and time relative to the PERC-ONLY group.

Group, Time, Task, and Type were not significant predictors, suggesting that before training, the PERC-ONLY group's production accuracy did not significantly differ between trained and untrained words in the word-list reading task, nor did it differ among tasks. Moreover, the production accuracy of the PERC-ONLY group on trained words in the word-list reading task did not significantly differ from the other groups. In addition, this group did not significantly improve on trained words in the word-list reading task after training.

Pairwise comparisons were performed to examine whether there were any significant differences between groups for each task at pre-test, and the results showed no significant differences before training. Further pairwise comparisons were carried out to investigate whether the initial identification rating scores significantly differed between trained and untrained words within each group. The results revealed no significant differences between word types.

Next, pairwise comparisons across groups and tasks at pre-test were performed to investigate whether there were any significant differences in participants' performances in each speaking task (Table 5.9). The results showed no significant differences across speaking tasks in the PERC-ONLY group. In the MIX group, participants' production was significantly more target-like in the word-list reading task compared to the picture description task ($p = 0.009$). In the PROD-ONLY group, participants' production was more accurate in the word-list reading task compared to the sentence reading task ($p = 0.001$) and picture description task ($p = 0.002$). Those in the PERC-PROD group

produced the target phonemes more intelligibly in the word-list reading task compared to the picture description task ($p = 0.006$). The difference in performances between the word-list reading and sentence reading tasks were approaching significance ($p = 0.057$).

Table 5.9 Experiment 3: Summary of Pairwise Comparisons across Groups and Tasks at Pre-test for the Identification Rating Score

<i>Groups</i>	<i>Tasks</i>		<i>Odd Ratios</i>	<i>SE</i>	<i>z</i>	<i>p</i>
PERC-ONLY	Word	Sentence	0.818	0.153	-1.072	0.532
		Picture	0.993	0.188	-0.039	0.999
	Sentence	Picture	1.214	0.227	1.035	0.554
MIX	Word	Sentence	1.423	0.251	2.002	0.112
		Picture	1.688	0.300	2.949	0.009
	Sentence	Picture	1.186	0.206	0.985	0.587
PROD-ONLY	Word	Sentence	1.926	0.360	3.505	0.001
		Picture	1.907	0.354	3.480	0.002
	Sentence	Picture	0.990	0.177	-0.056	0.998
PERC-PROD	Word	Sentence	1.543	0.293	2.289	0.057
		Picture	1.792	0.340	3.079	0.006
	Sentence	Picture	1.162	0.214	0.815	0.694

Table 5.10 presents a summary of pairwise comparisons between pre-test and post-test across groups and word types for the identification rating scores. The results revealed that participants in the PERC-ONLY group did not significantly improve their production on either word type from pre-test to post-test. In contrast, the MIX group ($p = 0.015$) and PROD-ONLY group ($p = 0.036$), participants improved only on trained words, while those in the PERC-PROD group significantly improved on both word types after training (Trained: $p = 0.022$; Untrained: $p = 0.036$). The p value for the PROD-ONLY group's improvement on untrained words was approaching significance ($p = 0.069$), suggesting a potential improvement.

Table 5.10 Experiment 3: Summary of Pre-test and Post-test Pairwise Comparisons across Groups and Speaking Tasks for the Identification Rating Score

<i>Groups</i>	<i>Types</i>	<i>Odd Ratios</i>	<i>SE</i>	<i>z</i>	<i>p</i>
PERC-ONLY	Trained	1.202	0.231	0.959	0.337
	Untrained	0.773	0.151	-1.321	0.187
MIX	Trained	0.645	0.116	-2.428	0.015
	Untrained	0.893	0.160	-0.633	0.527
PROD-ONLY	Trained	0.672	0.127	-2.101	0.036
	Untrained	0.706	0.135	-1.819	0.069
PERC-PROD	Trained	0.640	0.125	-2.285	0.022
	Untrained	0.663	0.130	-2.094	0.036

Table 5.11 summarises the results of pairwise comparisons between pre-test and post-test across groups and speaking tasks for the identification rating scores. The results showed that participants in the PERC-ONLY group did not significantly improve in either task, while those in the MIX group significantly improved only in the word-list reading task ($p = 0.004$). In contrast, in the PROD-ONLY and PERC-ONLY groups, participants significantly improved in the word-list reading task (PROD-ONLY: $p = 0.001$; PERC-PROD: $p = 0.000$) and picture description task (PROD-ONLY: $p = 0.032$; PERC-PROD: $p = 0.015$).

Table 5.11 Experiment 3: Summary of Pre-test and Post-test Pairwise Comparisons across Groups and Tasks for the Identification Rating Score

<i>Groups</i>	<i>Tasks</i>	<i>Odd Ratios</i>	<i>SE</i>	<i>z</i>	<i>p</i>
PERC-ONLY	Word	0.794	0.133	-1.375	0.169
	Sentence	1.142	0.189	0.804	0.421
	Picture	0.988	0.165	-0.073	0.942
MIX	Word	0.625	0.101	-2.901	0.004
	Sentence	0.899	0.141	-0.679	0.497
	Picture	0.777	0.123	-1.596	0.111
PROD-ONLY	Word	0.567	0.096	-3.337	0.001
	Sentence	0.816	0.133	-1.249	0.212
	Picture	0.706	0.114	-2.149	0.032
PERC-PROD	Word	0.536	0.092	-3.618	0.000
	Sentence	0.772	0.128	-1.562	0.118
	Picture	0.667	0.111	-2.434	0.015

5.3.5 Goodness rating task

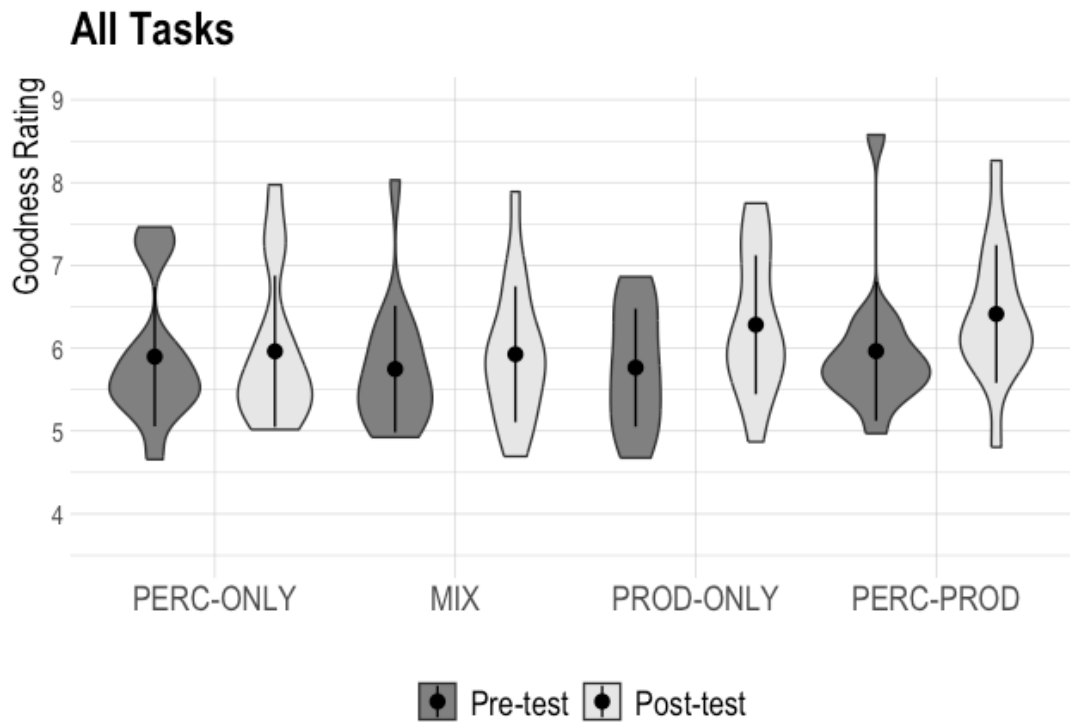
5.3.5.1 Descriptive statistics

Table 5.12 presents the mean score of goodness rating task and the standard deviations for each training group, speaking task, and word type. As mentioned above, the goodness of each token was rated on a scale of 1 to 9.

Table 5.12 Experiment 3: Mean Goodness Rating Scores and the Standard Deviations

			PERC-ONLY	MIX	PROD-ONLY	PERC-PROD
			Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)
All Tasks	Combined	Pre-test	5.84 (2.53)	5.75 (2.51)	5.76 (2.69)	5.96 (2.71)
		Post-test	5.95 (2.64)	5.93 (2.50)	6.26 (2.49)	6.40 (2.69)
	Trained	Pre-test	5.90 (2.42)	5.79 (2.51)	5.73 (2.62)	6.02 (2.67)
		Post-test	5.97 (2.58)	6.05 (2.37)	6.22 (2.44)	6.41 (2.67)
	Untrained	Pre-test	5.77 (2.63)	5.71 (2.50)	5.80 (2.76)	5.90 (2.75)
		Post-test	5.93 (2.70)	5.81 (2.61)	6.30 (2.54)	6.40 (2.72)
Word	Combined	Pre-test	5.76 (2.38)	6.08 (2.37)	6.34 (2.70)	6.38 (2.68)
		Post-test	6.10 (2.79)	6.45 (2.46)	6.82 (2.31)	7.34 (2.35)
	Trained	Pre-test	5.88 (2.17)	6.06 (2.45)	6.32 (2.62)	6.42 (2.64)
		Post-test	6.02 (2.73)	6.74 (2.15)	6.76 (2.18)	7.37 (2.41)
	Untrained	Pre-test	5.62 (2.60)	6.11 (2.29)	6.37 (2.79)	6.34 (2.74)
		Post-test	6.18 (2.86)	6.16 (2.72)	6.87 (2.44)	7.32 (2.30)
Sentence	Combined	Pre-test	5.96 (2.65)	5.54 (2.65)	5.39 (2.64)	5.67 (2.66)
		Post-test	5.76 (2.64)	5.82 (2.54)	6.08 (2.53)	5.93 (2.77)
	Trained	Pre-test	6.07 (2.53)	5.47 (2.59)	5.19 (2.68)	5.53 (2.74)
		Post-test	5.72 (2.62)	5.89 (2.45)	6.12 (2.56)	5.71 (2.85)
	Untrained	Pre-test	5.84 (2.78)	5.61 (2.72)	5.62 (2.59)	5.82 (2.59)
		Post-test	5.81 (2.68)	5.75 (2.63)	6.05 (2.51)	6.15 (2.69)
Picture	Combined	Pre-test	5.78 (2.55)	5.62 (2.47)	5.55 (2.65)	5.83 (2.74)
		Post-test	5.99 (2.47)	5.51 (2.41)	5.87 (2.54)	5.92 (2.69)
	Trained	Pre-test	5.72 (2.57)	5.86 (2.48)	5.70 (2.49)	6.13 (2.58)
		Post-test	6.16 (2.38)	5.51 (2.37)	5.80 (2.50)	6.15 (2.46)
	Untrained	Pre-test	5.85 (2.54)	5.37 (2.46)	5.41 (2.82)	5.52 (2.88)
		Post-test	5.81 (2.55)	5.51 (2.46)	5.95 (2.60)	5.68 (2.91)

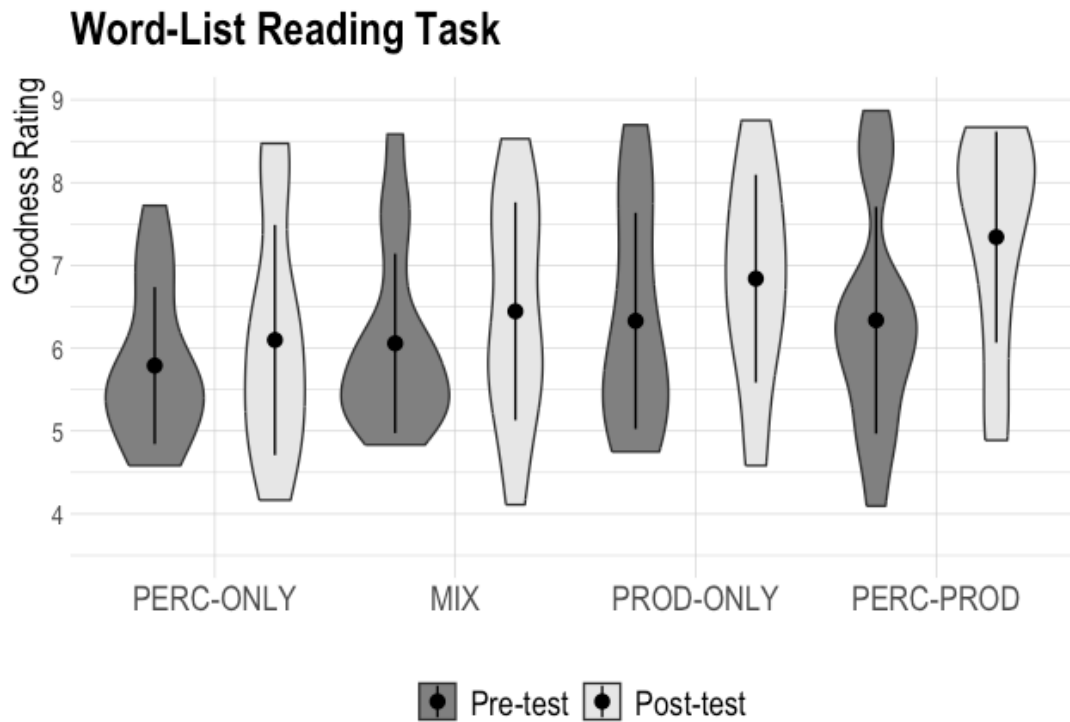
Figure 5.16 depicts the changes in goodness rating scores for all tasks combined, including both trained and untrained words, from pre-test to post-test. All groups started with scores just below 6 points. After training, the PROD-ONLY group showed the highest improvement (0.50 points), followed by the PERC-PROD group (0.44 points). The scores of the MIX group and PERC-ONLY group remained nearly the same, with increases of 0.18 and 0.11 points, respectively.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.16 Experiment 3: Identification Rating Task for All Tasks

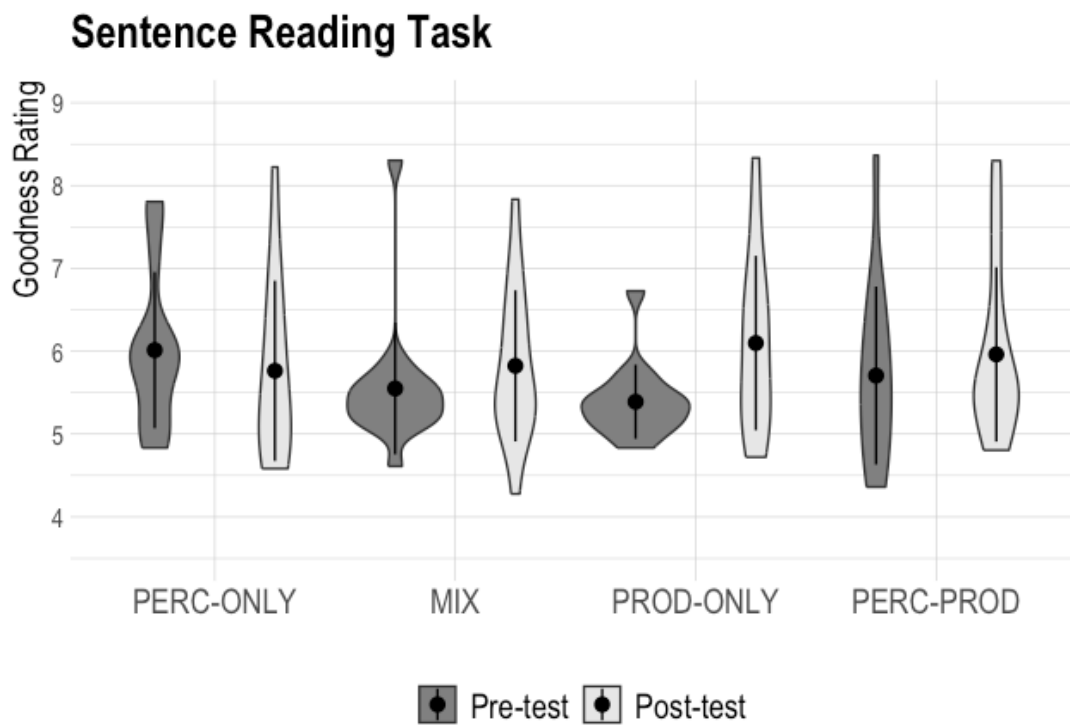
Figure 5.17 presents the pre- and post-test scores of the word-list reading task. At pre-test, only the PERC-ONLY group scored 5.76 points, while the other groups scored above 6 points. After training, the PERC-PROD group showed the highest improvement, increasing by 0.96 points to reach 7.34 points. The other groups showed similar improvements (PROD-ONLY: 0.48 points, MIX: 0.37 points, and PERC-ONLY: 0.34 points).



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.17 Experiment 3: Goodness Rating Task for the Word-List Reading Task

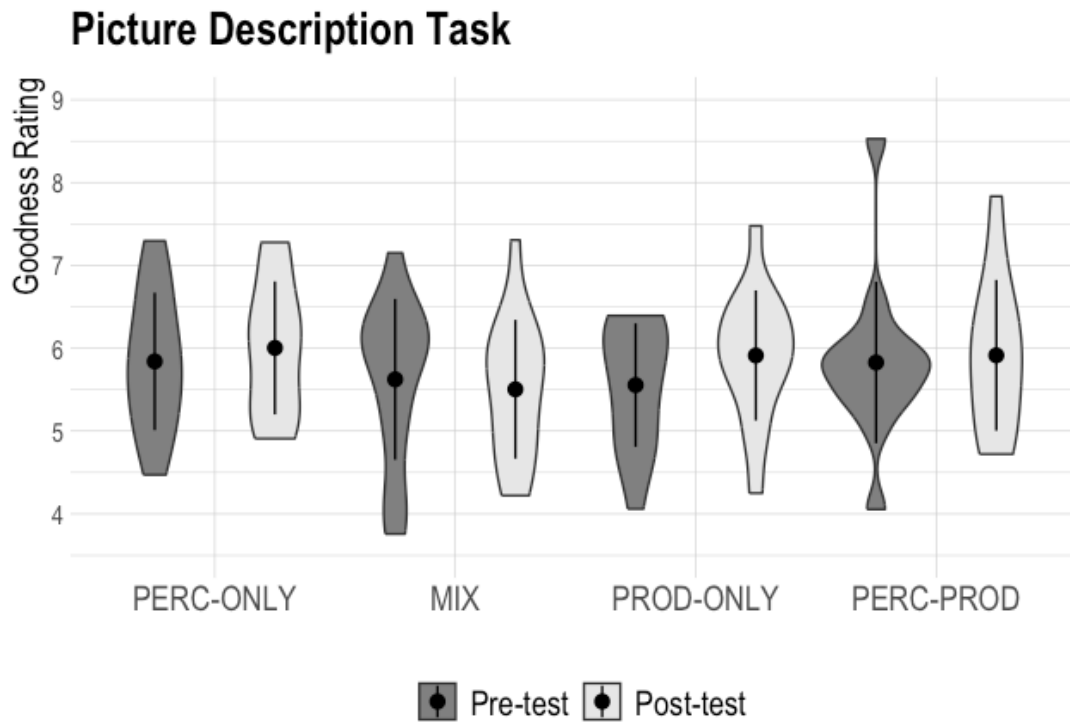
Figure 5.18 displays the pre- and post-test scores of the sentence reading task. At pre-test, the PERC-ONLY group received the highest score (5.96 points), while the other groups scored around 5.5 points. After training, the PROD-ONLY group improved by 0.69 points, followed by the MIX group (0.31 points) and the PERC-PROD group (0.26 points). In contrast, the PERC-ONLY group's mean score decreased by 0.20 points.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.18 Experiment 3: Goodness Rating Task for the Sentence Reading Task

The Figure 5.19 presents the pre- and post-test scores of the picture description task. All groups had pre-test scores between 5.55 and 5.84. After training, the PROD-ONLY group improved by 0.32 points, followed by the PERC-ONLY group (0.21 points). The PERC-PROD group's score remained nearly the same, with a slight improvement of 0.09 points. In contrast, the MIX group's mean score decreased by 0.11 points.



Note. The black dot of the point range shows the mean and the length of the bar shows the range of one standard deviation.

Figure 5.19 Experiment 3: Goodness Rating Task for the Picture Description Task

5.3.5.2 Inferential statistics

A linear mixed-effects model was fitted to the goodness rating scores with the PERC-ONLY group's pre-test scores of the word-list reading task with trained words as the reference level (see Appendix J). Neither the four-way interaction nor the three-way interactions were significant. In addition, the interactions Time $\times$ Type, Group $\times$ Type, and Task $\times$ Type were not significant. Type as a main effect was not significant either, suggesting that word type, whether trained or untrained, did not significantly influence participants' performances.

The final model revealed a significant two-way interaction between Group and Time (PROD-ONLY: $p = 0.045$), indicating that the improvement in the PROD-ONLY group was significantly larger than that of the PERC-ONLY group ($p = 0.045$). The Time and Task interaction was also significant (Picture: $p = 0.034$), suggesting that the improvement in the word-list reading task was significantly larger than in the picture description task. A significant interaction was found between Group and Task, indicating that participants in the PROD-ONLY and PERC-PROD groups performed significantly worse on the sentence reading task (PROD-ONLY: $p = 0.005$; PERC-PROD: $p < 0.001$) and picture description task (PROD-ONLY: $p = 0.002$; PERC-PROD: $p < 0.001$) compared to the word-list reading task, relative to the PERC-ONLY group. For participants in the MIX group, a significant decrease was observed only in the picture description task ($p = 0.002$).

Task was not a significant predictor, indicating that participants' performances in the PERC-ONLY group did not significantly differ between the word-list reading task and the other two tasks at pre-test. However, Group was a significant predictor (PERC-PROD: $p = 0.036$), suggesting that the PERC-PROD group's production was significantly more accurate compared to the PERC-ONLY group. Time did not significantly predict performance, indicating that participants in the PERC-ONLY group did not significantly improve their word-list reading scores from pre-test to post-test.

Pairwise comparisons were performed to examine whether there were any significant differences between groups for each task at pre-test, and the results showed no significant differences before training. Next, pairwise comparisons were conducted to examine whether there were any significant differences in goodness rating scores between speaking tasks within groups at pre-test (Table 5.13). The results showed no significant differences between tasks in the PERC-ONLY group before training. In the MIX group, the goodness rating scores for the word-list reading task were significantly higher than those for the picture description task, and the difference between the word-list reading and sentence reading tasks were approaching significance. In the PROD-ONLY and PERC-PROD groups, participants' goodness rating scores for the word-list reading task were significantly higher than those for the sentence reading and picture description tasks.

Table 5.13 Experiment 3: Summary of Pairwise Comparisons of Tasks within Groups at Pre-test for the Goodness Rating Score

<i>Groups</i>	<i>Tasks</i>		<i>Estimate</i>	<i>SE</i>	<i>z</i>	<i>p</i>
PERC-ONLY	Word	Sentence	-0.059	0.215	-0.272	0.960
		Picture	-0.163	0.217	-0.747	0.735
	Sentence	Picture	-0.104	0.216	-0.483	0.880
MIX	Word	Sentence	0.451	0.201	2.248	0.063
		Picture	0.501	0.202	2.475	0.036
	Sentence	Picture	0.049	0.203	0.244	0.968
PROD-ONLY	Word	Sentence	0.685	0.209	3.279	0.003
		Picture	0.656	0.207	3.171	0.004
	Sentence	Picture	-0.029	0.208	-0.138	0.990
PERC-PROD	Word	Sentence	0.927	0.212	4.365	<.0001
		Picture	0.804	0.212	3.787	0.000
	Sentence	Picture	-0.123	0.213	-0.579	0.831

Table 5.14 presents the summary of the improvement in goodness rating scores for each task and group. The results revealed that participants in the PERC-ONLY group did not significantly improve in any of the speaking tasks. In the MIX group, participants showed a significant improvement only in the word-list reading task. In contrast, participants in the PROD-ONLY and PERC-PROD groups significantly improved in the word-list reading and sentence reading tasks after training.

Table 5.14 Experiment 3: Summary of Pre-test and Post-test Pairwise Comparisons across Groups and Speaking Tasks for the Goodness Rating Score

<i>Groups</i>	<i>Tasks</i>	<i>Estimate</i>	<i>SE</i>	<i>z</i>	<i>p</i>
PERC-ONLY	Word	-0.293	0.191	-1.534	0.125
	Sentence	-0.027	0.190	-0.140	0.889
	Picture	0.102	0.192	0.531	0.596
MIX	Word	-0.402	0.180	-2.229	0.026
	Sentence	-0.136	0.181	-0.752	0.452
	Picture	-0.007	0.182	-0.041	0.967
PROD-ONLY	Word	-0.730	0.186	-3.937	0.000
	Sentence	-0.464	0.186	-2.495	0.013
	Picture	-0.336	0.185	-1.814	0.070
PERC-PROD	Word	-0.663	0.189	-3.516	0.000
	Sentence	-0.397	0.189	-2.101	0.036
	Picture	-0.268	0.189	-1.420	0.156

5.3.6 Summary of results on production learning

Table 5.15 summarises the p values from the pairwise comparisons in both identification rating and goodness rating scores. The results revealed that only the PERC-ONLY group did not significantly improve both identification rating and goodness rating scores. In other groups, participants significantly improved in the word-list reading task. Those in the PROD-ONLY group and PERC-PROD group significantly improved identification rating scores in the picture description task, but they improved goodness rating scores in the sentence reading task.

Table 5.15 Experiment 3: Summary of Pairwise Comparison Results (p values) for Production Learning

Identification Rating	PERC-ONLY	MIX	PROD-ONLY	PERC-PROD
Trained	×	0.015	0.036	0.022
Untrained	×	×	(0.069)	0.036
Word	×	0.004	0.001	0.000
Sentence	×	×	×	×
Picture	×	×	0.032	0.015
Goodness Rating				
Word	×	0.026	0.000	0.000
Sentence	×	×	0.013	0.036
Picture	×	×	×	×

Only the identification rating score results showed that the word type, whether trained or untrained, influenced participants' performances. Participants in the PERC-ONLY group did not significantly improve their identification rating scores, while the other three groups did for trained word stimuli, and the improvement generalised to untrained words only in the PERC-PROD group.

5.3.7 Relationship between PSTM and phonetic learning

First, an overview of participants' performances in the PSTM task is provided. Next, the results of a linear regression model examining the relationship between PSTM and discrimination accuracy prior to training are presented. Following this, the results of the second linear regression model with the gain score from the discrimination task as the dependent variable are discussed. Then, the results of the identification task data are provided in the same order. Finally, the results of the relationship between production data (i.e., identification rating and goodness rating scores) and the PSTM score are presented.

5.3.7.1 PSTM task score

Table 5.16 shows the mean scores, standard deviations, and minimum and maximum scores of the PSTM task, and Figure 5.20 depicts the PSTM score of each group. The mean score of the PERC-ONLY was 101.7, the MIX group 90.6, the PROD-ONLY group 90.1, and the PERC-PROD group 88.9, out of a maximum of 170 points. A linear regression model with the PSTM score as the dependent variable and Group as the independent variable revealed that there was no significant difference among the three groups.

Table 5.16 Experiment 3: Mean Scores, Standard Deviations, and Minimum and Maximum Scores of the PSTM Task

	Mean	<i>S.D.</i>	Min.	Max.
All groups	92.9	23.2	36.7	145.8
PERC-ONLY	101.7	22.8	64.2	145.8
MIX	90.6	19.3	54.2	126.0
PROD-ONLY	90.1	23.0	49.0	132.0
PERC-PROD	88.9	27.5	36.7	128.0

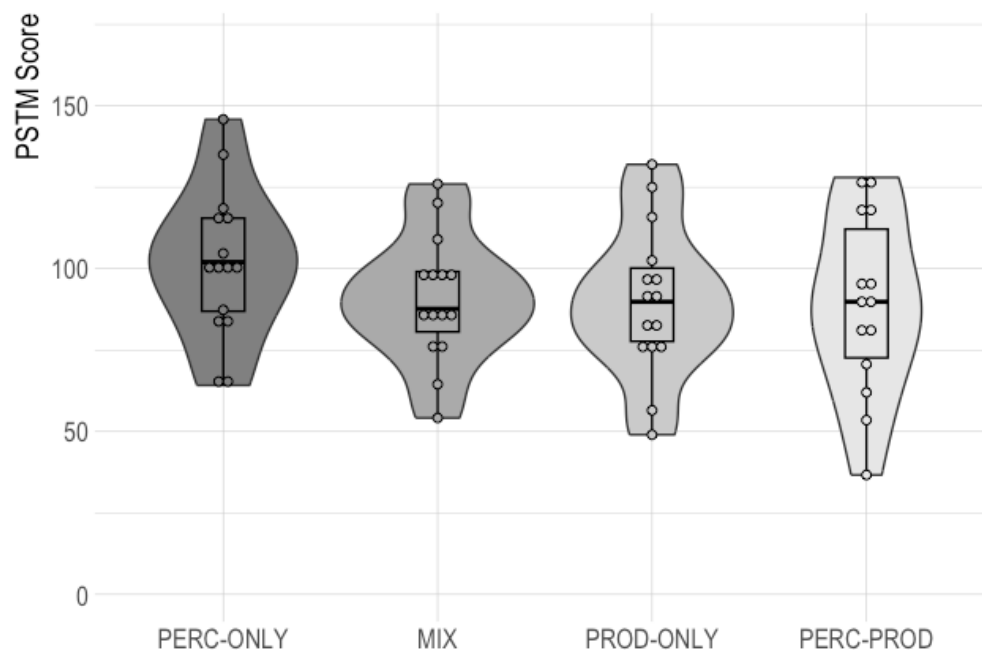


Figure 5.20 Experiment 3: PSTM Scores per Training Group

5.3.7.2 PSTM and discrimination scores

Figure 5.21 depicts the correlation between the PSTM score and the discrimination task pre-test score. A linear regression model with the pre-test score as the dependent variable and the PSTM score as the independent variable showed that PSTM was not a significant predictor of discrimination accuracy at pre-test ($\beta = 0.055$, $SE = 0.056$, $t = 0.981$, $p = 0.331$).

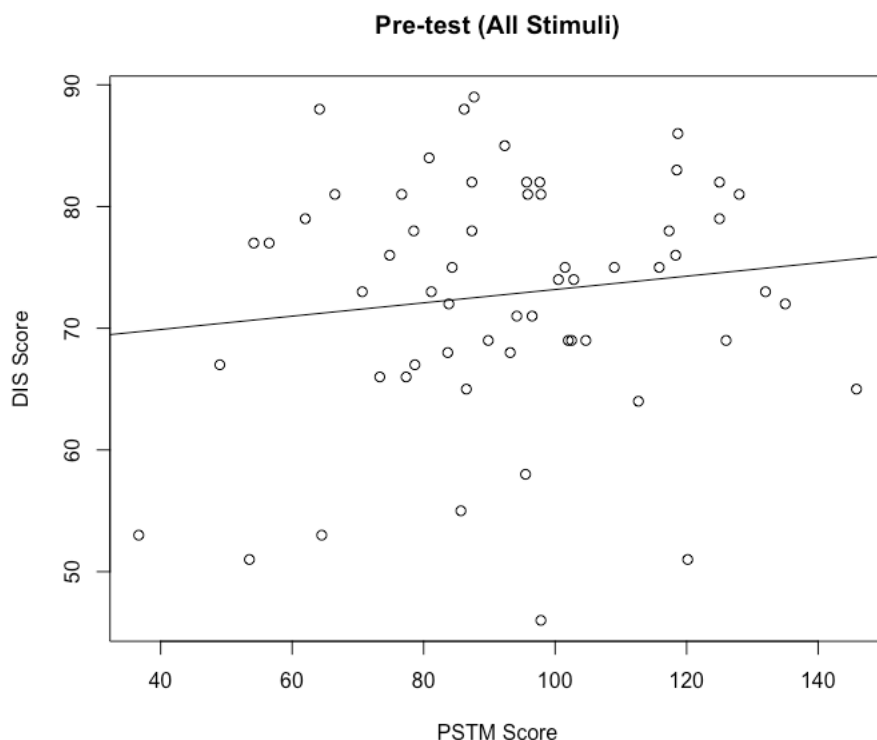


Figure 5.21 Experiment 3: Correlation between the PSTM Score and the Discrimination Task Pre-Test Score

Figure 5.22 illustrates the correlation between the PSTM score and the gain score of the discrimination task. A linear regression model revealed that PSTM was not a significant predictor ($\beta = 0.016$, $SE = 0.059$, $t = 0.271$, $p = 0.787$), suggesting that one's PSTM capacity does not predict the degree of learning in discrimination accuracy. Although one participant showed a much larger score increase than the others (outlier on the top left corner of the Figure), PSTM still did not emerge as a significant predictor when this participant was excluded ($\beta = 0.070$, $SE = 0.0499$, $t = 1.398$, $p = 0.168$).

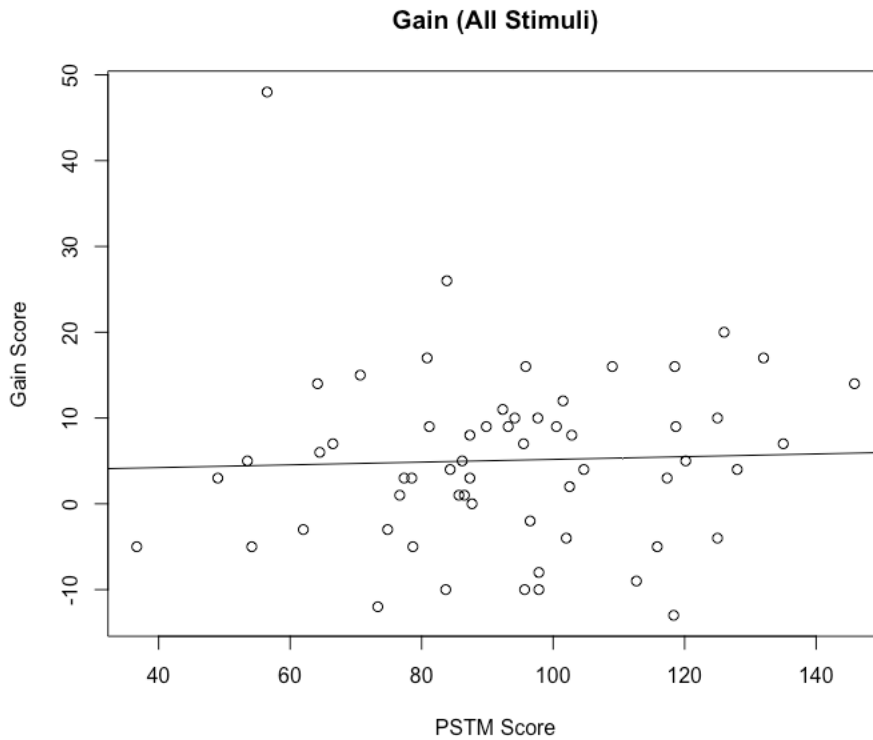


Figure 5.22 Experiment 3: Correlation between the PSTM Score and the Gain Score of the Discrimination Task

5.3.7.3 PSTM and identification scores

Figure 5.23 depicts the correlation between the PSTM score and the identification task pre-test score. A linear regression model revealed that PSTM was not a predictor of identification accuracy before training ($\beta = 0.061$, $SE = 0.059$, $t = 1.034$, $p = 0.306$).

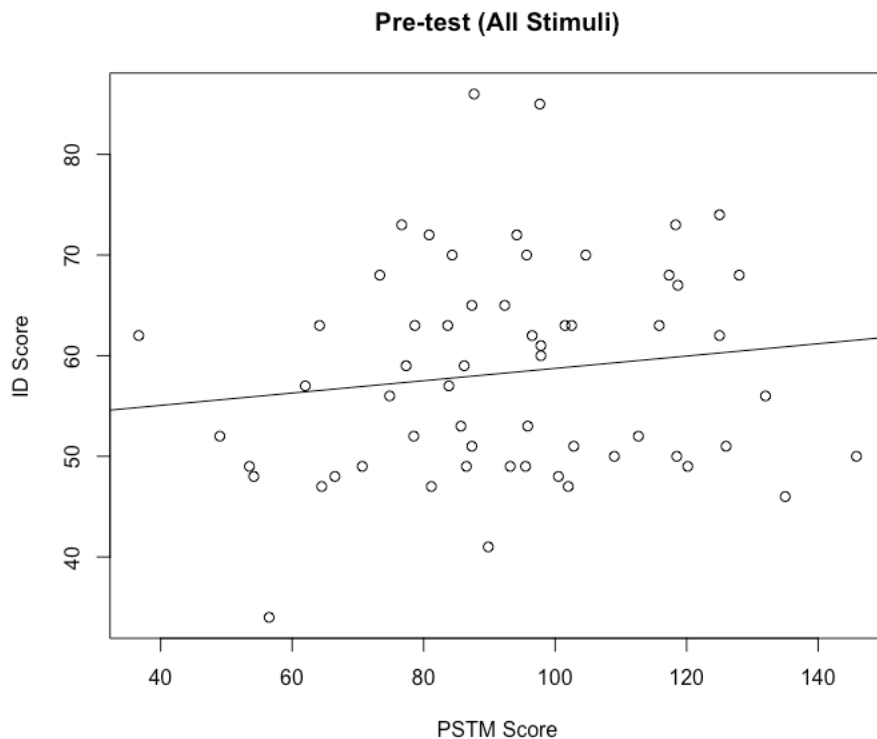


Figure 5.23 Experiment 3: Correlation between the PSTM Score and the Pre-Test Score of the Identification Task

Next, Table 5.17 presents the model output of a linear regression model, with the gain score in the identification task as the dependent variable and the PSTM score, Pre-test score, and Group as predictors. The model revealed a significant three-way interaction, suggesting that the relationship between PSTM and the gain from training varied depending on both the pre-test score and the training type. This variation was particularly notable in the PROD-ONLY group compared to the PERC-ONLY group. However, no significant interaction was found in the MIX and PERC-PROD groups.

Table 5.17 Experiment 3: Model Output of a Linear Regression Model with the PSTM score for the Identification Task

<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	-124.86	-294.31 – 44.58	0.145
PSTM	1.37	-0.22 – 2.97	0.090
Pre-test	2.53	-0.63 – 5.68	0.113
group [MIX]	136.04	-172.92 – 445.01	0.379
group [PROD-ONLY]	287.56 **	92.76 – 482.36	0.005
group [PERC-PROD]	245.34	-29.12 – 519.79	0.078
PSTM × Pre-test	-0.03	-0.06 – 0.00	0.078
PSTM × group [MIX]	-1.31	-4.52 – 1.90	0.416
PSTM × group [PROD-ONLY]	-2.85 **	-4.83 – -0.87	0.006
PSTM × group [PERC-PROD]	-2.71	-5.82 – 0.39	0.085
Pre-test × group [MIX]	-3.00	-9.00 – 2.99	0.318
Pre-test × group [PROD-ONLY]	-5.23 **	-8.80 – -1.67	0.005
Pre-test × group [PERC-PROD]	-4.54	-9.29 – 0.21	0.060
(PSTM × Pre-test) × group [MIX]	0.03	-0.03 – 0.09	0.340
(PSTM × Pre-test) × group [PROD-ONLY]	0.05 **	0.02 – 0.09	0.005
(PSTM × Pre-test) × group [PERC-PROD]	0.05	-0.00 – 0.10	0.062
Observations	59		
R ² / R ² adjusted	0.392 / 0.180		

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In the PERC-ONLY group, PSTM was not a significant predictor of the gain score. The two-way interaction between PSTM and Group was significant, suggesting that PSTM had a different effect on gain scores in the PROD-ONLY group compared to the PERC-ONLY group. The Pre-test × Group interaction was also significant, and it was a negative relationship. These findings suggest that the pre-test score significantly predicted gain score in the PROD-ONLY group, and participants with higher pre-test scores showed less improvement.

To investigate whether PSTM was a significant predictor for the PROD-ONLY group, the model was relevelled to make the PROD-ONLY group as the reference level. The relevelled model showed that PSTM and the interaction between PSTM and Pre-test were significant, with a negative relationship between PSTM and the gain scores. This suggests that PSTM capacity negatively influenced the degree of perceptual learning, but the influence varied depending on the pre-test score in the PROD-ONLY group.

Figure 5.24 visualises the interaction between PSTM, Pre-test score, and Group on the gain score. The plot depicts the relationship between the predicted gain scores and PSTM scores at three different pre-test scores (50, 60, and 70) for all groups. At a pre-test score of 50, a negative relationship between the PSTM and gain scores was observed for the MIX group, suggesting that participants with the larger PSTM capacity benefited more from MIX training. In contrast, the PROD-ONLY and PERC-PROD groups showed a negative relationship between the PSTM and gain scores, indicating that participants with larger PSTM capacity improved less than those with smaller PSTM capacity. The PERC-ONLY group exhibits a flat relationship, implying that PSTM had little effect on improvement.

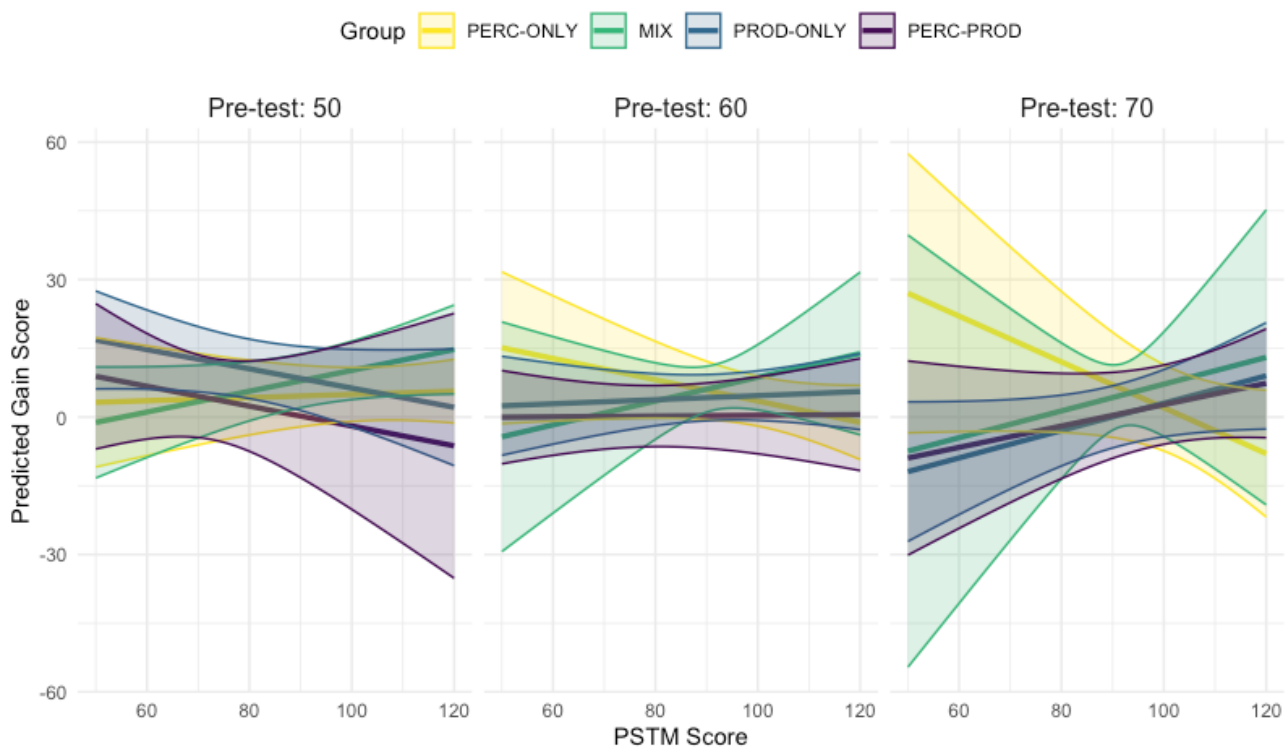


Figure 5.24 Experiment 3: Interaction between the PSTM and Pre-test Scores on the Gain Score in the Identification Task across Groups

At a pre-test score of 60, the MIX group continued to exhibit a positive relationship between the PSTM and gain scores. The PERC-ONLY group showed a negative relationship, similar to the pattern observed at the pre-test score of 60. In the PROD-ONLY and PERC-PROD groups, there was a flat relationship between the PSTM and gain scores.

At a pre-test score of 70, a relationship between the PSTM and gain scores was positive in the MIX, PROD-ONLY, PERC-PROD groups. In contrast, the PERC-ONLY group exhibited negative relationship between the PSTM and gain scores, indicating that higher PSTM scores lead to less improvements. In summary, these findings suggest that participants with larger PSTM capacity in the MIX group benefited more from training. However, in the PROD-ONLY and PERC-PROD groups, participants with larger PSTM capacity were negatively impacted at lower pre-test scores but positively impacted at higher pre-test scores. In contrast, participants in the PERC-ONLY group with larger PSTM capacity were negatively impacted on their perceptual learning, particularly at higher pre-test scores.

5.3.7.4 PSTM and identification rating scores

Next, the associations between PSTM and the production scores were analysed. Figure 5.25 presents the correlation between the PSTM and identification rating scores. A linear regression model showed that PSTM was not a significant predictor of the identification rating scores before training ($\beta = 0.001$, $SE = 0.001$, $t = 0.895$, $p = 0.374$)

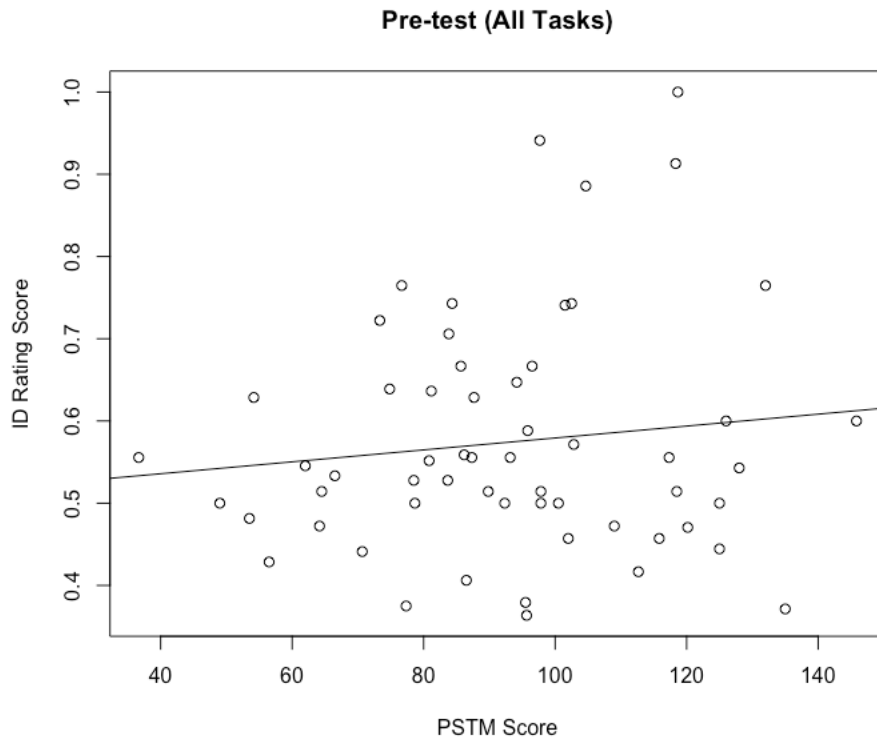


Figure 5.25 Experiment 3: Correlation between the PSTM Score and the Pre-Test Score of the Identification Rating Scores on Production

Figure 5.26 illustrates the correlation between the PSTM and gain scores of the identification rating on production. A linear regression model with the gain score as the dependent variable and PSTM score, Pre-test score, and Group as independent variables revealed that the three-way interaction and two-way interactions were not significant (Table 5.18). As a main effect, the PSTM score was a significant predictor across all groups ($p = 0.011$), suggesting that participants with higher PSTM scores showed greater improvements in the gain score compared to those with lower PSTM scores. Group and Pre-test score were also significant predictors, indicating that participants with lower pre-test scores improved more than those with higher pre-test scores, and the gain in the PERC-ONLY group was significantly smaller compared to that in the PROD-ONLY and PERC-PROD groups.

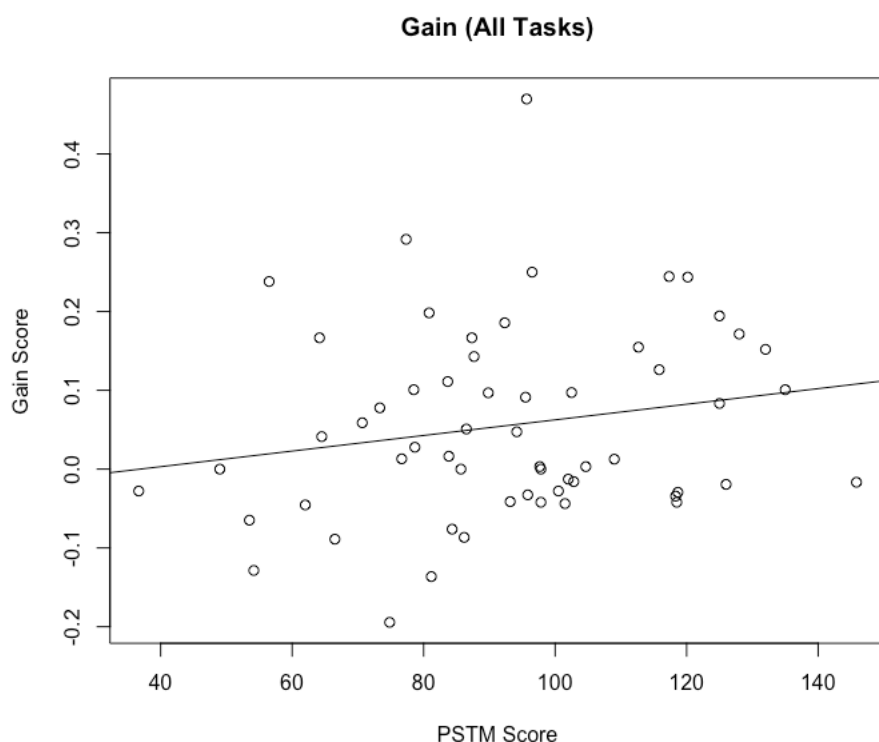


Figure 5.26 Experiment 3: Correlation between the PSTM Score and Gain Score of the Identification Rating Scores on Production

Table 5.18 Experiment 3: Model Output of a Linear Regression Model with the PSTM score for the Identification Rating Score with All Tasks

<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.03	-0.14 – 0.20	0.704
PSTM	0.00 *	0.00 – 0.00	0.011
Pre-test	-0.35 ***	-0.55 – -0.15	0.001
group [MIX]	0.07	-0.01 – 0.15	0.102
group [PERC-PROD]	0.11 *	0.03 – 0.19	0.010
group [PROD-ONLY]	0.10 *	0.02 – 0.18	0.020
Observations	58		
R ² / R ² adjusted	0.306 / 0.239		

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

5.3.7.5 PSTM and goodness rating scores

Figure 5.27 presents the correlation between the PSTM and goodness rating score at pre-test. In line with the results of the identification rating scores, a linear regression model revealed that PSTM did not significantly predict goodness rating score before training ($\beta = 0.008$, $SE = 0.004$, $t = 1.93$, $p = 0.059$), although the p value was approaching significance.

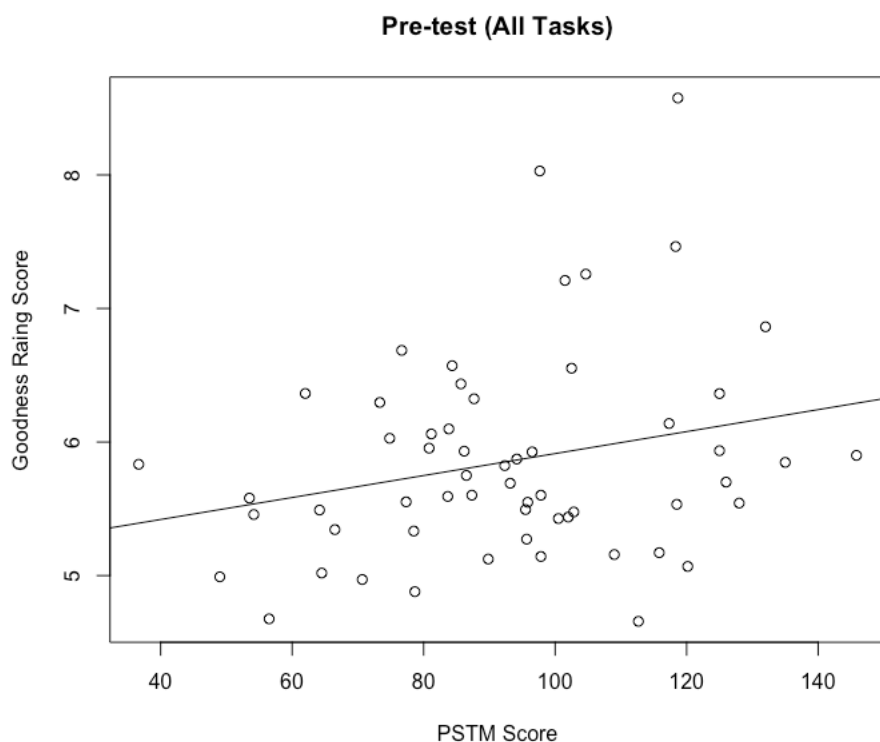


Figure 5.27 Experiment 3: Correlation between the PSTM Score and the Pre-test Score of the Goodness Rating Scores on Production

Figure 5.28 depicts the correlation between the PSTM score and gain score of goodness rating on production. A linear regression model showed that PSTM was not a significant predictor of gain score ($\beta = 0.002$, $SE = 0.003$, $t = 0.718$, $p = 0.476$).

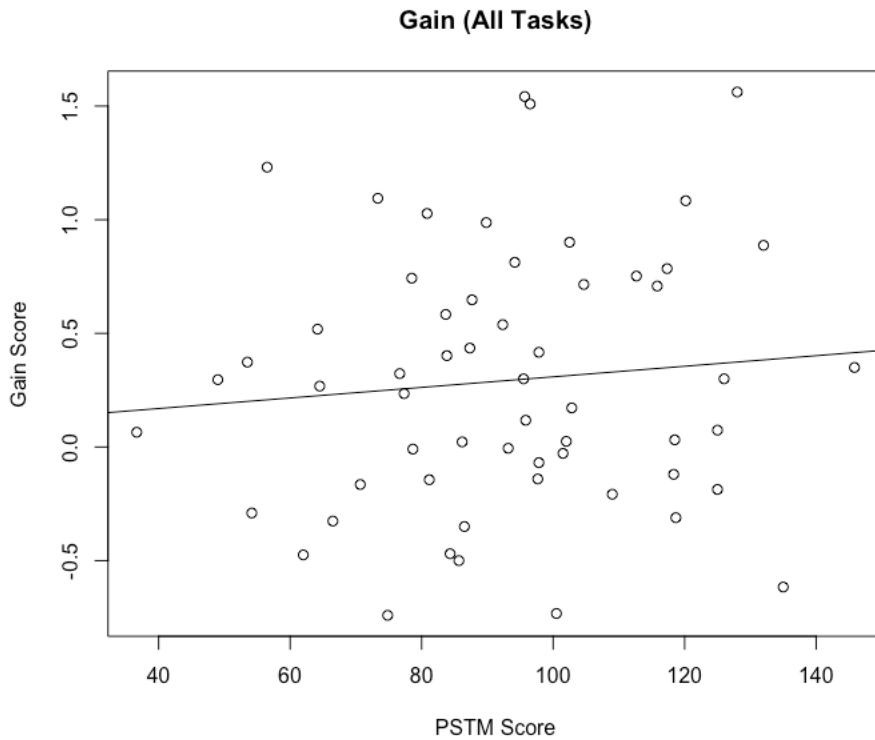


Figure 5.28 Experiment 3: Correlation between the PSTM Score and the Gain Score of the Goodness Rating Scores on Production

5.3.8 Summary of results on PSTM

Participants' PSTM scores did not significantly predict either discrimination or identification accuracy before training. However, for identification accuracy only, a significant three-way interaction was found between PSTM score, Pre-test score, and Group, suggesting different trends across groups: in the MIX group, PSTM capacity positively impacted perceptual learning, while it had a negative impact at lower pre-test scores in the PERC-ONLY group. In contrast, in the PROD-ONLY and PERC-PROD groups, PSTM capacity was negatively associated with gains in discrimination accuracy at lower pre-test scores and it shifted to a positive association as the pre-test scores increased. In terms of production accuracy, none of the interactions were significant, indicating that all groups showed similar relationships between PSTM and production learning. Regarding main effects, the PSTM score was not a significant predictor of the gain score before training. As for the degree of learning, the PSTM score significantly predicted gain score only in identification rating across all groups.

5.3.9 Learners' Perceptions of Phonetic Training

Among those who completed all training sessions, 43 participants (12 in the PERC-ONLY group, 10 in the MIX group, 11 in the PROD-ONLY group, and 10 in the PERC-PROD group) filled out an exit questionnaire regarding their perceptions of the phonetic training that participants received.

Table 5.19 summarises participants' responses to the exit questionnaires. With the scale ranging from 1 to 10, responses between 1 and 3 were considered 'ineffective,' scores between 4 and 6 were regarded as 'neutral,' while scores from 6 to 10 were deemed 'effective.' First, the questionnaire asked participants whether they wanted to improve their English skills and pronunciation. Most participants reported their desire to improve their English skills and pronunciation (Figure 5.29). Four participants selected 5, just below the midpoint, for English skills and pronunciation, respectively. One participant from the PERC-PROD group consistently answered 1, the lowest for both questions. Apart from this participant, individuals across all groups exhibited similar levels of desire to improve their English skills and pronunciation.

Table 5.19 Experiment 3: Summary of the results of the Exit Questionnaire

Group	Question	1	2	3	4	5	6	7	8	9	10
All groups	I want to improve my English skills.	1	0	0	0	1	0	7	11	5	18
	I want to improve my English pronunciation.	1	0	0	0	3	2	7	8	4	18
PERC-ONLY	I want to improve my English skills.	0	0	0	0	0	0	2	3	0	7
	I want to improve my English pronunciation.	0	0	0	0	0	0	2	2	1	7
	The training task was effective in improving my English pronunciation.	0	0	1	2	1	3	2	1	1	1
	I want to continue using the training task to improve my English pronunciation.	0	0	1	2	2	1	1	2	1	2
MIX	I want to improve my English skills.	0	0	0	0	1	0	2	2	1	4
	I want to improve my English pronunciation.	0	0	0	0	1	0	4	1	0	4
	The training task was effective in improving my English pronunciation.	1	0	0	0	2	2	3	2	0	0
	I want to continue using the training task to improve my English pronunciation.	1	0	4	0	1	2	1	1	0	0
PROD-ONLY	I want to improve my English skills.	0	0	0	0	0	0	0	5	2	4
	I want to improve my English pronunciation.	0	0	0	0	1	1	1	3	1	4
	The training task was effective in improving my English pronunciation.	0	0	2	0	1	1	1	2	3	1
	I want to continue using the training task to improve my English pronunciation.	1	0	2	1	0	1	3	0	2	1
PERC-PROD	I want to improve my English skills.	1	0	0	0	0	0	3	1	2	3
	I want to improve my English pronunciation.	1	0	0	0	1	1	0	2	2	3
	The training task was effective in improving my English pronunciation.	0	0	1	0	1	3	2	3	0	0
	I want to continue using the training task to improve my English pronunciation.	0	0	2	0	3	0	3	2	0	0
	The training task was effective in improving my English pronunciation. (PERC-ONLY training)	0	0	0	0	0	2	4	2	2	0
	I want to continue using the training task to improve my English pronunciation. (PROD-ONLY training)	0	0	1	1	2	1	2	2	1	0

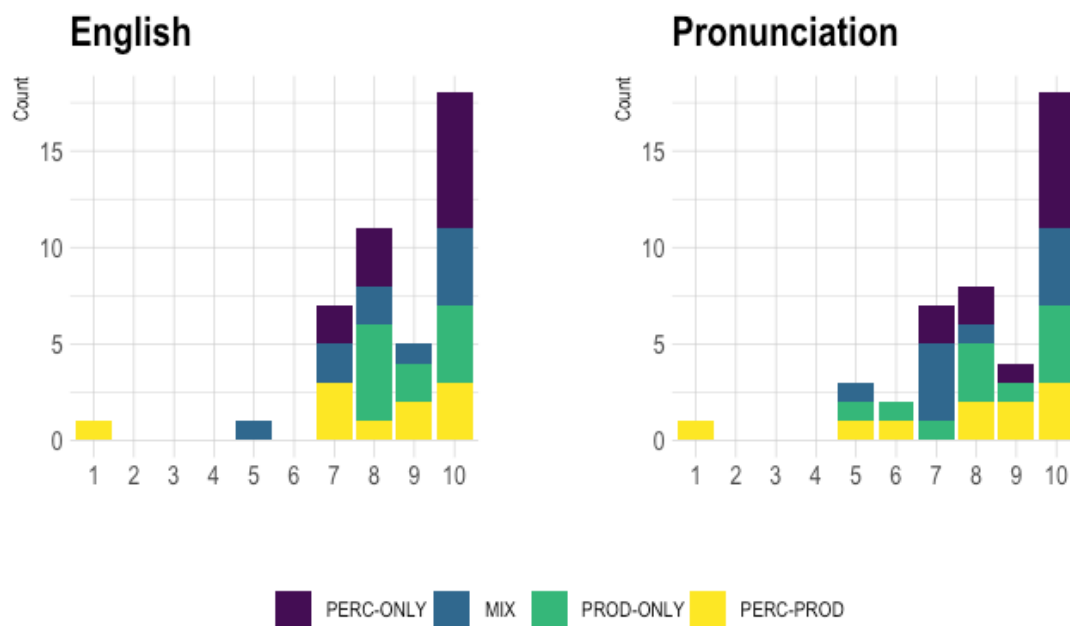


Figure 5.29 Experiment 3: Participants' Desires to Improve Their English Skills and Pronunciation

5.3.9.1 PERC-ONLY training

Next, participants' perceptions of each training task were analysed by group. Table 5.30 depicts participants' perceptions of the PERC-ONLY training task. In the PERC-ONLY group, 25% of participants ($n = 3$) rated the training task as effective, 66.7% rated it as neutral ($n = 8$), and 8.3% ($n = 1$) rated it as ineffective.

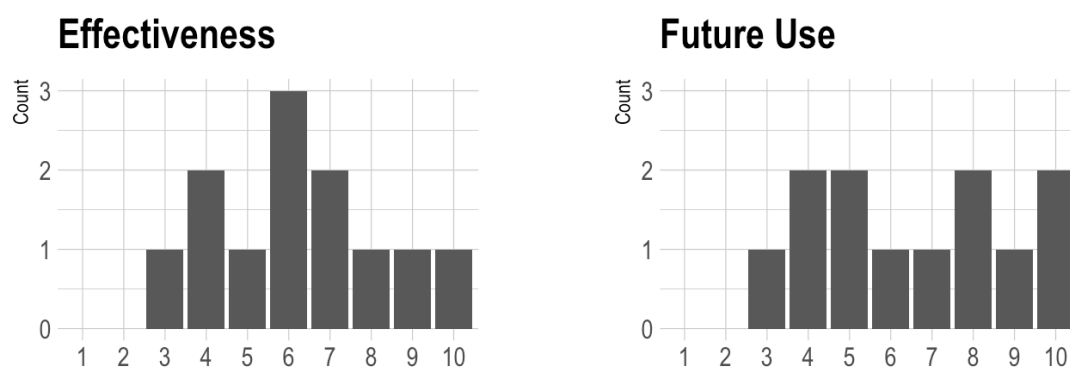


Figure 5.30 Experiment 3: Participants' Perceptions towards PERC-ONLY Training

Participants who rated the task as effective noted that exposure to the target sounds helped them hear the difference between the two sounds. Participants answered in Japanese and their responses were translated into English by the researcher.

たくさん聞いてなんとなく l と r の違いがわかった。
(I listened to a lot and understood the difference between l and r to some extent.)

2つの単語を聞き分けるためにアクセントや発音を意識して聞いていたから。
(Because I was consciously listening to the stress and pronunciation to distinguish two words.)

続けていくうちに発音の違いがわかってきたから。
(Because I started to hear the pronunciation differences as I continued.)

One participant mentioned that they were unable to distinguish between the target sounds by the end of the training. However, most negative responses may have stemmed from a misunderstanding of the question's intent. Some participants noted that the task would have been more helpful if they could know whether their answers were correct, although immediate feedback was provided during the training. This misunderstanding might have arisen because this questionnaire was completed after the post-test, where participants did not receive immediate feedback. In addition, since the post-test was conducted in an English class, participants might have assumed that post-test perceptual tasks were part of the training.

l と r の違いが分からないまま終わってしまった。
(It ended without me knowing the difference between l and r.)

自分の思っているものが正解なのかがはっきりしなかったから。
(It was not clear if what I was thinking [my answer] was correct.)

One participant suggested that it would have been better if they could hear both sounds again after answering the question in each trial.

どちらかの発音だけじゃなくて、回答した後に2つとも発音を聞いて違いがわかるようになっていたら良かったと思う。
(It would have been good if I could listen to both two pronunciations to hear the difference after answering, instead of just one of them.)

Participants' responses regarding their desire to continue using the training task in the future were consistent with their beliefs about its effectiveness. A total of 58.3% of the participants ($n = 7$) indicated that they would continue using the PERC-ONLY training task. They reported that they found the training task effective and observed improvements in their perceptual accuracy during training.

効果的だと思ったから。(Because I thought it was effective.)

似た発音を意識して聴き分けられるようになるから。
(Because it helps me distinguish between similar pronunciations by focusing on them.)

途中から聴き分けやすくなったから。
(Because I started to find it easy to hear the difference during the course [of training].)

はじめより、少しながら聞き分けられるようになった単語が増えたので、力がつく
と思ったから。
(Because I thought I could distinguish more words than I could at first and improve my
skills.)

少しは分かるようになったと思ったから。
(Because I thought I became able to understand a little.)

さらに続けていくことで発音の上達につながると思うから。
(Because I think it will help me to improve my pronunciation if I continue further.)

A total of 41.7% of the participants ($n = 5$) rated the task below the midpoint. Among the negative responses, one participant criticised that PERC-ONLY training as ineffective for pronunciation learning. Another participant expressed a preference for training the involves sound production rather than a listening-focused approach.

自分で声に出してみるトレーニングもやってみたいからです。
(Because I want to try training to produce sounds aloud.)

発音の練習にはあまりならないと思ったから。
(Because I did not think it would be much of a pronunciation practice.)

One participant mentioned that they did not see an opportunity to engage in PERC-ONLY training, implying that they were unsure where to access this type of phonetic training outside of class.

する機会はないと思うから。(Because I do not think there is an opportunity.)

At the end of the survey, one participant expressed that they enjoyed the training, and another found the pronunciation training beneficial. Participants also reported that they had not encountered this type of training before. One suggested that a longer training period would be more effective, while only one participant found the training challenging.

楽しかった。(It was fun.)

週に1回発音のトレーニングができてよかった。

(It was good to receive pronunciation training once a week.)

高校でもしないようなものだったので新鮮でした。
(It was refreshing to me because it was something [I] had not done in high school.)

ずっと続けていると分かるようになっていった。
(I thought I would know the difference if I continued doing it for a long time.)

徐々に慣れていったが、難しかったです。
(I gradually got used to it, but it was difficult.)

5.3.9.2 MIX training

Figure 5.31 presents participants' perceptions of the MIX training task. Most participants ($n = 9$) in the MIX group rated the effectiveness of the MIX training between 5 and 8, while one participant rated it as ineffective.

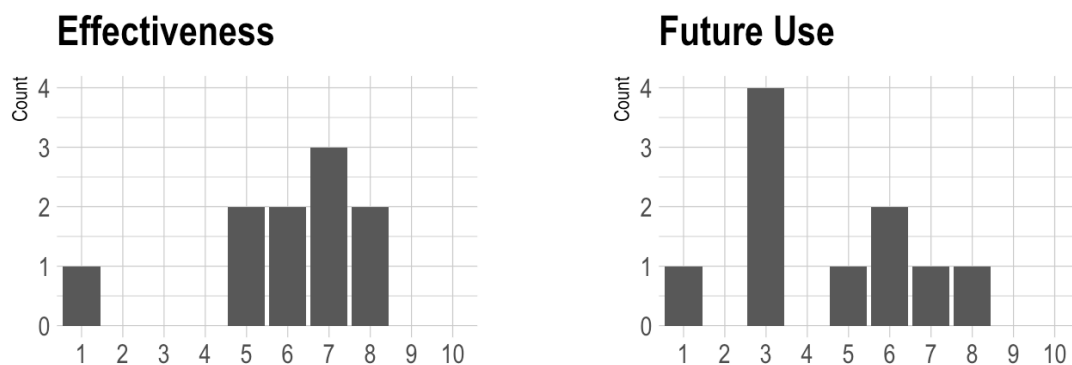


Figure 5.31 Experiment 3: Participants' Perceptions towards MIX Training

Like the positive feedback for the PERC-ONLY training, participants mentioned that MIX training facilitated their perceptual learning, and they noticed an improvement in their perceptual accuracy as the training progressed.

L と R の発音が分かりやすく聞こえたから。
(Because I heard the pronunciation of L and R clearly.)

r と l の発音の違いを特に意識することができたから。
(Because I was able to become especially aware of the difference between the pronunciation of r and l.)

耳が慣れてくると判別しやすくなっていったから。
(Because it became easier to distinguish as my ears became accustomed to it.)

A couple of participants commented on its effectiveness in improving their pronunciation.

音声を聞いてすぐに発音するため、発音が身につけやすい
(It is easy to learn pronunciation as I hear audio and pronounce it immediately.)

L と R の発音を中心に、まねるようにして発音を学習できたと思う。
(I think I learned pronunciation by imitating, focusing on L and R pronunciation.)

Some participants commented negatively on the effectiveness of the training due to the speakers' 'quirky voice', although one participant mentioned that it was easy to hear the difference between the target sounds. Another participant criticised the lack of novelty in the training. In addition, one participant reported that the training proceeded without them fully understanding the difference between the target sounds.

癖があってわかりにくい音声の人もいて、感覚が掴みにくいように感じた
(Some people had a difficult-to-understand voice with a quirky sound, and I found it difficult to get a sense of what was going on.)

発音の区別は中学校で習うので。
(Because pronunciation differences are learned in junior high school.)

違いがわからないままトレーニングをしていて、よくわからなかったから。
(Because I did not understand well as training proceeded without knowing the differences)

Regarding whether participants would continue doing MIX training, half of them ($n = 5$) rated its future use between 5 and 8. Four participants rated it 3, and one rated it 1. Some participants indicated that they would continue if they had an opportunity and time. Others mentioned their desire to practice regularly, noting that more extended training could enhance learning.

機会があるならやりたいと思うため。
(Because I want to do it if there is an opportunity.)

効果はあったから。時間があればやり続け[て]みたい。
(Because it was effective. I want to continue if I have time.)

r の発音は日本語では使わないので、定期的に練習したいから。
(Because I want to practice regularly as I do not use the pronunciation of r in Japanese)

聞き続けていくとわかるようになれそうだから。
(Because I feel that I will be able to understand it if I keep listening.)

Other participants ($n = 2$) expressed negative opinions on the future use of MIX training, mentioning that the training task was difficult and that they did not feel they improved much. One

participant suggested that the difficulty stemmed from the ambiguous pronunciation of trained stimuli. A different participant expressed a preference for pronunciation practice with contextual sentences, while another pointed out the challenge of finding a quiet place for training.

難しく、分かりにくかったから。
(Because it was difficult and hard to understand.)

あまり自分ではできるようになっている実感がなかった。
(I did not feel that much that I was able to do better.)

問題出題者の発音がかなり曖昧に思えるので。
(Because I think the pronunciation of the question-submitter sounds quite ambiguous.)

単語ごとで練習するよりも文の流れで発音を練習するほうが自分に合っていると考えているため
(Because I believe that practising pronunciation in the flow of a sentence is better suited for me than practising word by word.)

静かな環境でやらないといけないの[で、]学習できる時間が限られるから。
(Because I have to do it in a quiet environment, time to study is limited.)

Regarding general comments, one participant noted the effectiveness of the high variability of trained stimuli, although they were not informed about the underlying theories of the training. A couple of participants said that they wished they had summative or immediate feedback. Although they did not receive summative feedback, immediate feedback was provided during training. As mentioned in the PERC-ONLY training section, it is likely that these participants were referring to the perceptual test rather than training itself when filling out the survey. It is important to note, however, that other participants clearly understood that this survey asked about training, as indicated by one participant's comment on the high variability of stimuli's voices.

同じ単語でもしっかりと色々な人の音声を用意されていたので、単語の聞き取り練習にはとても良いと感じました。
(I found it to be a great way to practice listening to words, as it provided a solid variety of people's voices for the same word.)

自分のリスニング力がどれくらい上達したかがあまり分からなかったですが、英語に触れられる良い機会になったと思います。
(I was not really sure how much my listening skills had improved, but I think it was a good opportunity to be exposed to English.)

一回だけでもいいので正解が分かるとうれしいです
(I would be happy to know the correct answer even once.)

5.3.9.3 PROD-ONLY training

Figure 5.32 depicts participants' perceptions of the PROD-ONLY training task. Results revealed that 54.5% of participants rated the effectiveness of PROD-ONLY training using ASR between 8 and 10 ($n = 6$). Meanwhile, 27.3% of participants rated it as neutral ($n = 3$), and another 18.2% rated it as 3 ($n = 2$).

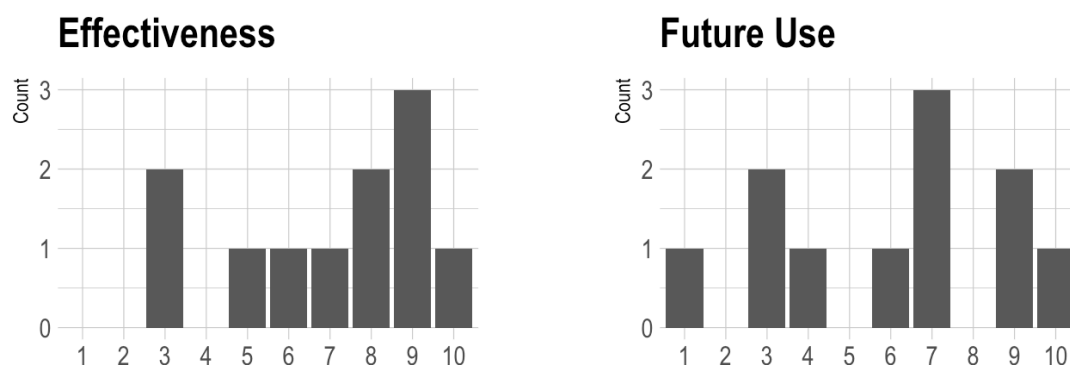


Figure 5.32 Experiment 3: Participants' Perceptions towards PROD-ONLY Training

The majority of participants ($n = 6$) made positive comments about the effectiveness of the PROD-ONLY training task using ASR. They reported focusing on the pronunciation of the target phonemes during training. Some participants ($n = 4$) noted that the ASR system's outputs facilitated pronunciation learning, and they observed improvement in their pronunciation by checking the ASR's feedback. One participant mentioned that they reflected on how to improve their pronunciation during training, suggesting that the ASR-based training stimulated their metacognitive skills in pronunciation learning.

r や l の発音を気にしなければならなかったから。
(Because I had to focus on the pronunciation of r and l.)

単語を見て音読することで、発音を効率的に意識できたから。
(Because I was able to become more aware of pronunciation efficiently by looking at the words and reading them aloud.)

回数を重ねるにつれて、提示された単語と聞き取られた単語が一致しやすくなったから。
(Because as the number of times increased, presented words and perceived words became more likely to match.)

自分の発音が正しいかどうかわかるから
(Because I know whether my pronunciation is right)

たくさん発音をして次第に機械が聞き取れるようになったから。
(Because by pronouncing a lot, the machine gradually became able to hear them.)

何回もトレーニングを繰り返すことでどうすれば発音が良くなるのかを考えながら発音をすることができたから。
(Because I was able to repeat the training many times and think about how to improve my pronunciation.)

Participants who rated the training as less effectiveness reported that their pronunciation did not change significantly. One participant mentioned that they were pronouncing words without knowing whether their pronunciation was target-like. Although they were instructed to check the ASR's feedback during training, it seems that some participants either did not frequently check the feedback or did not fully trust its accuracy.

最初と最後であまり変わらなかったから。
(Because it did not change much from the beginning to the end.)

いっぱい発音したので、身についたと思うが、正解がよく分からないまま発音していたので、すごく効果的だったとはあまり思えない。
(I think I learned them because I pronounced them a lot, but I do not really think it was very effective because I was pronouncing them without really knowing the correct answer.)

Regarding their plan to use the PROD-ONLY training with ASR in the future, 27.3% of the participants indicated that they would continue the training task ($n = 3$), and 36.4% selected a rating of 6 or 7, slightly above the midpoint ($n = 4$). The remaining participants stated that they would not continue the training task ($n = 4$).

Three participants expressed a desire to continue PROD-ONLY training due to the effective training design, such as the ASR's feedback and the inclusion of stimuli containing /l/ and /r/. Others mentioned that the training helped them understand the pronunciation differences between the target phonemes and encouraged them to think about pronunciation when speaking. Additionally, one participant reported wanting to improve their pronunciation but felt they lacked opportunities to do so.

自分が話した英語がどのように認識されるかわかるから
(Because I know how my spoken English is perceived.)

特に L と R の発音の違いなど、意識すべき単語がまとめられていて良いと思ったから。
(Because I thought it was a good compilation of words to be aware of, especially the difference in pronunciation between L and R.)

R と L の発音の違いが段々とわかるようになったから。
(Because I gradually came to understand the difference between R and L pronunciations.)

話す時に考えて発音するきっかけになると思うから。
(Because I think it will give me a chance to think when I speak and pronounce it.)

発音を矯正したいから。
(Because I want to correct my pronunciation.)

発音する機会があまりないから。
(I do not have many opportunities to pronounce them.)

In their final comments, one participant noted that they found the ASR training effective and would recommend it to junior high and high school students. However, others expressed negative comments about the training, mentioning difficulties such as the pronunciation challenge, the training length, and the need to find a quiet place for practice.

英単語を正確に読むことは難しいと思った。
(I found it difficult to read English words accurately.)

少し量が多い。
(A little too much.)

静かな場所でやらなきゃいけないので、できる場所が限られてたところがちょっとめんどくさかった。
(I had to do it in a quiet place, so it was a bit of a hassle that there were only a few places where I could do it.)

5.3.9.4 PERC-PROD training

Figure 5.33 illustrates the perceptions of the participants in the PERC-PROD group towards PERC-ONLY training and PROD-ONLY training tasks. Thirty percent of the participants rated PERC-ONLY training as effective ($n = 3$), 60% rated it as neutral ($n = 6$), and 10% rated it as inefficient ($n = 1$).

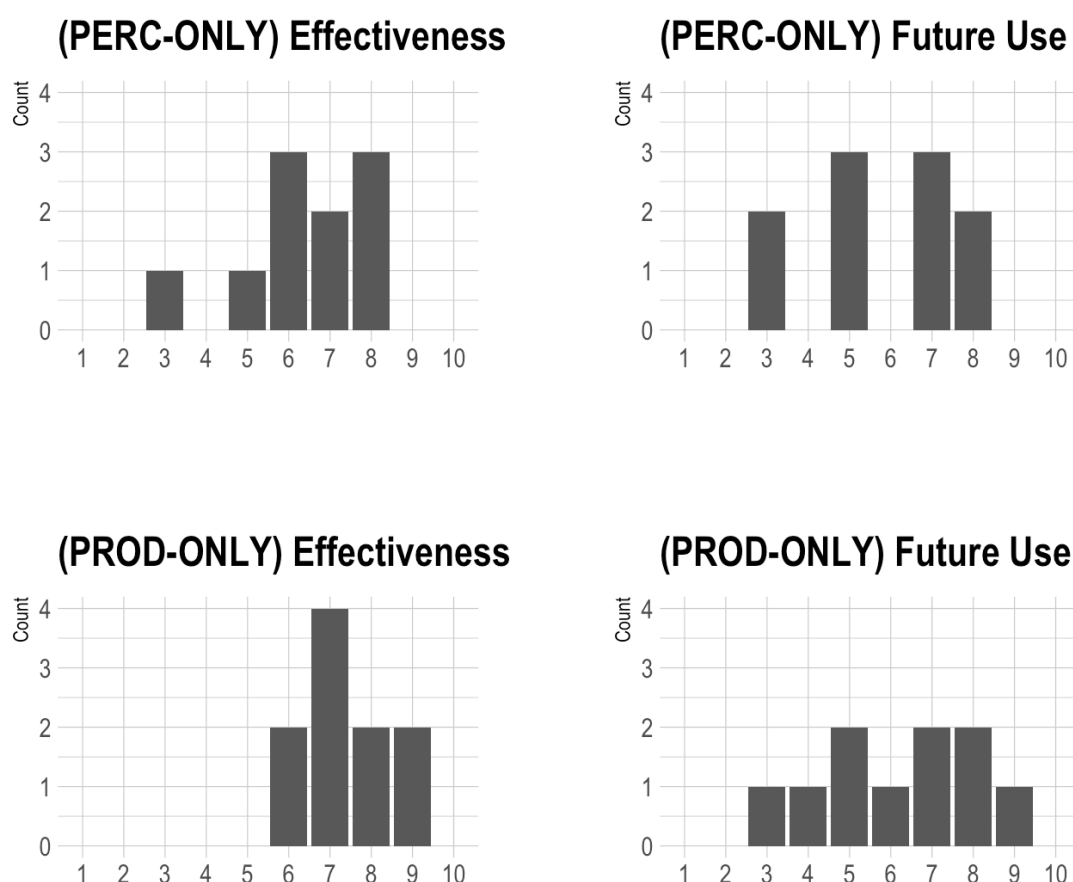


Figure 5.33 Experiment 3: Participants' Perceptions towards PERC-PROD Training

Participants reported that they found the training effective because it allowed them to focus on the differences between the target sounds. They also believed that listening to words in isolation was helpful. One participant mentioned that listening to L1 speakers would aid in pronunciation learning. Another noted that they began paying closer attention to pronunciation, suggesting that training facilitated the development of their metacognitive skills.

聞き分けるのには、とても効果的だったと思ったから。
(Because I thought it was very effective for distinguishing [sounds])

L と R の発音を文脈判断ではなく意識的に聞き取る必要があるため。
(Because it is necessary to consciously listen to the pronunciation of L and R instead of relying on contextual judgment.)

わかる問題も以前よりは多くなったように思うから。手軽にできたから。
(Because I think I understood more questions than before. Because it was easy to do.)

ネイティブの人の発音を聞くことは英語を聞き取るための近道だと考えるから。
(Because I believe that listening to native speakers' pronunciation is a shortcut to listening to English.)

注意深く音を気にするようになったから。
(Because I have become more attentive to sound.)

In contrast, some participants said the training was ineffective because they were unsure if their pronunciation was improving, indicating a need for summative feedback. Another participant pointed out a limitation of the training, the 2-alternative forced-choice design. It suggests that participants might have wanted a more varied approach, such as incorporating pronunciation practice or using a wider range of sounds. One participant expressed uncertainty about the effectiveness of the training as they proceeded without fully understanding the differences in pronunciation.

自分の英語の発音が伸びているかわからないから。
(Because I do not know if my English pronunciation is improving.)

二択を判定する要素しか鍛えられないため。
(Because only skills to judge from two choices can be trained.)

たまに違いがあまりわからないけれど選んでいたのも、効果的だったか分からない。
(I do not know if it was effective because sometimes I could not tell much difference, but I chose it.)

About their potential future use of the PERC-ONLY training task, participants in the PERC-PROD group mentioned that it would help them improve their perception and production. A couple of participants also reported that the PERC-ONLY training task was easy to follow.

音を聞き分ける練習になったから。
(Because it gave me practice in distinguishing sounds.)

聴き続けば、発音が良くなるようにも感じたから。
(Because I felt that my pronunciation would improve if I kept listening to it.)

トレーニング方法があまり時間がかからず簡単でやりやすかったから。
(The training method was simple and easy to follow without taking too much time.)

However, some participants ($n = 5$) did not agree that PERC-ONLY training would improve their pronunciation, with one mentioning they did not want to continue because perceiving the /l/ and /r/ distinction is not a necessary skill.

確かにネイティブの人の発音を聞くことは大切だが、それだけで発音が完璧になる訳では無いので、only トレーニングだけを行うのは物足りない気がする。(It is certainly important to listen to native speakers' pronunciation, but since it is not the only way to perfect pronunciation, I think it is not enough to do ONLY training solely.)

これ続けると自分の発音が伸びるかわからないから。
(I do not know whether my pronunciation will improve if I keep doing this.)

ぜったいいるスキルでもなさそう
(It does not seem to be a skill I need.)

As for general comments, one participant highlighted that the training made them realise their lack of perceptual skills, and others commented that they struggled with hearing the difference between the target phonemes and selecting the correct answer.

意外と聴き分ける力が足りないと実感できたのでいい経験だった。
(It was a good experience because it made me realise that my perceptual skills were surprisingly lacking.)

L と R の識別だが、どっちか分からない微妙な問題があったように感じた。単に僕が聞き取れてないのかもしれないですが。
(Regarding identifying L and R, I felt there was a confusing question where I could not tell which was which. I might not be simply able to hear well, though)

なかなか正解しない
(It is hard to get it right.)

With regard to the PROD-ONLY training task using ASR, 40% of the participants in the PERC-PROD group rated it as effective ($n = 4$), 60% rated it as neutral ($n = 6$), and none of them rated it negatively. Participants reported that the feedback from ASR on their pronunciation was helpful, suggesting that they checked ASR's outputs during training, which helped them learn. A couple of participants mentioned that repeated PROD-ONLY training would facilitate pronunciation learning.

自分の発音を客観的な視点でチェックできるため。
(Because I can check my pronunciation from an objective point of view.)

実際に自分の声が認識されることで、どんなときに正しい発音ができたか分かるから。
(Because I understand when my pronunciation was correct as my voice being actually recognised.)

自分の発音した言葉がどのように聞き取られているかが分かり、発音のトレーニングになるから。

(Because I can see how my pronunciation is being heard, and it is good for pronunciation improvement.)

耳で通して発音してみるのと声に出して発音してみるのでは全然違うことが分かった。これを繰り返し行えば発音の上達も見込めると思う。

(I found that pronouncing the words via ear is completely different from pronouncing them out loud. I think that if I do this repeatedly, I can expect to improve my pronunciation.)

トレーニングをしていくうちに発音のコツがわかったように思う時があったから。

(Because there were times when I thought I got the knack of pronunciation as I went through the training.)

One participant rated its effectiveness as low due to a lack of opportunities to listen to correct sounds. Another mentioned their frustration with ASR's 'misperception.'

正しい発音が聞けない為、どうやったら正しく認識される発音になるのか分からなかった

(Because I could not hear the correct pronunciation, I did not know how to make the pronunciation correctly recognised)

たまに google の AI が明らかな聴きミスをしていて気に触ることがあったが基本的にはいい経験だった。

(Sometimes the google AI made obvious listening mistakes that bothered me, but basically it was a good experience.)

Participants in the PERC-PROD group rated their intended future use of the PROD-ONLY training with ASR between 3 and 9. Three participants noted that they would continue using it due to its effectiveness and the enjoyable game-like nature of the training.

繰り返せば、正しい発音ができるようになるような気がするから。

(Because I feel that if I repeat it, I will be able to pronounce it correctly.)

ゲーム感覚で面白いため。

(Because it is interesting in a game-like sense.)

Although many participants found PROD-ONLY training effective, several ($n = 7$) expressed negative comments about its future use. They highlighted issues with ASR's misperceptions, which caused frustration during the training. A couple of participants also mentioned that the number of trials was too many.

明らかにそんなこと言ってないのに誤認識されることが多々あったので、なかなか合わないことでモチベーションが少し下がってしまったから。

(Because I was often misidentified even though I obviously did not say such a thing, and I lost a little motivation because my pronunciation did not match very well.)

自分の発音が良くないのだと思うが、聞きとられていないことが多くあり、何度もやり直しをしていたから。

(I think my pronunciation was not good, but there were many times that my pronunciation was not being heard, and I had to redo them many times.)

たまに私は絶対に言っていないと思う単語が認識されたりするので、発音が悪すぎるのかイヤホンや言い方が悪かったが分からなかったから。

(Because I did not know if I pronounced them too badly or if I was using bad earphones, when words I thought I never said were recognised sometimes.)

発音すべき量が多い

(There are many to pronounce)

As for general comments on PROD-ONLY training using ASR, one participant mentioned that they felt happy when their pronunciation was recognised correctly, while another reported that the ASR training helped them become aware of their pronunciation habits.

合わない時は凹んだが、ちゃんと発音してちゃんと認識された時は嬉しかった。

(When my pronunciation did not match, I felt down, but when I pronounced it correctly and it was recognised properly, I was happy.)

自分の発音の癖を知ることができたのでよかった。

(It was good to know about my pronunciation habits.)

Lastly, the exit questionnaire asked participants in the PERC-PROD group which training task they found more effective: PERC-ONLY or PROD-ONLY. A majority, 63.6% of participants ($n = 7$), chose PROD-ONLY training, while only one preferred PERC-ONLY training. Three participants stated that both training types were equally effective. Participants generally believed that pronunciation practice is more effective than perception practice, implying that perceptual training is ineffective for improving pronunciation. One participant commented that they believed they could hear the sounds if they could produce them, and another expressed hopes for further advancements in the ASR system.

自分で話す方が伸びる気がするから。

(Because I feel like I can improve more if I speak on my own.)

Only トレーニングはあまり発音を伸ばすのに使えんと思えなかった

(I did not think the Only training could be used to improve pronunciation very much.)

話すことができれば聞き取ることはできると考えるため。

(Because I think that if I can speak, I can hear.)

今後も発展が期待できるため

(Because we can expect further development in the future)

The participant who supported PERC-ONLY training over PROD-ONLY training noted that listening to the pronunciation differences taught them which parts of their pronunciation to focus on.

発音の差を聞けるので自分の発音で意識する部分が分かってよかった。

(It was good to hear the difference in pronunciation so I could understand what I need to be aware of in my pronunciation.)

A few participants ($n = 2$) described that both training types had their respective strengths.

One participant suggested that two sessions were too short for phonetic learning.

聞いて学ぶのと発音して学ぶのには、両者ともに良い点があると思ったから。

(I thought that learning by listening and learning by pronouncing have both good points.)

どちらのトレーニングも行ったが、それぞれ2回ずつだったのであまり効果を得られた実感はしなかったから。

(I did both training sessions, but only twice each, so I did not feel that I got much benefit from them.)

5.4 Discussion

This section begins with a summary of the results, followed by a discussion of the findings presented in the previous section and a comparison with those of previous studies on L2 speech learning. The discussion addresses each research question in turn.

Regarding perceptual learning, results revealed a significant improvement in discrimination accuracy for stimuli that appeared during training in the PERC-ONLY, MIX, and PERC-PROD groups. However, the learning generalised to a new voice only in the PERC-PROD group. In contrast, participants in the PROD-ONLY group did not significantly improve on the trained words spoken by a trained speaker, but they improved on the trained words spoken by an untrained speaker and untrained words spoken by a trained speaker. In the identification task, all groups except for the PERC-ONLY group significantly improved on the trained words spoken by a trained speaker. The learning generalised to a new voice only in the MIX group, and this trend was also observed in the PROD-ONLY group. In the PERC-ONLY group, participants showed significant improvement on trained words when spoken by an untrained speaker, as well as on untrained words when spoken by a trained speaker; however, they did not improve on the stimuli used during training.

Production results revealed that the PERC-ONLY group did not significantly improve the accuracy of the target phonemes in any of the speaking tasks based on both the identification rating and goodness rating tasks. In contrast, the other three groups significantly increased their identification rating scores in the word-list reading task compared to the PERC-ONLY group. Only

the PROD-ONLY and PERC-PROD groups significantly improved their production in the picture description task. None of the groups showed significant improvements in the sentence reading task. The goodness rating results generally aligned with the identification rating results. Participants in the PERC-ONLY group did not significantly improve their goodness rating scores in any of the speaking tasks. In contrast, those in the other groups significantly improved in the word-list reading task, with only the PROD-ONLY and PERC-PROD groups showing significant improvements in the sentence reading task.

Focusing on PSTM's role in phonetic learning, the PSTM score was not a significant predictor of either discrimination or identification accuracy before training. However, for identification accuracy only, the influence of PSTM capacity appeared to vary depending on the pre-test level among the groups. In the MIX group, PSTM capacity positively impacted perceptual learning, while it had a negative impact at lower pre-test scores in the PERC-ONLY group. In contrast, PSTM capacity was negatively associated with the gain in discrimination accuracy at lower pre-test scores in the PROD-ONLY and PERC-PROD groups, and it shifted to a positive association as the pre-test scores increased. In terms of production accuracy, the PSTM score was not a significant predictor of the gain score before training. Regarding the degree of learning in production, PSTM capacity was significantly predicted gain score only in identification rating across all groups.

The exit questionnaire revealed that most participants possessed a desire to improve their English skills and pronunciation. The majority of participants rated the training tasks as effective or neutral, noting that training gave them opportunities to practice perception and/or production and focus on the differences between the target phonemes, fostering metacognitive skills in phonetic learning. However, some participants expressed reluctance to continue due to the intensity of the training and the lack of feedback. Among the training types, PROD-ONLY training using ASR was the most favoured. Indeed, participants in the PERC-PROD group preferred PROD-ONLY training over PERC-ONLY training, as they believed that pronunciation practice was more effective for phonetic learning. ASR training was generally well-received; however, some participants expressed frustration due to misperceptions by the ASR system.

(1) Does perceptual training in an identification task format improve L2 phonetic perception and production?

The findings revealed that PERC-ONLY training significantly improved discrimination accuracy only for stimuli that appeared in training, suggesting no generalisation to new words or a new voice. However, it did improve identification accuracy for trained words spoken by an untrained speaker and untrained words spoken by a trained speaker, with no significant improvement observed for the trained stimuli. These findings suggest that participants' learning may not have been robust enough to show consistent improvements in perception performances across stimulus types, including both trained and untrained stimuli. The findings contradict those of previous studies that have shown more robust learning and generalisation to new words and voices (e.g., Bradlow et al., 1997; Iverson et al., 2012; Lively et al., 1993; Logan et al., 1991; Shinohara & Iverson, 2018; Thomson, 2011).

However, some previous studies on HVPT have also reported a lack of generalisation effects or minimum improvement. For example, Sonu et al. (2013) conducted perceptual training on Japanese geminates with L1 Korean speakers. Although participants improved their perceptual accuracy for trained stimuli, they still had difficulty perceiving Japanese geminates after training. Additionally, a study by Perrachione et al. (2011) indicated that the effectiveness of HVPT can depend on individual perceptual abilities, with some learners, particularly those with weaker perceptual skills, finding HVPT less effective compared to LVPT, suggesting that not all learners benefit equally from HPVT.

One possible explanation for this inconsistency is the intensiveness of training. In Experiment 3, the number of trials per session was increased with the aim of enhancing learning outcomes. However, this increase did not appear to enhance perceptual learning, when comparing the results with those of Experiment 2. In this experiment, the PERC-ONLY group's discrimination accuracy increased by 3.8% and their identification accuracy increased by 6.6% with all stimuli combined, while in Experiment 2, the PERC-ONLY group improved by 4.2% in discrimination accuracy and 7% in identification accuracy. These findings suggest that training over a longer period might be necessary for robust learning as discussed in the discussion for Experiment 1 (Chapter 3.4).

It is also possible that participants' lack of concentration during training led to lower learning outcomes compared to previous studies. The PERC-ONLY training consisted of a task that lasted 10 to 15 minutes and which could be perceived as monotonous by participants. Such training, particularly with an increased number of trials (i.e., 150 trials per session), might have been too tedious to maintain concentration throughout, leading to lower learning outcomes compared to previous studies. Some studies have shown significant improvements in perception with a large number of trials, such as 200 trials in Thomson (2011) and 300 trials in Shinohara and Iverson

(2018). However, in these studies, participants received training in a lab-based or quasi-lab-based environment, which provided optimal conditions for concentration. In this Experiment, participants completed training at their own pace in locations of their choice, and for some, the environments might not have been ideal for maintaining concentration. In fact, one participant mentioned in the exit questionnaire that it was challenging for them to find a quiet environment for training. Moreover, the training period was from June to July 2022, which was around the end of the COVID-19 pandemic, a time when many people were already fatigued by the increased use of computers. These factors might have led to a lack of focus on training, leading to less learning outcomes.

Taken together, while participants may benefit from longer perceptual training session in a lab-based environment, actual learners in language classes might benefit from shorter training session by focusing on training. Previous research indicates that spaced learning leads to better learning outcomes than massed learning (Kim & Webb, 2022). Therefore, incorporating additional two training sessions, rather than increasing the number of trials per session, may have yielded better learning outcomes.

Another key finding on PERC-ONLY training is that perceptual learning did not transfer to production learning, in contrast to many previous studies (Bradlow et al., 1999; Bradlow et al., 1997; Huensch & Tremblay, 2015; Inceoglu, 2016; Iverson et al., 2012; Lambacher et al., 2005; Lengeris, 2018; Lopez-Soto & Kewley-Port, 2009; Motohashi-Saigo & Hardison, 2009; Shinohara & Iverson, 2018; Thomson, 2011; Wong, 2014). One explanation is the overall reduced learning in perception, suggesting that for more robust learning to occur, a transfer from perceptual to production learning may be necessary.

Another factor to consider is the difficulty of the target contrast. Studies have demonstrated that English /l/-/r/ contrasts are particularly challenging for L1 Japanese speakers (Goto, 1971; MacKain et al., 2008; Saito, 2011; Sheldon & Strange, 1982). According to Flege's SLM (1995a), these phonemes are difficult for L1 Japanese speakers to acquire because Japanese has a flap /r/, which is similar to the English /l/ and /r/. PAM (Best, 1995; Best & Tyler, 2007) also predicts that distinguishing this contrast is challenging because both sounds can be assimilated into the single phoneme flap /r/. Indeed, studies have found that many L1 Japanese speakers did not achieve robust accuracy even after undertaking perceptual training (Bradlow et al., 1999; Bradlow et al., 1997; Shinohara & Iverson, 2018; Ueda & Hashimoto, 2019). For instance, Ueda and Hashimoto (2019) found that L1 Japanese learners of English significantly improved their perception and production accuracy of the English /b/-/v/ and /s/-/θ/ contrasts following classroom-based perceptual training.

However, no significant improvement was observed for the /l/–r/ contrast. Therefore, the difficulty of the contrast itself could be a reason for the lack of improvement in production.

This finding of the null effect of perceptual training on production learning provides an important perspective on the relationship between perception and production learning. This experiment revealed that, despite significant improvement in perception, production did not show significant improvement. This indicates that learning in these two modalities does not mirror each other and may be independent, especially at the initial stage of phonetic category formation. The relationship between perception and production is further discussed in General Discussion (Chapter 6.2).

(2) Does production training using the ASR system improve L2 phonetic perception and production?

The results revealed that PROD-ONLY training using the ASR system significantly improved participants' production accuracy, replicating findings from previous studies on production training in general (e.g., Akahane-Yamada et al., 1997; Hardison, 2004; Hattori & Iverson, 2009; Herd et al., 2013; Hirata, 2004) and production training using an ASR technology (e.g., Chen et al., 2020; García et al., 2020; Inceoglu et al., 2020; Liakin et al., 2013, 2015; McCrocklin, 2019; Spring & Tabuchi, 2022). This experiment further supports the idea that production training, particularly with the use of an ASR system, can be an effective tool for improving challenging pronunciation. Moreover, given that the PERC-ONLY group did not significantly improve their production accuracy, the results suggest that production training brings about better learning outcomes than perceptual training. These findings align with some previous studies (Catford & Pisoni, 1970; Lacabex et al., 2008; Linebaugh & Roche, 2015).

Furthermore, this improvement in production also transferred to perceptual improvement, aligning with previous research findings (Herd et al., 2013; Hirata, 2004; Kartushina et al., 2015; Kissling, 2014; Sakai, 2016; Thorin, 2020). These results suggest that production learning influences perceptual learning, challenging the theory that perception precedes production (Flege, 1995a). Other studies, however, have found that production training did not lead to improvements in perceptual accuracy (Hattori, 2009; Tateishi & Winters, 2015). Notably, in production training studies using ASR systems, Liakin et al. (2013, 2015) found no transfer effect from production learning to perceptual learning. A potential explanation for this discrepancy is the difference in participants' proficiency in the target language. The participants in Liakin et al.'s studies were beginner learners, whereas participants in the current experiment had prior exposure to the target phonemes before

training. This prior exposure may have facilitated the transfer of learning between production and perception. Considering all of these mixed findings, it can be suggested that learning phases of the two modalities' are independent or may change over time during the course of learning (Nagle & Baese-Berk, 2021).

It is also important to note that while both PROD-ONLY training and MIX training significantly improved perceptual accuracy, PROD-ONLY training led to better learning outcomes in production. The MIX group's improvement appeared only in the word-list reading task, not in other speaking tasks. This occurred despite both groups practicing pronouncing words with the target phonemes the same number of times during training. The main differences between the two training types are whether a perceptual training element was involved and immediate feedback on pronunciation was provided.

Regarding the perceptual training element, this experiment showed that learners could improve their perceptual accuracy through pronunciation practice without directly hearing the target sounds. However, unlike in Sakai (2016), where participants listened to white noise and could not hear their own speech during production-only training, participants in this experiment could hear themselves, allowing for ecological validity (van Lier, 1997). It could be argued that hearing their own speech without hearing L1 speakers' model pronunciation can disrupt perceptual learning. However, Baese-Berk and Samuel (2016) found that hearing one's own inaccurate speech does not disrupt perceptual learning. Therefore, as participants' pronunciation improved, hearing their own speech may have influenced their perception positively rather than negatively.

In terms of feedback, this study suggests that immediate feedback on pronunciation is crucial, supporting previous research (Kartushina et al., 2015; Neri et al., 2006; Saito & Lyster, 2012; Wang & Young, 2015). One possible explanation for the importance of immediate feedback is that learners, particularly beginners, may struggle to accurately judge whether their pronunciation is target-like. Research has shown that lower proficiency learners often overestimate their pronunciation accuracy, which further emphasises the need for feedback (Trofimovich et al., 2014). Their perceptual ability may not be fully developed, making them unaware of their mispronunciations. Immediate feedback may be necessary for them to notice and improve their pronunciation.

(3) Does production during perceptual training in an identification task format disrupt L2 phonetic perceptual and production learning?

The MIX group significantly improved their perceptual accuracy as much as PERC-ONLY training, contradicting the findings of Experiments 1 and 2, as well as previous studies (Baese-Berk, 2010, 2019; Baese-Berk & Samuel, 2016; Wright, 2021). Baese-Berk and Samuel (2022) found that switching between perception and production tasks was cognitively costly, leading to interference in perceptual learning. This could explain why disruption was observed in Experiments 1 and 2. In these experiments, one training session consisted of 100 trials, which was 50 trials fewer than those in Experiment 3. Such shorter training sessions may have made switching between the two tasks too complex for participants. In contrast, in Experiment 3, the longer training sessions might have allowed participants to become more accustomed to the task format and maintain better focus on the training, resulting in better learning. Baese-Berk (2019) found that participants who repeated stimuli aloud achieved robust perceptual learning and caught up with those who underwent perception-only training after an extra day of training. Thus, providing enough time for participants to adapt to the training task might also be crucial for optimising learning outcomes when incorporating oral repetition during perception training.

However, the results of this experiment also support that oral repetition during perceptual training does not bring about additional learning compared to perception-only training. This finding aligns with those from studies conducted by Thorin et al. (2018), Thorin (2020), and Lu et al. (2015). A potential reason could be attributed to a lack of feedback as discussed in Discussion of Experiment 1 (Chapter 3.4).

Regarding production learning, the MIX group significantly improved their production accuracy, which is consistent with findings from previous studies (Baese-Berk, 2010, 2019; Thorin et al., 2018). However, they improved only in the word-list reading task, which was the most controlled speaking condition. Moreover, the improvement did not generalise to untrained words. These findings suggest that participants in the MIX group produced trained words more accurately when they could pay attention to each word carefully compared to more spontaneous conditions. This indicates that although their production became more accurate, their production ability was not yet robust. Since ASR-based PROD-ONLY training showed improvement in less controlled speaking tasks, it implies that feedback may be necessary for constructing phonetic categories and achieving more robust improvements in both perception and production accuracy, as discussed earlier.

(4) Does prior exposure to target phonemes alleviate the disruption of production to L2 perceptual and production learning?

The results of this experiment revealed that perceptual learning was not disrupted by production during perceptual training, and both groups showed significant improvements in both discrimination and identification accuracy. However, similar to the MIX group, the improvement did not generalise to untrained word stimuli. This finding suggests that more intensive learning might be necessary for robust learning as discussed above. Nevertheless, considering that half the amount of perceptual training in the PERC-PROD group yielded an equivalent degree of perceptual learning as the PERC-ONLY group, it is likely that production training that followed perceptual training facilitated perceptual learning. This indicates a link between production learning and perceptual learning, as observed in the PROD-ONLY group.

In terms of production learning, while the MIX group's improvement only appeared in the most controlled speaking task with trained words based on an identification rating, the PERC-PROD group's improvement appeared in less controlled speaking tasks with untrained words. This finding is similar to that of the PROD-ONLY group. Considering that participants in the PERC-PROD group produced target phonemes only half the number of times compared to the MIX group or PROD-ONLY group, the results suggest that initial exposure to target phonemes may enhance production learning. This supports findings from Baese-Berk and Samuel's (2016) study, which also highlight the positive impact of prior exposure on phonetic learning.

This finding suggests that early exposure to target sounds allow learners to form phonetic categories, which could serve as a foundation for production practice and positively impact perceptual learning. This idea aligns with Krashen's Input Hypothesis (1982), which posits that learners acquire a language by receiving comprehensible input slightly above their level. The Input Hypothesis also explains that there is a phase where learners focus on listening before they begin producing the language. This silent period allows learners to internalise the input and build mental representations of the linguistic aspects of the language. Moreover, Kissling (2014) found that learners' perceptual accuracy can predict learning outcomes in production after phonetics instruction, indicating the significance of perceptual learning before production practice. Thus, the hypothesis supports the notion that the initial phase of exposure to target sounds plays a significant role in pronunciation learning.

However, it is also important to note that participants in the PERC-ONLY group did not improve production accuracy, while those in the PERC-PROD group did. Given the PERC-ONLY group's lack of improvement in production, this highlights the necessity of production training for enhancing production accuracy. This experiment's results contribute to an emerging understanding of how perception and production interact during language learning, suggesting that production

training can effectively enhance perceptual learning when appropriately timed in the training sequence. Indeed, particularly for beginners, it is important to first focus on perception and gradually incorporate production practice as their phonetic categories are being constructed. In contrast, for learners who have already been exposed to the target language, such as advanced learners or those in immersion settings, the initial exposure phase might not be necessary because production is less likely to disrupt phonetic learning after phonetic categories of the target language have been formed to some degree.

(5) Is phonetic learning related to learners' PSTM?

As observed in Experiments 1 and 2, the PSTM score was not a significant predictor of the pre-test score of either the discrimination task or the identification task. As mentioned in the Discussion of Experiment 1, this finding aligns with the results of Safronova and Mora (2012) and Ghaffarvand Mokari and Werner (2019) but differs from other studies (Cerviño-Povedano et al., 2011; Inceoglu, 2019; Kondo, 2019; MacKay et al., 2001). The contradictory findings might be explained by the inconsistency between mora-based items in the PSTM task and syllable-based words in training and the proficiency of the target phonemes before training (see Chapter 3.4).

Moreover, PSTM capacity was not found to significantly predict production accuracy before training. This finding aligns with some studies (Kissling, 2014; Slevc & Miyake, 2006; Venkatagiri & Levis, 2007) but contrasts with others (Kondo, 2012; Nagle, 2013). The same reasons that may explain the lack of a connection between PSTM capacity and perceptual accuracy before training, as discussed above, could also apply to production accuracy.

It is also possible that the relationship shifts over depending on the stage of language learning. Previous studies found that the relationship between PSTM capacity and production accuracy may vary depending on the learners' proficiency of a target language. For example, Zahler and Lord (2022) showed that PSTM emerged as a significant predictor for undergraduate students but not for graduate students. Moorman (2017) found that it was a significant predictor only for advanced students, while Serafini and Sanz (2015) revealed that the significant relationship only appeared in the initial stage of learning. Moreover, the results of both Experiment 2 and this experiment suggest that the way PSTM capacity influences perceptual learning through training varied depending on the initial level of perceptual accuracy. Therefore, the relationship between PSTM capacity and production accuracy may change depending on the learning stage, and this

Experiment might not have captured a strong association because participants' learning history varied before training.

Regarding the change in the influence of PSTM capacity on perceptual learning, a similar finding was observed in Experiment 2: some groups showed either a positive or negative relationship, while the relationship changed from a negative relationship to a positive relationship as the pre-test score increased. As discussed in Chapter 4.4, a negative relationship was observed at lower pre-test scores potentially because improvement is likely to occur regardless of PSTM capacity in the initial learning stage. However, improving accuracy is getting more challenging as learning proceeds, and this may be the phase where PSTM capacity plays a role. Only in the PERC-ONLY, a negative relationship was observed at the higher pre-test scores. This negative relationship might be due to the lack of improvement in perception after training in the PERC-ONLY group.

It is also important to note that results revealed a significant relationship between PSTM capacity and the degree of production learning (gain scores). However, it was limited to the results from identification rating. The inconsistency cannot be explained solely from the results of this experiment. Moreover, although the results revealed no significant relationship between PSTM capacity and production accuracy before training, the *p* value was approaching significance for the goodness rating score. Considering these factors, one's PSTM capacity may play a significant role in both perceptual and production learning, and it is important to conduct further research on the relationship between PSTM and phonetic learning, particularly including production learning.

(6) How do learners perceive each phonetic training?

Overall, the majority of participants rated each training method as either effective or neutral. Many participants highlighted that listening to words while focusing on the distinction of the target contrast was effective, and they gradually developed an understanding of the English /l/-/r/ distinction as training progressed. Both PERC-ONLY and PROD-ONLY training appeared to stimulate learners' metacognitive skills in phonetic learning. Metacognitive skills, defined as one's conscious awareness of their own (learning) process and their ability to monitor and evaluate it (Flavell, 1979), have been shown to be one of the most important skills for successful learning (Wang et al., 1990) and play an essential role in pronunciation learning (He, 2011). Some participants observed that training made them aware of their lack of ability in distinguishing target contrasts and encouraged them to think about how to improve their pronunciation. Similarly, previous studies have found that production training fosters learners' autonomy (McCrocklin, 2016)

and increases their awareness of intelligibility (Mroz, 2018). These findings suggest that phonetic training not only enhances learners' metacognitive skills but also encourages them to improve their own pronunciation, which is a key aspect of phonetic learning.

This survey also underscored the significance of feedback in the learning process. Several participants indicated that they wanted feedback after completing the post-test perceptual tasks, highlighting the learners' desire for more feedback. Others mentioned that they were uncertain whether their pronunciation was improving. Although immediate feedback was provided during training, the absence of summative feedback that shows their overall progress may have left participants uncertain about their learning. Offering summative feedback that tracks learners' learning could give them a clearer sense of their improvement and potentially boost their motivation.

Regarding the PROD-ONLY training with ASR technology, it was perceived as the most effective method, with 54.5% of participants finding it effective. This finding aligns with those from Chen et al. (2020), where 55% of participants also found ASR-based training useful. Additionally, participants in the PERC-PROD group rated PROD-ONLY training as more effective than PERC-ONLY training. Some participants doubted the effectiveness of perceptual training for improving pronunciation and expressed a preference for tasks that allowed them to practice pronouncing the target sounds. These responses suggest that learners may favour more interactive tasks over passive ones. Therefore, when introducing perceptual training in class, it is important to advise learners of the purpose of perceptual training; otherwise, learners may not understand its value compared to pronunciation practice.

As for the future use of phonetic training, some participants expressed a desire to continue because they found it both enjoyable, similar to a game, and effective. However, others were reluctant to continue due to its accessibility. For example, some participants found it difficult to find a suitable environment to do training tasks. In this experiment, participants were required to use their PCs to complete training. If training was accessible on smartphones, which is a more commonly used device for Japanese students, it could reduce barriers to continue. For example, Thomson has recently launched an app called English Accent Coach Asteroids (2024). This app incorporates a gamification element, allowing learners to practice English vowels by listening to each sound and selecting which sound they heard. If learners make too many mistakes, the game ends and they need to start a new game. In addition to immediate feedback, summative feedback is provided, enabling learners to review which vowels they need to improve on (see Chapter 7.1 for more about gamification).

Another issue raised by participants was the difficulty of the distinction between English /l/ and /r/, and some were unsure whether they were improving, particularly during perceptual training. Providing summative feedback could help learners recognise and track their progress. Furthermore, more intensive learning may be necessary considering learning outcomes in this experiment. However, some participants felt that the training was too long. In this experiment, the number of trials per session was increased to optimise learning outcomes, but adding more sessions with fewer trials per session might have been more effective. In fact, one participant in the PERC-PROD group noted that two sessions per training type were insufficient to achieve noticeable improvement in each modality.

Regarding production training using ASR technology, some participants expressed frustration with the ASR system's misperceptions. This issue has been observed in other studies as well (Chen et al., 2020; Inceoglu et al., 2020; Liakin et al., 2017; McCrocklin, 2016; Spring & Tabuchi, 2021). From the learners' perspective, the technology is not always accurate, leading to situations where they believe they are pronouncing words intelligibly but receive incorrect feedback from the ASR system. Despite this, numerous studies have shown that current ASR technologies mostly align with human raters (Guskaroska, 2019; Inceoglu et al., 2023; McCrocklin & Edalatishams, 2020; McCrocklin et al., 2019). To address this issue, it may be beneficial to demonstrate the ASR system's accuracy to learners and advise them to overlook certain inaccurate feedback, such as homophones, and focus on the target sounds they are practicing.

These findings suggest several important considerations for improving phonetic training methods from the learners' perspective. However, one limitation of this survey is the sample size as not many participants completed it. In addition, some participants mistakenly thought that the pre- and post-tests were part of training, which may have biased their responses. While this experiment offers valuable insights into how participants perceived each training task, further research with a larger sample size and clearer instructions regarding the research procedure is needed to more thoroughly explore learners' perceptions and beliefs about phonetic training.

5.5 Summary of Experiment 3

Experiment 3 investigated the effects of four phonetic training types (i.e., PERC-ONLY, PROD-ONLY, MIX, and PERC-PROD training) on L2 learners' phonetic perception and production. All groups significantly improved perceptual accuracy, while only three groups that involved production training significantly improved production accuracy, suggesting the necessity of

production practice for improving pronunciation. The PROD-ONLY group showed transfer from production learning to perceptual learning, suggesting that production learning may precede perception learning. Although the MIX group improved without disruption, no additional learning effects were noted. The PERC-PROD group's significant production learning, potentially surpassing that of the PROD-ONLY group, emphasises the value of prior exposure to target sounds before production learning.

PSTM capacity was not a significant predictor of either perceptual or production accuracy before training. As for learning gains, the relationship between PSTM capacity and perceptual learning may vary based on initial perceptual ability, with greater PSTM potentially having a positive influence as learning proceeds. A positive relationship between PSTM capacity and production learning was observed only for identification rating. These findings highlight the need for further research that takes into account the different learning phases in both perception and production.

The exit survey revealed a preference for interactive tasks, such as ASR-based production training, although some participants were frustrated with ASR's misperceptions. All training types were generally considered effective and contributed to the development of metacognitive skills.

Chapter 6 General Discussion

The results of Experiments 1 and 2 revealed that perception-only training led to the most significant improvements in phonetic learning. Disruption occurred particularly when participants repeated the stimulus after answering a perceptual question, but not when repeating immediately after hearing it. These findings contrast with previous studies that showed consistent disruption (Baese-Berk, 2010; Baese-Berk & Samuel, 2016). A possible explanation could be the difference in participants' proficiency levels; while Baese-Berk and Samuel's (2016) participants were novice learners, this study involved lower-intermediate learners, potentially leading to varying effects of perceptual and production training. The lack of feedback during production might also explain these disruptions.

The findings from Experiment 3 revealed that although all groups significantly improved perceptual accuracy, perception-only training did not improve production, highlighting the necessity of production practice for improving pronunciation. In contrast, while perceptual training with oral repetition of stimuli improved production in the most controlled speaking task, production-only and perception-to-production training with ASR's feedback led to improvements in both controlled and less controlled speaking tasks. This suggests an important role of feedback in production learning. However, only in the condition with initial exposure did production learning generalise to untrained words based on identification ratings, despite the shorter duration of production training. This suggests that initial exposure to target phonemes may result in better learning outcomes in production than training focused solely on production.

The results of the association between PSTM capacity and pre-training performances revealed no significant relationship with either perception or production accuracy. Regarding the degree of learning after phonetic training, there were mixed findings, including positive, non-significant, and negative relationships. However, these findings suggest that the relationship between PSTM capacity and perceptual learning may shift over the course of learning.

6.1 Production effects in perceptual training

This research revealed mixed findings regarding the effect of production on perceptual learning. In Experiments 1 and 2, disruption in perception occurred when production followed a perceptual question. Researchers have considered several reasons for the disruption effect of production, including the heavy cognitive cost of storing a perceptual answer while pronouncing target sounds (Thorin et al., 2018) and learners hearing their own inaccurate pronunciation (Baese-

Berk & Samuel, 2016). Thorin et al. (2018) argued that disruption occurred only in studies conducted by Baese-Berk and Samuel (2016) because participants needed to store their perceptual answer in their short-term memory while producing a word. In contrast, participants in Thorin et al.'s study answered a perceptual question before producing a word. The researchers claimed that the training design in Baese-Berk and Samuel was more cognitively costly, leading to disruption in perceptual learning. Experiments 1 and 2 examined whether the timing of the oral repetition affected perceptual learning. The results revealed that the disruption occurred particularly when production followed a perceptual question, rejecting the hypothesis that the timing of production was the cause.

Another possible factor for the disruption is hearing inaccurate pronunciation by producing target phonemes while participants' perceptual ability is not fully developed. However, Baese-Berk and Samuel (2016) found that producing an irrelevant letter also disrupted perceptual learning, suggesting that speech production itself can disrupt perceptual learning.

Baese-Berk (2019) hypothesised that the dual nature of the training task between perception and production may lead to disruption to perceptual learning. Indeed, Baese-Berk and Samuel (2022) found that waiting for several seconds before repeating a word aloud attenuated the disruption effect of production. More specifically, both short and long delays improved perceptual accuracy similar to perception-only training, but the long delay was more effective. Therefore, the task switching cost may provide a good explanation for why production during perceptual training disrupted perceptual learning.

In tasks that require switching between different cognitive activities, attention control is essential for managing the mental resources required to shift focus. Monsell (2003) explained that when people switch between tasks, their performance often slows down, a phenomenon known as 'switch cost.' This delay happens because of the need to reconfigure attention and adjust to the new task. Although preparing in advance can reduce the switch cost, residual effects of task switching remain, highlighting the complex demands of attention control. In fact, some studies have shown that better attention control leads to better perceptual ability (Darcy et al., 2014; Gökgöz-Kurt, 2016).

In this research, however, disruption was not observed in the discrimination task when production preceded a perceptual question. The disruption was observed more consistently when production followed a perceptual question, maybe due to the higher level of switching cost compared to the training where production preceded a perceptual question. In the condition where participants repeated a stimulus aloud before answering a perceptual question, speech perception and production components were completed before switching to answering a perceptual question by pressing a key. In contrast, in the other condition, the keypress answering task occurred between the perception and

production elements. From a task switching cost perspective, the condition where production followed a perceptual question might have been more cognitively demanding, at least in this research, leading to disruption in perceptual learning.

It is important to note that the disruption in perceptual learning disappeared in Experiment 3. A potential explanation for the different findings is the varying length of training. In Experiment 3, the number of trials per session increased compared to Experiments 1 and 2. This longer training may have attenuated the task switching cost that caused the disruption, as participants had more time to adapt to the training design, which was more complex than simple perception-only training. Baese-Berk (2019) found that an extra day of training mitigated disruption, and a combined training group caught up with the perception-only training group after the additional training day. In addition, studies where no disruption was observed involved more intensive training compared to those that showed disruptive effects of production (Thorin, 2020; Thorin et al., 2018). These findings suggest that when incorporating oral repetition of stimuli during perceptual training, the training period might need to be long enough for learners to adapt to the task to fully benefit from the training.

Nonetheless, disruption did not lead to additional perceptual learning. One potential reason for the null effect of oral repetition is the lack of feedback on production. Studies have shown that immediate feedback is essential for improving pronunciation (Dlaska & Krekeler, 2013; Flege, 1988; Kartushina et al., 2015; Lee et al., 2015; Neri et al., 2006; Saito & Lyster, 2012; Tejedor-García et al., 2017; Wang & Young, 2015). Thorin et al. (2018) argued that the absence of feedback during production practice could contribute to the null or disruptive effects in learning. This underscores the importance of integrating feedback mechanisms to maximise the effectiveness of production practice in phonetic learning. Therefore, for future research, it would be interesting to investigate whether providing feedback to production during combined training could result in better learning outcomes than perception-only training.

6.2 Relationship between speech perception and production learning

This study presents contradictory findings regarding the relationship between perceptual learning and production learning. The results from Experiments 1 and 2 showed that production during perceptual training disrupted perceptual learning, particularly when producing the target phonemes after a perceptual question. These disruption effects suggest a negative relationship between perception and production learning; production training negatively impacted perceptual learning, as observed in other studies (Baese-Berk, 2010, 2019; Baese-Berk & Samuel, 2016).

However, Experiment 3 revealed that the MIX training was as effective as perception-only training in perceptual learning, indicating that production did not lead to additional learning effects, and may not facilitate perceptual learning.

However, training focused solely on production significantly improved perceptual accuracy, supporting the idea that production learning can transfer to perceptual learning. These findings align with previous studies showing that production learning can precede perceptual learning, as some learners were able to produce sound that they cannot yet perceive accurately (Goto, 1971; Sheldon & Strange, 1982; Yamada et al., 1994). Moreover, a notable finding is that perception-only training did not significantly improve production accuracy.

These mixed findings challenge previous hypotheses about a straightforward relationship between perception and production in phonetic learning. Flege's SLM (1995a) proposed that perceptual learning precedes production, suggesting that learners cannot produce sounds that they are unable to perceive accurately. In contrast, the Motor Theory (Liberman & Mattingly, 1985) posits a direct link between perception and production with shared phonetic categories. Since the three experiments showed that, in some cases, production training can facilitate perceptual learning, and in other cases, it does not, the results do not fully support either the Motor Theory or the SLM. Rather, these findings suggest that perception and production learning may be more independent. However, the finding that exposure to target sounds before production training enhanced production learning implies that some connection between the two modalities does exist.

The mixed findings regarding the relationship between the two modalities in the three experiments suggest that this relationship is more complex and cannot be fully explained by simple models (Baese-Berk & Samuel, 2022). Given this complexity, some researchers argue that the relationship may evolve over time throughout the learning process (Baese-Berk, 2019; Baese-Berk & Samuel, 2016). Nagle and Baese-Berk (2021) propose that the relationship between perception and production varies based on the learners' developmental stage, with a weaker connection in the early stages of learning.

Yamada et al. (1994) found a positive correlation between perception and production of English /l/, /r/, and /w/ sounds among Japanese learners, but the correlation weakened when excluding participants who had more exposure to the target sounds. Similarly, Jia et al. (2006) found a weaker relationship among beginners compared to advanced learners, suggesting that perception and production are not strongly linked at early stages but become more aligned as learning progresses. Indeed, Liakin et al. (2013, 2015) did not observe a transfer from production learning to perceptual learning when beginner learners underwent production training, whereas this study found

a transfer from production learning to perceptual learning among participants with more learning experience in the target language.

These findings indicate that production learning can precede perceptual learning, with perceptual ability gradually catching up. However, improving pronunciation takes time, particularly for challenging sounds. To speed up the learning process, focussing on perceptual learning early on may support phonetic learning when perceptual development is still slower, as observed in the PERC-PROD group in this study.

However, most studies supporting the notion that production precedes perception are largely limited to the acquisition of English /l/–/r/ among L1 Japanese speakers (Goto, 1971; Sheldon & Strange, 1982; Yamada et al., 1994). This unique finding may be explained by the way Japanese learners of English are taught how to produce English /l/ and /r/ sounds. Goto (1971) suggested that L1 Japanese learners might rely on the physical sensations of their speech organs, much like deaf individuals might. For example, in school, they often receive an explanation about the tongue positions for these /l/ and /r/ sounds. This learning method might allow learners to become able to produce the sounds by memorising the tongue positions, but it may not help constructing robust phonetic categories for these sounds.

This may explain why some studies did not observe a transfer from production learning to perceptual learning (Hattori, 2009; Tateishi & Winters, 2015). For example, Tateishi and Winters (2015) trained L1 Japanese speakers to pronounce English /l/ and /r/ using ultrasound imaging to visualise tongue shapes and movements. In Hattori (2009), participants received one-on-one pronunciation training with explicit instructions, followed by practice using real-time spectrograms to check the F3 values for their production. During the training sessions, participants were instructed to check their tongue position and shape when their pronunciation did not match the target.

In both studies, the primary focus was on training participants to adjust their tongue position to achieve target-like pronunciation, rather than on the phonetic differences between sounds. This specific training method might explain why production learning sometimes precedes perceptual learning. In this research, production training may have improved perceptual accuracy because participants were not explicitly instructed on how to move their tongues, encouraging them to focus more on the sounds themselves. Baese-Berk (2019) argued that the connection between perception and production may vary depending on the similarities between a learner's L1 and the target language—some sounds might stand out more in perception, while others may be easier to produce. To better understanding the relationship between perceptual learning and production learning, further research is needed with different cases involving various sounds and/or L1s.

In summary, these findings suggest that perception and production abilities may not be tightly connected at the initial stage of phonetic learning, but they gradually start to influence each other as learning progresses. To fully investigate this complex learning trajectory, cross-linguistic longitudinal studies and consideration of individual differences are necessary, as Nagle and Baese-Berk (2021) suggested, cross-sectional studies may misrepresent the relationship between perception and production.

6.3 Relationship between PSTM and phonetic learning

The current three experiments consistently revealed that PSTM capacity did not predict perceptual accuracy before training. These findings align with those from studies conducted by Safronova and Mora (2012) and Ghaffarvand Mokari and Werner (2019) but contradict previous studies that showed PSTM capacity as a significant predictor (Cerviño-Povedano et al., 2011; Inceoglu, 2019; Kondo, 2019; MacKay et al., 2001).

One possible explanation for why PSTM was not related to perceptual accuracy before training could be the different learning histories among participants. In this research, participants were Japanese learners of English who had studied English in school for at least six years. Although none of them reported living or staying overseas for more than three months, the onset time of learning English varied. Therefore, this varied learning experience might explain the disconnection between PSTM and perceptual accuracy observed in this research.

Another explanation for the disconnection could be the differences in phonological systems between participants' L1 (Japanese) and the target language (English). Japanese is mora-based, while English is syllable-based. In studies that have shown a positive relationship between PSTM scores and perceptual accuracy, both the participants' L1 and L2 were syllable-based. Therefore, the PSTM task stimuli were composed of sequences of syllables. In contrast, the PSTM task stimuli used in this research consisted of sequences of three-mora words, following the PSTM task design from Kondo (2019). However, as processing words within an unfamiliar phonological system is more challenging for learners (Dupoux et al., 1999), it is possible that speech processing requires more cognitive resources when the words are based on an unfamiliar phonological system. Moreover, the differences in ease of processing between L1 and L2 might also show individual differences. Therefore, although PSTM should be measured using L1-based nonwords stimuli, it might be worth measuring participants' PSTM in both their L1 and the target language to examine potential performance differences between the two types of stimuli and to investigate whether either one is more closely related to phonetic learning.

Regarding the gain score in perception, mixed findings were observed, but two major insights emerged from the three experiments. One is that significant improvement may be a condition for observing a positive relationship between PSTM and perceptual learning. In Experiment 2, PSTM emerged as a significant predictor of identification accuracy only in the PERC-ONLY group, which was the only group showing a significant improvement in identification accuracy. This indicates that significant improvement could be a prerequisite for observing a significant relationship.

Another major finding is that the relationship between PSTM capacity and the degree of perceptual learning changes over the course of learning. Particularly, in the initial stage of phonetic learning, a negative relationship is likely to be observed, and this relationship shifts towards a positive one as learning progresses. Beginners are likely to improve their perception regardless of PSTM capacity, but as learning becomes more challenging, PSTM capacity may play an important role. In fact, Moorman (2017) showed that PSTM was a significant predictor of production learning only for advanced learners. This shift in the relationship over time may explain the mixed findings in previous research; some studies found a positive relationship (Cerviño-Povedano et al., 2011; Inceoglu, 2019; Kondo, 2019; MacKay et al., 2001), while others did not (Aliaga-Garcia et al., 2011; Ghaffarvand Mokari & Werner, 2019). However, Serafini and Sanz (2015) revealed that the significant relationship appeared only in the early stages of learning. These mixed findings suggest the importance of considering learning stages to fully capture the relationship between PSTM and L2 phonetic learning.

Regarding production, PSTM capacity was not a significant predictor before training, which is in line with findings from studies conducted by Kissling (2014), Slevc and Miyake (2006), and Venkatagiri and Levis (2007), but not with others (Kondo, 2012; Nagle, 2013), although it was approaching significance for goodness rating. However, in terms of the degree of learning, PSTM capacity was significantly related to the amount of learning only in the identification rating and not in the goodness rating. The contradictory findings may be attributed to the differences in what the two rating tasks measure. The identification rating focused on categorical intelligibility, while the goodness rating focused on the quality of the production of target phonemes. Therefore, it is possible that participants' improvement in intelligibility, rather than production quality, is more closely related to PSTM capacity.

As the current three experiments revealed mixed findings on the relationship between PSTM and L2 phonetic learning, more research is needed to further explore the role of PSTM. One idea for

future research is to control the word-familiarity variable for stimuli during phonetic training. Previous research has shown that PSTM plays an important role in learning unfamiliar items, but not in associating already-known items (Baddeley et al., 1988), suggesting that using unfamiliar words as stimuli in phonetic training might lead to different findings. For example, participants could receive phonetic training using only familiar words in one condition, and nonwords or unfamiliar words in the other condition. Some studies have found that using nonwords for L2 phonetic training is more effective than using existing words (Mora et al., 2022; Thomson & Derwing, 2016). Therefore, phonetic training with nonword stimuli might reveal clearer outcomes regarding training effects and their relationship with PSTM.

Another idea is to examine additional components of cognitive abilities that may contribute to a better understanding of individual differences in L2 phonetic acquisition. For example, studies have suggested that acoustic short-term memory, another component of working memory, may play a role in L2 speech learning (Kondo, 2019; Safronova, 2016; Tanaka & Nakamura, 2004). Acoustic short-term memory stores and processes acoustic information, which is not limited to linguistic features. In terms of speech processing, it involves the retention of suprasegmental features, such as pitch and tone. Other potential factors might be related to attention, such as attention and inhibitory control (Lev-Ari & Peperkamp, 2014; Mora & Darcy, 2016; Safronova, 2016; Zamuner, Yeung, et al., 2017). Attention control refers to the ability to focus on relevant information by filtering out distractions, which is essential for managing cognitive tasks. Inhibitory control is the cognitive ability to suppress irrelevant responses or information to focus on a specific task (Green, 1998). In bilinguals, this involves suppressing one language to prevent interference while using the other. Therefore, examining more cognitive features like those may help untangle the complexity of individual differences in L2 speech learning.

Chapter 7 Conclusion

This section discusses the pedagogical implications of the findings, outlines the limitations of this research, and suggests directions for future research.

7.1 Pedagogical implications

Several pedagogical implications emerged from the findings of this research. First, focusing on perception, particularly in the initial stages of learning, may be essential for successful phonetic acquisition. Experiments 1 and 2 showed that perception-only training improved the perceptual accuracy more effectively than training that involved production practice, and Experiment 3 revealed that participants who focused on perception before production practice improved their production accuracy effectively. These findings emphasise the importance of perceptual training at the initial phase of learning. This can be achieved by incorporating a lot of opportunities to listen to the target language in class, or more directly, by introducing perceptual training in class or as homework. Several online platforms, such as English Accent Coach (Thomson, 2012a) and English Accent Coach Asteroids (Thomson, 2024), offer free perceptual training in an HPVT format, providing learners with opportunities to practice phonemes.

However, as learning progresses, pronunciation practice becomes essential for improving pronunciation accuracy. When implementing pronunciation practice, it is important not only to have students repeat words but also to provide feedback on their pronunciation. This research found that while repeating model speech aloud during perceptual training can sometimes disrupt phonetic learning, production practice with feedback enhanced learning. In classrooms, language teachers often rely on listen-and-repeat methods for pronunciation practice. While repetition can be beneficial, providing effective feedback enables learners to identify specific pronunciation errors that they may not otherwise recognise on their own (Saito & Lyster, 2012).

There are various forms of feedback, ranging from teacher-provided feedback to technology-based ones like ASR. Given that ASR technology is readily accessible to learners today, it can serve as an excellent alternative to teacher feedback, especially when teachers are unable to provide individual feedback due to large class sizes and/or time constraints. For example, teachers can introduce ASR technology at the beginning of a course and assign pronunciation practice using ASR as homework. This allows students to receive feedback on their pronunciation even outside the classroom, fostering autonomy in their learning and motivating them to practice pronunciation more (McCrocklin, 2016).

Finally, long and monotonous training sessions tend to be less effective. In Experiment 3, the number of trials per session increased by 50 compared to Experiments 1 and 2 to aim for better learning. However, this increase did not yield the expected improvement. In addition, participants expressed boredom with the training, stating that its length made them feel less willing to continue. To ensure concentration and enhance engagement for better learning outcomes, incorporating various techniques to maintain student motivation may be key. One approach is employing a spaced learning design. A meta-analysis on L2 acquisition found that spaced learning yields better learning results than massed practice (Kim & Webb, 2022). Another method is integrating gamification elements, such as using apps (Saito et al., 2022; Thomson, 2024). Su et al. (2021) found that mobile-based vocabulary learning was more effective than non-mobile-based learning in improving engagement and learning outcomes. By combining these techniques, learners might remain motivated throughout training, leading to more effective learning.

7.2 Limitations and further research

A significant limitation of this research is the small sample size. As Plonsky's (2013) meta-analysis highlights, many studies in second language acquisition suffer from small sample sizes, which is problematic since findings based on small samples can be biased and may not reliably capture the true effects of treatment. In the series of experiments conducted for this thesis, data collection occurred online due to travel restrictions from the COVID-19 pandemic. This led to challenges in recruiting L1 Japanese participants. Additionally, a considerable amount of data had to be removed due to technological issues, such as low microphone quality and unstable Wi-Fi connections. For future research, it is recommended to replicate these experiments with an appropriate sample size, determined through power analysis, to confirm the reliability of the findings presented in this thesis.

Another limitation is the recruitment of L1 Japanese speakers from actual English classrooms. While this approach enhances the ecological validity of the research by capturing more authentic learning experiences (van Lier, 1997), it may not be the most ideal setting for examining L2 phonetic learning trajectories, particularly regarding the production effect during perceptual training. A lab-based condition might offer a more controlled environment, where participants can focus more intensely on phonetic training and the pre- and post-tests, completing tasks under the direct supervision of the researcher. This controlled setting would allow for stricter regulation of variables, potentially leading to more precise findings.

Moreover, controlling for participants' L2 proficiency and their prior exposure to the target language can be challenging in a classroom-based study. In this research, all participants had studied English for at least six years in school prior to training. Although none had spent more than three months in an English-speaking country, some had learned English outside of formal schooling and began their studies earlier than others. This variation in prior exposure makes it difficult to draw uniform conclusions about phonetic learning outcomes.

To address these challenges, future studies could benefit from working with beginner learners to better control for prior language exposure. For example, recruiting L1 English speakers for a study targeting L2 Japanese phonemes could offer a clearer view of the production effect and the relationship between PSTM and phonetic learning. This approach would help minimise variability in participants' prior knowledge of the target language and provide a more controlled understanding of the effects of phonetic training.

In addition, future research should consider the variation of English accents used in training stimuli. In this research, recordings from L1 Australian English speakers were used for perceptual training, and participants selected Australian English for ASR on Google Docs during production training to maintain consistency. However, Japanese learners of English are typically more familiar with American English. Australian English differs from American English in terms of pronunciation, particularly with vowel sounds (Cox, 1999) and non-rhoticity (Loneragan & Cox, 2010). Although the target phonemes in this research appeared only in word-initial positions, where pronunciation differences between the two accents are minimal, unfamiliarity with Australian English pronunciation may have caused confusion for some participants during training.

For example, several participants reported frustration due to ASR misperceptions, which may have been influenced by the vowel space differences between Australian English and their own pronunciation, which might be more closely aligned with American English. This unfamiliarity could also have affected their performance in perceptual training. Therefore, replicating this research using stimuli recorded by L1 American English speakers would offer valuable insights into whether these accent differences impact phonetic learning outcomes.

Finally, in this research, the improvements in perception and production for English /l/ and /r/ phonemes were combined for analysis. However, previous studies have shown mixed findings regarding which phoneme is more challenging. For example, Aoyama et al. (2004) found that Japanese learners of English improved more on /r/ than /l/, while Iverson et al. (2005) reported that /l/ is easier than /r/. Therefore, examining perception and production improvements for each phoneme separately could be beneficial in future research.

References

- Abbs, B., Gupta, P., & Khetarpal, N. (2008). Is overt repetition critical to expressive word learning? The role of overt repetition in word learning with and without semantics. *Applied Psycholinguistics*, 29(4), 627-667. <https://doi.org/10.1017/s0142716408080272>
- Akahane-Yamada, R., Adachi, T., & Kawahara, H. (1997). Second language production training using spectrographic representations as feedback. *Journal of the Acoustical Society of Japan (E)*, 18(6), 341-343. <https://doi.org/10.1250/ast.18.341>
- Aliaga-Garcia, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent research in second language phonetics/phonology: perception and production* (pp. 2-31). Cambridge Scholars Publishing.
- Aliaga-Garcia, C., Mora, J. C., & Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznan Studies in Contemporary Linguistics*, 47(1), 1-14. <https://doi.org/10.2478/psicl-2011-0002>
- Antolík, T. K., Pillot-Loiseau, C., & Kamiyama, T. (2019). The effectiveness of real-time ultrasound visual feedback on tongue movements in L2 pronunciation training. *Journal of Second Language Pronunciation*, 5(1), 72-97. <https://doi.org/10.1075/jslp.16022.ant>
- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the Case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, 32(2), 233-250. [https://doi.org/10.1016/s0095-4470\(03\)00036-6](https://doi.org/10.1016/s0095-4470(03)00036-6)
- Artieda, G., & Clements, B. (2019). A comparison of learner characteristics, beliefs, and usage of ASR-CALL systems. In F. Meunier, J. Van de Vyver, L. Bradley, & S. Thouéšny (Eds.), *CALL and Complexity – Short Papers from EUROCALL 2019* (pp. 19-25). <https://doi.org/10.14705/rpnet.2019.38.980>
- Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience. *Child Development*, 52(4), 1135-1145. <https://doi.org/10.1111/j.1467-8624.1981.tb03159.x>
- Baddeley, A. D. (1992). Working memory. *Science*, 255, 556-559. <https://www.jstor.org/stable/2876819>
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417-423. [https://doi.org/https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/https://doi.org/10.1016/S1364-6613(00)01538-2)
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47-89). Academic Press. [https://doi.org/https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/https://doi.org/10.1016/S0079-7421(08)60452-1)
- Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, 27, 586-595.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 575-589. [https://doi.org/https://doi.org/10.1016/S0022-5371\(75\)80045-4](https://doi.org/https://doi.org/10.1016/S0022-5371(75)80045-4)
- Baese-Berk, M. M. (2010). *An examination of the relationship between speech perception and production* [Northwestern University].
- Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, and Psychophysics*, 81(4), 981-1005. <https://doi.org/10.3758/s13414-019-01725-4>

- Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, 89, 23-36. <https://doi.org/10.1016/j.jml.2015.10.008>
- Baese-Berk, M. M., & Samuel, A. G. (2022). Just give it time: Differential effects of disruption and delay on perceptual learning. *Atten Percept Psychophys*, 84(3), 960-980. <https://doi.org/10.3758/s13414-022-02463-w>
- Bai, B., & Yuan, R. (2019). EFL teachers' beliefs and practices about pronunciation teaching. *ELT Journal*, 73(2), 134-143. <https://doi.org/10.1093/elt/ccy040>
- Berko, J., & Brown, R. (1960). Psycholinguistic research methods. In P. H. Mussen (Ed.), *Handbook of Research Methods in Child Development* (pp. 517-557). John Wiley & Sonss, Inc.
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H. C. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. MIT Press.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience. Issues in Cross-Language Research* (pp. 171-204). York Press. [https://doi.org/10.1016/0378-4266\(91\)90103-S](https://doi.org/10.1016/0378-4266(91)90103-S)
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-34). John Benjamins. <https://doi.org/10.1121/1.1332378>
- Bird, S., & Gick, B. (2018). Ultrasound biofeedback in pronunciation teaching and learning. In *Proceedings of the International Symposium on Applied Phonetics (ISAPh 2018)* (pp. 5-11). <https://doi.org/10.21437/ISAPh.2018-2>
- Bliss, H., Abel, J., & Gick, B. (2018). Computer-assisted visual articulation feedback in L2 pronunciation instruction: A review. *Journal of Second Language Pronunciation*, 4(1), 129-153. <https://doi.org/10.1075/jslp.00006.bli.Computer-Assisted>
- Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review*, 21(1), 149-154. <https://doi.org/10.3758/s13423-013-0485-1>
- Boersma, P. (2001). *Praat, a system for doing phonetics by computer*. In <http://www.praat.org/>
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. i. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics*, 61(5), 977-985. <https://doi.org/10.3758/BF03206911>
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. i. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of Acoustic Society of America*, 101(4), 2299-2310. <https://doi.org/10.1121/1.418276>
- Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly*, 22(4), 593-606. <https://www.jstor.org/stable/3587258>
- Bryfonski, L. (2023). Is seeing believing? The role of ultrasound tongue imaging and oral corrective feedback in L2 pronunciation development. *Journal of Second Language Pronunciation*, 9(1), 103-129. <https://doi.org/10.1075/jslp.22051.bry>
- Burgess, J., & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, 28, 191-215.
- Burns, A. (2006). Integrating research and professional development on pronunciation teaching in a national adult ESL program. *TESL Reporter*, 39(2), 34-41.
- Burri, M., & Baker, A. (2021). 'I feel ... slightly out of touch': a longitudinal study of teachers learning to teach English pronunciation over a six-year period. *Applied Linguistics*, 42(4), 791-809. <https://doi.org/10.1093/applin/amab009>

- Carlet, A. I., & Cebrian, J. (2019). Assessing the effect of perceptual training on L2 vowel identification, generalization and long-term effects. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A Sound Approach to Language Matters: In Honor of Ocke-Schwen Bohn* (pp. 91-119). Department of English, School of Communication & Culture, Aarhus University. <https://doi.org/10.7146/aul.322.218>
- Carley, P., & Mees, I. M. (2019). *American English Phonetics and Pronunciation Practice* (First ed.). Routledge.
- Catford, J., & Pisoni, D. B. (1970). Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal*, 54(7), 477-481. <https://www.jstor.org/stable/321767>
- Cerviño-Povedano, E., & Mora, J. C. (2011). Investigating Catalan-Spanish bilingual EFL learners' over-reliance on duration: Vowel cue weighting and phonological short-term memory. In K. Dziubalska-kołaczyk, M. Wrembel, & M. Kul (Eds.), *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010* (pp. 56-64). Peter Lang.
- Cerviño-Povedano, E., & Mora, J. C. (2015). Spanish EFL learners' categorization of /i:-ɪ/ and phonological short-term memory. *Procedia - Social and Behavioral Sciences*, 173, 18-23. <https://doi.org/10.1016/j.sbspro.2015.02.024>
- Cerviño-Povedano, E., Mora, J. C., & Aliaga-Garcia, C. (2011). Phonological short-term memory and language dominance in Catalan/Spanish bilingual EFL learner. In *The 5th Phonetics and Phonology in Iberia Conference PaPI-2011* (pp. 21-22).
- Chen, W.-H., Inceoglu, S., & Lim, H. (2020). Using ASR to improve Taiwanese EFL learners' pronunciation: Learning outcomes and learners' perceptions. In O. Kang, S. Staples, K. Yaw, & H. Kevin (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference* (pp. 37-48).
- Cox, F. (1999). Vowel change in Australian English. *Phonetica*, 56, 1-27. <https://doi.org/https://doi.org/10.1159/000028438>
- Cramer, D., & Howitt, D. (2004). *The SAGE Dictionary of Statistics: A Practical Resource for Students in the Social Sciences*. SAGE Publications.
- Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*. The MIT Press. <https://doi.org/https://doi.org/10.7551/mitpress/9012.001.0001>
- Dahlen, K., & Caldwell-Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal*, 97(4), 902-916. <https://doi.org/10.1111/j.1540-4781.2013.12045.x>
- Daidone, D. (2017). *Choosing a perception task*. <http://www.ddaidone.com/blog/choosing-a-perception-task>
- Daidone, D., & Darcy, I. (2021). Vocabulary size is a key factor in predicting second language lexical encoding accuracy. *Frontiers in Psychology*, 12, 688356. <https://doi.org/10.3389/fpsyg.2021.688356>
- Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation "toolbox". In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 3-108). Iowa State University.
- Darcy, I., Mora, J. C., & Daidone, D. (2014). Attention control and inhibition influence phonological development in a second language. *Concordia Working Papers in Applied Linguistics*, 5, 115-129.
- Darcy, I., Park, H., & Yang, C.-L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, 40, 63-72. <https://doi.org/10.1016/j.lindif.2015.04.005>
- de Jong, K., Hao, Y.-C., & Park, H. (2009). Evidence for featural units in the acquisition of speech production skills: Linguistic structure in foreign accent. *Journal of Phonetics*, 37(4), 357-373. <https://doi.org/10.1016/j.wocn.2009.06.001>

- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research*. John Benjamins Publishing Company.
- Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly*, 34(3), 592-603.
<https://doi.org/https://www.jstor.org/stable/3587748>
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29(3), 359-380.
<https://doi.org/10.1093/applin/amm041>
- Dlaska, A., & Krekeler, C. (2013). The short-term effects of individual corrective feedback on L2 pronunciation. *System*, 41(1), 25-37. <https://doi.org/10.1016/j.system.2013.01.005>
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C. C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology*, 25(6), 1568-1578.
<https://doi.org/10.1037/0096-1523.25.6.1568>
- Elimat, A. K., & AbuSeileek, A. F. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation. *The JALT CALL Journal*, 10(1), 21-47.
- Ellis, R. (2005). Principles of instructed language learning. *System*, 33(2), 209-224.
<https://doi.org/10.1016/j.system.2004.12.006>
- Escudero, P. (2005). *Linguistic perception and second language acquisition: explaining the attainment of optimal phonological categorization* [Doctoral dissertation, Utrecht University].
- Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: a meta-analysis. *Acta Psychologica*, 142(1), 1-5. <https://doi.org/10.1016/j.actpsy.2012.10.001>
- Felker, E., Broersma, M., & Ernestus, M. (2021). The role of corrective feedback and lexical guidance in perceptual learning of a novel L2 accent in dialogue. *Applied Psycholinguistics*, 42(4), 1029-1055. <https://doi.org/10.1017/s0142716421000205>
- Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, 51(2), 419-439.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive—development inquiry. *American Psychologist*, 34(10), 906-911.
- Flege, J. E. (1988). Using Visual Information to Train Foreign-Language Vowel Production. *Language Learning*, 38(3), 365-407. <https://doi.org/10.1111/j.1467-1770.1988.tb00417.x>
- Flege, J. E. (1995a). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 229-273). York Press. <https://doi.org/10.1111/j.1600-0404.1995.tb01710.x>
- Flege, J. E. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, 16(4), 425-442. <https://doi.org/10.1017/s0142716400066029>
- Flege, J. E. (1999a). Age of learning and second language speech. In D. Birdsong (Ed.), *Second Language Acquisition and the Critical Period Hypothesis* (pp. 101-132).
- Flege, J. E. (1999b). The relation between L2 production and perception. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)* (Vol. 2, pp. 1273-1276).
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 319-355). Mouton de Gruyter.
<https://doi.org/10.1515/9783110895094.319>
- Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437-470.
<https://doi.org/10.1006/jpho.1997.0052>
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of Acoustical Society of America*, 97(5), 3125-3134.

- Foote, J. A., Trofimovich, P., Collins, L., & Urzúa, F. S. (2013). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal*, 44(2), 181-196. <https://doi.org/10.1080/09571736.2013.784345>
- Forrin, N. D., Macleod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, 40(7), 1046-1055. <https://doi.org/10.3758/s13421-012-0210-8>
- French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, 29(3), 463-487. <https://doi.org/10.1017/s0142716408080211>
- Fujita, H., Ueda, R., & Hashimoto, K. (2019). Does repeated exposure to segmental sounds improve perceptual ability in non-Native speakers? In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2563-2566). Australasian Speech Science and Technology Association inc.
- Fujiyuki, C., Kitamura, M., Eguchi, S., Yang, S., & Akahane-Yamada, R. (2017). Training native speakers of Japanese to distinguish English /ɪ/ and /I/: Order effect of the perception training and production training. *Acoustical Science and Technology*, 38(4), 229-231. <https://doi.org/10.1250/ast.38.229>
- García, C., Nickolai, D., & Jones, L. (2020). Traditional versus ASR-based pronunciation instruction: An empirical study. *CALICO Journal*, 37(3), 213-232. <https://doi.org/10.1558/cj.40379>
- Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry*, 39(1), 3-27. <https://doi.org/10.1017/s0021963097001753>
- Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, 81(4), 439-454. <https://doi.org/10.1111/j.2044-8295.1990.tb02371.x>
- Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition*, 16(2), 110-119. <https://doi.org/https://doi.org/10.3758/BF03213478>
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40(2), 177-190. <https://doi.org/10.1037/0012-1649.40.2.177>
- Ghaffarvand Mokari, P., & Werner, S. (2019). On the Role of Cognitive Abilities in Second Language Vowel Learning. *Language and Speech*, 62(2), 260-280. <https://doi.org/10.1177/0023830918764517>
- Gökgöz-Kurt, B. (2016). *Attention control and the effects of online training in improving connected speech perception by learners of English as a second language* [Doctoral dissertation, University of South Carolina].
- Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19(5), 515-523. <https://doi.org/10.1111/j.1467-9280.2008.02117.x>
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, 9(3), 317-323. [https://doi.org/https://doi.org/10.1016/0028-3932\(71\)90027-3](https://doi.org/https://doi.org/10.1016/0028-3932(71)90027-3)
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2), 67-81. <https://doi.org/10.1017/s1366728998000133>
- Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics*, 72, 17-34. <https://doi.org/10.1016/j.wocn.2018.10.005>
- Grenon, I., Sheppard, C., & Archibald, J. (2018). Discrimination training for learning sound contrasts. In *Proceedings of the International Symposium on Applied Phonetics (ISAPh)* (pp. 51-56). <https://doi.org/10.21437/isaph.2018-9>

- Griffiths, S. K., Brown, W. S., Gerhardt, K. J., Abrams, R. M., & Morris, R. J. (1994). The perception of speech sounds recorded within the uterus of a pregnant sheep. *Journal of the Acoustical Society of America*, 96(4), 2055-2063. <https://doi.org/10.1121/1.410147>
- Guskaroska, A. (2019). The potential of ASR for facilitating vowel pronunciation practice for Macedonian learners. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*. Iowa State University.
- Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, 14(1), 31-38. <https://doi.org/https://doi.org/10.1111/1467-9280.01415>
- Hamada, A., & Tsushima, T. (2001). Relation between perception and production ability during a speech training course. In J. White (Ed.), *Proceedings of the Fourth Conference on Foreign Language Education and Technology* (pp. 579-582). Japan Association for Language Education and Technology.
- Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, 8(1), 34-52. <http://llt.msu.edu/vol8num1/hardison/>
- Hattori, K. (2009). *Perception and production of English /r/-/l/ by adult Japanese speakers* [University College London].
- Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America*, 125(1), 469-479. <https://doi.org/10.1121/1.3021295>
- Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, 23(1), 65-94. <https://doi.org/10.1177/0267658307071601>
- He, L. (2011). Metacognition in EFL pronunciation learning among Chinese tertiary learners. *Applied Language Learning*, 21, 1-27. <https://doi.org/10.5167/uzh-128569>
- Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Waniek-Klimczak, E., Levey, D., Cunningham, U., & Curnick, L. (2012). The English pronunciation teaching in Europe survey: Selected results. *Research in Language*, 10(1), 5-27. <https://doi.org/10.2478/v10015-011-0047-4>
- Henshaw, F. (2011). Effects of feedback timing in SLA: A computer-assisted study on the Spanish subjunctive. In C. Sanz & R. P. Leow (Eds.), *Implicit and Explicit Language Learning: Conditions, Processes, and Knowledge in SLA and Bilingualism* (pp. 73-84). Georgetown University Press.
- Hepper, P. G., & Shahidullah, B. S. (1994). The development of fetal hearing. *Fetal and Maternal Medicine Review*, 6, 167-179. <https://doi.org/10.1136/fn.71.2.f81>
- Herd, W., Jongman, A., & Sereno, J. (2013). Perceptual and production training of intervocalic /d, r, r/ in American English learners of Spanish. *J Acoust Soc Am*, 133(6), 4247-4255. <https://doi.org/10.1121/1.4802902>
- Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL*, 15(1), 3-20. <https://doi.org/10.1017/s0958344003000211>
- Hirata, Y. (2004). Computer assisted pronunciation training for native english speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning*, 17(3-4), 357-376. <https://doi.org/10.1080/0958822042000319629>
- Hu, X., Ackermann, H., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2013). Language aptitude for pronunciation in advanced second language (L2) learners: behavioural predictors and neural substrates. *Brain and Language*, 127(3), 366-376. <https://doi.org/10.1016/j.bandl.2012.11.006>

- Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics*, 52, 105-120. <https://doi.org/10.1016/j.wocn.2015.06.007>
- Hwang, H., & Lee, H.-Y. (2015). The effect of high variability phonetic training on the production of English vowels and consonants. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, 1041-1045.
- Hyunsu, P., & Joo-Kyeong, L. (2022). Bidirectional effects of training order on perception and production accuracy of English L2 vowels. *Studies in Phonetics, Phonology and Morphology*, 28(2), 229-257. <https://doi.org/10.17959/sppm.2022.28.2.229>
- Iino, A. (2018). Use of HVPT in CALL to assess less successfully recognized phonemes for Japanese EFL learners. *Hosei University Tama Ronju*, 34, 129-143. <https://doi.org/http://doi.org/10.15002/00014476>
- Iino, A. (2019). Effects of HVPT on perception and production of English fricatives by Japanese learners of English. In *CALL and Complexity – Short Papers from EUROCALL 2019* (pp. 186-192). <https://doi.org/10.14705/rpnet.2019.38.1007>
- Iino, A., & Thomson, R. I. (2018). Effects of web-based HVPT on EFL learners' recognition and production of L2 sounds. In *Future-Proof CALL: Language Learning as Exploration and Encounters – Short Papers from EUROCALL 2018* (pp. 106-111). <https://doi.org/10.14705/rpnet.2018.26.821>
- Iino, A., Yabuta, Y., & Wistner, B. (2020). Effects and users' reactions to the use of CAPT and HVPT on Japanese EFL learners' segmental perception and production. In K.-M. Frederiksen, S. Larsen, L. Bradley, & S. Thouësny (Eds.), *CALL for Widening Participation: Short Papers from EUROCALL 2020* (pp. 107-113). Research-publishing.net. <https://doi.org/10.14705/rpnet.2020.48.1173>
- Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics*, 37(5), 1175-1199. <https://doi.org/10.1017/s0142716415000533>
- Inceoglu, S. (2019). Individual differences in L2 speech perception: The role of phonological memory and lipreading ability. *The Modern Language Journal*, 103(4), 782-799. <https://doi.org/10.1111/modl.12591>
- Inceoglu, S. (2022). A web-based high variability phonetic training application for French vowels. In J. Levis & A. Guskaroska (Eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference, held June 2021 virtually at Brock University, St. Catharines, ON*. <https://doi.org/10.31274/psllt.13336>
- Inceoglu, S., Chen, W.-H., & Lim, H. (2023). Assessment of L2 intelligibility: Comparing L1 listeners and automatic speech recognition. *ReCALL*, 35(1), 89-104. <https://doi.org/10.1017/s0958344022000192>
- Inceoglu, S., & Gnevshva, K. (2020). Ultrasound imaging in the foreign language classroom: Outcomes, challenges, and students' perceptions. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference, Northern Arizona University, September 2019* (pp. 115-126).
- Inceoglu, S., Lim, H., & Chen, W.-H. (2020). ASR for EFL Pronunciation Practice: Segmental Development and Learners' Beliefs. *The Journal of AsiaTEFL*, 17(3), 824-840. <https://doi.org/10.18823/asiatefl.2020.17.3.5.824>
- Inoi, S. i. (2010). 外国語活動における小学校教員の負担感及び英語の好き嫌い [The burden perceived by elementary school teachers in foreign language education and their preferences toward English]. *TELES Journal*, 30, 45-54. <https://ndlsearch.ndl.go.jp/books/R0000000004-I10667733>

- Inoi, S. i. (2017). A report of changes in primary school teachers' perceptions of English pronunciation through attending an English pronunciation training course. *Bulletin of the Faculty of Education, Ibaraki University, Educational Science*, 66, 271-281.
- Irwin, O. C., & Curry, T. (1941). Vowel elements in the crying vocalization of infants under ten days of age. *Child Development*, 12(2), 99-109.
- Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r/-/l/ to Japanese adults. *J Acoust Soc Am*, 118(5), 3267-3278. <https://doi.org/10.1121/1.2062307>
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. i., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), B47-B57. [https://doi.org/10.1016/s0010-0277\(02\)00198-1](https://doi.org/10.1016/s0010-0277(02)00198-1)
- Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(1), 145-160. <https://doi.org/10.1017/s0142716411000300>
- Jakobson, R. (1968). *Child Language: Aphasia and Phonological Universals*. Mouton.
- Jakobson, R. (1971). *Studies on Child Language and Aphasia*. Mouton.
- Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics*, 40(4), 205-215. <https://doi.org/10.3758/BF03211500>
- Jia, G., Strange, I. W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of Acoustical Society of America*, 119, 1118-1130. <https://doi.org/10.1121/1.2151806>
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *J Acoust Soc Am*, 138(2), 817-832. <https://doi.org/10.1121/1.4926561>
- Kaushanskaya, M., & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes*, 26(1), 121-148. <https://doi.org/10.1080/01690965.2010.486579>
- Kent, R. D. (1981). Articulatory-acoustic perspectives on speech development. In R. E. Stark (Ed.), *Proceedings of the Language Behavior in Infancy and Early Childhood. Proceedings of a Pediatric Round Table held at the Santa Barbara Hotel. Santa Barbara, California* (pp. 105-126). Elsevier North Holland.
- Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning*, 72(1), 269-319. <https://doi.org/10.1111/lang.12479>
- Kissling, E. M. (2014). What predicts the effectiveness of foreign-language pronunciation instruction? Investigating the role of perception and other individual differences. *The Canadian Modern Language Review*, 70(4), 532-558. <https://doi.org/10.3138/cmlr.2161>
- Kochiyama, M., Arimoto, J., & Nakanishi, N. (2013). English pronunciation teaching in teacher's license courses. *Language Education & Technology*, 50, 119-130.
- Kondo, A. (2012). Phonological memory and L2 pronunciation skills. In A. Stewart & N. Sonda (Eds.), *JALT2011 Conference Proceedings*, 535-541.
- Kondo, A. (2019). The effects of verbal and acoustic short-term memory on Japanese EFL learners' sound recognition skills. *Language Education & Technology*, 56, 77-101. https://doi.org/https://doi.org/10.24539/let.56.0_77
- Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon Press Inc.
- Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, 4(6), 812-822. [https://doi.org/10.1016/0959-4388\(94\)90128-7](https://doi.org/10.1016/0959-4388(94)90128-7)
- Kuhl, P. K. (2000). A new view of language acquisition. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 97, pp. 11850-11857). National Academy of Sciences. <https://doi.org/10.4324/9780203929124>

- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831-843. <https://doi.org/10.1038/nrn1533>
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363(1493), 979-1000. <https://doi.org/10.1098/rstb.2007.2154>
- Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "Perceptual Magnet Effect". In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Research* (pp. 121-154). York Press.
<http://www.cs.indiana.edu/~port/teach/641/Kuhl.Iverson.lingc.exptc.percpt.magnt.pdf>
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2). <https://doi.org/10.1111/j.1467-7687.2006.00468.x>
- Kühnert, B., & Pillot-Loiseau, C. (2022). Teaching pronunciation with direct visual articulatory feedback: Pedagogical considerations for the use of ultrasound in the classroom. *Pronunciation Matters: Current Perspectives on Teaching and Learning L2 Phonology*, 55. <https://hal.science/hal-03838029>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). "lmerTest" package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-26. <https://doi.org/10.18637/jss.v082.i13>
- Lacabex, E. G., & Gallardo-del-Puerto, F. (2020). Explicit phonetic instruction vs. implicit attention to native exposure: phonological awareness of English schwa in CLIL. *International Review of Applied Linguistics in Language Teaching*, 58(4), 419-442. <https://doi.org/10.1515/iral-2017-0079>
- Lacabex, E. G., Lecumberri, M. L. G., & Cooke, M. (2008). Identification of the contrast full vowel-schwa: Training effects and generalization to a new perceptual context. *Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies*, 55, 173-196. <https://doi.org/10.5007/2175-8026.2008n55p173>
- Ladefoged, P. (2003). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Blackwell Publishing.
- Lambacher, S. (1999). A call tool for improving second language acquisition of English consonants by Japanese learners. *International Journal of Phytoremediation*, 21(1), 137-156. <https://doi.org/10.1076/call.12.2.137.5722>
- Lambacher, S., Martens, W., Kakehi, K., Marasinghe, c. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26, 227-247-227-247.
- Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). Routledge. <https://doi.org/https://doi.org/10.4324/9781315775661>
- Law, I. L. G., Grenon, I., Sheppard, C., & Archibald, J. (2019). Which is better : Identification or discrimination training for the acquisition of an English coda contrast. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 919-923).
- Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: when adults learn new words. *Cognitive Psychology*, 55(4), 306-353. <https://doi.org/10.1016/j.cogpsych.2007.01.001>
- Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, 38(1), 35-64. <https://doi.org/10.1017/s0272263115000194>
- Lee, B., Plonsky, L., & Saito, K. (2020). The effects of perception- vs. production-based pronunciation instruction. *System*, 88. <https://doi.org/10.1016/j.system.2019.102185>

- Lee, D. Y., & Baese-Berk, M. M. (2021). Non-native English listeners' adaptation to native English speakers. *JASA Express Letters*, 1(10), 105201. <https://doi.org/10.1121/10.0006558>
- Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), 345-366. <https://doi.org/10.1093/applin/amu040>
- Lee, Y. (2021). Effectiveness of mobile assisted pronunciation training in the acquisition of English vowels, /o/ and /ɔ/. *Foreign Languages Education*, 28(3), 53-76. <https://doi.org/10.15334/fle.2021.28.3.53>
- Lengeris, A. (2018). Computer-based auditory training improves second-language vowel production in spontaneous speech. *Journal of the Acoustical Society of America*, 144(3), EL165-EL171. <https://doi.org/10.1121/1.5052201>
- Lengeris, A., & Nicolaidis, K. (2014). English consonant confusions by Greek listeners in quiet and noise and the role of phonological short-term memory. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*(September), 534-538. <https://doi.org/10.21437/interspeech.2014-131>
- Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, 2(12), 59-67. <https://doi.org/10.1080/21548331.1967.11707799>
- Lenth, R. V. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. In <https://CRAN.R-project.org/package=emmeans>
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093-1096. <https://doi.org/10.1016/j.jesp.2010.05.025>
- Lev-Ari, S., & Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics*, 47, 36-46. <https://doi.org/10.1016/j.wocn.2014.09.001>
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369-377. <https://www.jstor.org/stable/3588485>
- Levis, J. M., Sonsaat, S., Link, S., & Barriuso, T. A. (2016). Native and nonnative teachers of L2 pronunciation: Effects on learner performance. *TESOL Quarterly*, 50(4), 894-931. <https://doi.org/10.1002/tesq.272>
- Liakin, D., Cardoso, W., & Liakina, N. (2013). Mobile speech recognition software: A tool for teaching second language pronunciation. *OLBI Working Papers*, 5, 85-99. <https://doi.org/10.18192/olbiwp.v5i0.1120>
- Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, 32(1), 1-25. <https://doi.org/10.1558/cj.v32i1.25962>
- Liakin, D., Cardoso, W., & Liakina, N. (2017). Mobilizing instruction in a second-language context: Learners' perceptions of two speech technologies. *Languages*, 2(3). <https://doi.org/10.3390/languages2030011>
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461. <https://doi.org/https://doi.org/10.1037/h0020279>
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358-368. <https://doi.org/https://doi.org/10.1037/h0044417>
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revisited. *Cognition*, 21, 1-36. [https://doi.org/0010-0277/85/\\$11.30](https://doi.org/0010-0277/85/$11.30)
- Linebaugh, G., & Roche, T. (2015). Evidence that L2 production training can enhance perception. *Journal of Academic Language & Learning*, 9(1), A1-A17. <https://journal.aall.org.au/index.php/jall/article/view/326>

- Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, 6(3), 1-10.
<https://doi.org/10.1111/1467-7687.00275>
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, 94, 1242-1255.
<https://doi.org/10.1121/1.408177>
- Lively, S. E., Pisoni, D. B., Akahane-Yamada, R., Tohkura, Y. i., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, 96(4), 2076-2087.
<https://doi.org/10.1121/1.410149>
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese Listeners To Identify English /R/ And /l/: A First Report. *Journal of the Acoustical Society of America*, 89(2), 874-886-874-886. <https://doi.org/10.1121/1.1894649>
- Lonergan, J., & Cox, F. (2010). Is there any evidence of rhoticity in historical Australian English? In L. de Beuzeville & P. Peters (Eds.), *Proceedings of the Australian Linguistic Society Conference 2008* (pp. 1-17). Australian Linguistic Society, University of Sydney.
- Lopez-Soto, T., & Kewley-Port, D. (2009). Relation of perception training to production of codas in English as a Second Language. *Proceedings of Meeting on Acoustics*, 6, 1-15.
<https://doi.org/10.1121/1.3262006>
- Lotto, A. J., Sato, M., & Diehl, R. L. (2004, June 11-13). Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. *From Sound to Sense: 50+ Years of Discoveries in Speech Communication*, Cambridge, MA.
- Lu, S., Wayland, R., & Kaan, E. (2015). Effects of production training and perception training on lexical tone perception—A behavioral and ERP study. *Brain Research*, 1624, 28-44.
<https://doi.org/10.1016/j.brainres.2015.07.014>
- Lüdecke, D. (2024). sjPlot: Data Visualization for Statistics in Social Science. <https://CRAN.R-project.org/package=sjPlot>
- Macdonald, S. (2002). Pronunciation—Views and practices of reluctant teachers. *Prospect: An Australian journal of TESOL*, 17(3), 3-18.
- MacKain, K. S., Best, C. T., & Strange, W. (2008). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2(4), 369-390.
<https://doi.org/10.1017/s0142716400009796>
- MacKay, I. R. A., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica*, 58, 103-125. <https://doi.org/10.1159/000028490>
- MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3), 671-685. <https://doi.org/10.1037/a0018785>
- Magnuson, J. S., Nusbaum, H. C., Akahane-Yamada, R., & Saltzman, D. (2021). Talker familiarity and the accommodation of talker variability. *Attention, Perception, & Psychophysics*, 83(4), 1842-1860. <https://doi.org/10.3758/s13414-020-02203-y>
- Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, 19(23), 1994-1997.
<https://doi.org/10.1016/j.cub.2009.09.064>
- Martin, I. A. (2020a). Pronunciation can be acquired outside the classroom: Design and assessment of homework-based training. *The Modern Language Journal*, 104(2), 457-479.
<https://doi.org/10.1111/modl.12638>
- Martin, I. A. (2020b). Pronunciation development and instruction in distance language learning. *Language Learning & Technology*, 24(1), 86-106. <https://doi.org/https://doi.org/10125/44711>

- Masaki, S., Akahane-Yamada, R., Tiede, M. K., Shimada, Y., & Fujimoto, I. (1996). An MRI-based analysis of the English /r/ and /l/ articulations. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*,
- Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Boston University Conference on Language Development* (pp. 522-533). Cascadia Press.
- Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us. *Proceedings of the 25th Annual Boston University Conference on Language Development*.
- McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, 57, 25-42. <https://doi.org/10.1016/j.system.2015.12.013>
- McCrocklin, S. (2019). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation*, 5(1), 98-118. <https://doi.org/10.1075/jslp.16034.mcc>
- McCrocklin, S., & Edalatshams, I. (2020). Revisiting popular speech recognition software for ESL speech. *TESOL Quarterly*, 54(4), 1086-1097. <https://doi.org/10.1002/tesq.3006>
- McCrocklin, S., Humaidan, A., & Edalatshams, I. (2019). ASR dictation program accuracy: Have current programs improved? In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference* (pp. 191-200). Iowa State University.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134-140. [https://doi.org/10.1016/s1364-6613\(03\)00028-7](https://doi.org/10.1016/s1364-6613(03)00028-7)
- Moon, C. M., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatrica, International Journal of Paediatrics*, 102(2), 156-160. <https://doi.org/10.1111/apa.12098>
- Moorman, C. M. (2017). *Individual differences and linguistic factors in the development of mid vowels in L2 Spanish learners: A longitudinal study* [Doctoral dissertation, Georgetown University].
- Mora, J. C., & Darcy, I. (2016). The relationship between cognitive control and pronunciation in a second language. In T. Isaacs & P. Trofimovich (Eds.), *Second Language Pronunciation Assessment - Interdisciplinary Perspectives* (pp. 95-120). Multilingual Matters. <https://doi.org/10.1016/b978-0-08-088583-4.50007-8>
- Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-Garcia, C. (2022). Training the pronunciation of L2 vowels under different conditions: the use of non-lexical materials and masking noise. *Phonetica*, 79(1), 1-43. <https://doi.org/10.1515/phon-2022-2018>
- Motohashi-Saigo, M., & Hardison, D. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology*, 13(2), 29-47. <https://doi.org/10.125/44179>
- Moussalli, S., & Cardoso, W. (2016). Are commercial 'personal robots' ready for language learning? Focus on second language speech. In *CALL communities and culture – short papers from EUROCALL 2016* (pp. 325-329). <https://doi.org/10.14705/rpnet.2016.eurocall2016.583>
- Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, 51(3), 617-637. <https://doi.org/10.1111/flan.12348>
- Mroz, A. (2020). Aiming for advanced intelligibility and proficiency using mobile ASR. *Journal of Second Language Pronunciation*, 6(1), 12-38. <https://doi.org/10.1075/jslp.18030.mro>
- Munro, M. J., & Derwing, T. M. (2006a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73-97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>

- Munro, M. J., & Derwing, T. M. (2006b). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), 520-531. <https://doi.org/10.1016/j.system.2006.09.004>
- Munro, M. J., Derwing, T. M., & Saito, K. (2024). Aspiring to aspirate: L2 acquisition of English word-initial /p/ over 10 years. *Language and Speech*. <https://doi.org/10.1177/00238309241264296>
- Nagle, C. (2013). A reexamination of ultimate attainment in L2 phonology: Length of immersion, motivation, and phonological short-term memory. In E. Voss, S.-J. D. Tai, & Z. Li (Eds.), *Selected Proceedings of the 2011 Second Language Research Forum* (pp. 148-161). Cascadilla Proceedings Project.
- Nagle, C. L. (2022). Individual difference factors for second language pronunciation. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge Handbook of Second Language Acquisition and Individual Differences* (pp. 269-281). <https://doi.org/10.4324/9781003270546-23>
- Nagle, C. L., & Baese-Berk, M. M. (2021). Advancing the state of the art in L2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*, 44(2), 580-605. <https://doi.org/10.1017/s0272263121000371>
- Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning. *Studies in Second Language Acquisition*, 37(4), 677-711. <https://doi.org/10.1017/s0272263114000825>
- Neri, A., Cucchiarini, C., & Strik, H. (2006). ASR-based corrective feedback on pronunciation: Does it really work ? ISCA Interspeech, Pittsburgh, PA.
- Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, 21(5), 393-408. <https://doi.org/10.1080/09588220802447651>
- Newman, R. S. (2017). Communication development in infancy. In J. B. Gleason & N. B. Ratner (Eds.), *The Development of Language* (Ninth ed., pp. 26-44). Pearson.
- Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, 50(6), 1496-1509. [https://doi.org/10.1044/1092-4388\(2007/103\)](https://doi.org/10.1044/1092-4388(2007/103))
- Noguchi, M., Yamane, N., Tsuda, A., Kazama, M., Kim, B., & Gick, B. (2015, September). *Towards protocols for L2 pronunciation training using ultrasound imaging* [Poster presentation]. The 7th annual Pronunciation in Second Language Learning and Teaching conference, Texas A&M University.
- Nozawa, T. (2015). Effects of attention and training method on the identification of American English vowels and coda nasals. Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK.
- O'Brien, I., Segalowitz, N., Collentine, J. O. E., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, 27(3), 377-402. <https://doi.org/10.1017/s0142716406060322>
- O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, 29(04). <https://doi.org/10.1017/s027226310707043x>
- Ogden, R. (2017). *An Introduction to English Phonetics* (Second ed.). Edinburgh University Press.
- Okuno, T., & Hardison, D. M. (2016). Perception-production link in L2 Japanese vowel duration: Training with technology. *Language Learning and Technology*, 20(2), 61-80.
- Olson, D. J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. *Language Learning & Technology*, 18, 173-192. <https://doi.org/10.125/44389>

- Ortega, L. (2009). *Understanding Second Language Acquisition*. Taylor & Francis Group.
<http://ebookcentral.proquest.com/lib/anu/detail.action?docID=564558>
- Otsuka, T., & Ueda, H. (2012). Pronunciation instruction and its achievements in Japanese junior and senior high schools. *Journal of Osaka Jogakuin University*, 8, 1-27.
- Ottem, E. J., Lian, A., & Karlsen, P. J. (2007). Reasons for the growth of traditional memory span across age. *European Journal of Cognitive Psychology*, 19(2), 233-270.
<https://doi.org/10.1080/09541440600684653>
- Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, 130(1), 461-472.
<https://doi.org/10.1121/1.3593366>
- Plonsky, L. (2013). Study quality in SLA: An Assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, 35(4), 655-687. <https://doi.org/10.1017/s0272263113000399>
- Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal*, 100(2), 538-553.
<https://doi.org/10.1111/modl.12335>
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2), 421-435.
- Price, P. J. (1981). *A cross-linguistic study of flaps in Japanese and in American English* [Doctoral dissertation, University of Pennsylvania].
- Qian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning and Technology*, 22(1), 69-96. <https://doi.org/10.125/44582>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. In R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rato, A. (2014). Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. *Concordia Working Papers in Applied Linguistics*, 5, 529-546.
- Reber, S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855-863. [https://doi.org/10.1016/S0022-5371\(67\)80149-X](https://doi.org/10.1016/S0022-5371(67)80149-X)
- Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: evidence from Japanese geminates. *The Journal of the Acoustical Society of America*, 134(2), 1324-1335. <https://doi.org/10.1121/1.4812767>
- Safronova, E. (2016). *The role of cognitive ability in the acquisition of second language perceptual phonological competence* [Doctoral dissertation, University of Barcelona].
- Safronova, E., & Mora, J. C. (2012). Acoustic and phonological memory in L2 vowel perception. *Proceedings of 22nd EUROSLA*, 384-390.
http://www.aedean.org/pdf/atatimecrisis/Safronova_Mora_AEDEAN35.pdf
- Saito, K. (2011). Identifying problematic segmental features to acquire comprehensible pronunciation in EFL settings: The case of Japanese learners of English. *RELJ Journal*, 42(3), 363-378. <https://doi.org/10.1177/0033688211420275>
- Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, 46(4), 842-854.
<https://doi.org/10.1002/tesq.67>
- Saito, K., Hanzawa, K., Petrova, K., Kachlicka, M., Suzukida, Y., & Tierney, A. (2022). Incidental and multimodal high variability phonetic training: Potential, limits, and future directions. *Language Learning*. <https://doi.org/10.1111/lang.12503>

- Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, 62(2), 595-633. <https://doi.org/10.1111/j.1467-9922.2011.00639.x>
- Saito, K., & Munro, M. J. (2014). The early phase of /ɹ/ production development in adult Japanese learners of English. *Language and Speech*, 57(4), 451-469. <https://doi.org/10.1177/0023830913513206>
- Sakai, M. (2016). *(Dis)connecting perception and production: training adult native speakers of Spanish on the English /i-/ɪ/ distinction*
- Sakai, M., & Moorman, C. (2017). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1), 187-224. <https://doi.org/10.1017/s0142716417000418>
- Schwartz, R. G., & Leonard, L. B. (1982). Do children pick and choose? an examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, 9(2), 319-336. <https://doi.org/10.1017/s0305000900004748>
- Serafini, E. J., & Sanz, C. (2015). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, 38(4), 607-646. <https://doi.org/10.1017/s0272263115000327>
- Setter, J. (2019). *Your Voice Speaks Volumes: It's Not What You Say, But How You Say It*. Oxford University Press.
- Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3(3), 243-261. <https://doi.org/10.1017/S0142716400001417>
- Shinohara, Y. (2014). *Perceptual Training of English /r/ and /l/ for Japanese Adults, Adolescents and Children* [Doctoral dissertation, University College London].
- Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r-/l/. *Journal of Phonetics*, 66, 242-251. <https://doi.org/10.1016/j.wocn.2017.11.002>
- Shinohara, Y., & Iverson, P. (2021). The effect of age on English /r-/l/ perceptual training outcomes for Japanese speakers. *Journal of Phonetics*, 89. <https://doi.org/10.1016/j.wocn.2021.101108>
- Slevc, R. L., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, 17(8), 675-681. <https://doi.org/https://doi.org/10.1111/j.1467-9280.2006.01765.x>
- Sonu, M., Kato, H., Tajima, K., Akahane-Yamada, R., & Sagisaka, Y. (2013). Non-native perception and learning of the phonemic length contrast in spoken Japanese: training Korean listeners using words with geminate and singleton phonemes. *Journal of East Asian Linguistics*, 22(4), 373-398. <https://doi.org/10.1007/s10831-013-9107-1>
- Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, 25(2), 293-321. <https://doi.org/10.1017/s0142716404001146>
- Spring, R., & Tabuchi, R. (2021). Assessing the practicality of using an automatic speech recognition tool to teach English pronunciation online. *STEM Journal*, 22(2), 93-104. <https://doi.org/10.16875/stem.2021.22.2.93>
- Spring, R., & Tabuchi, R. (2022). The role of ASR training in EFL pronunciation improvement: An in-depth look at the impact of treatment length and guided practice on specific pronunciation points. *Computer Assisted Language Learning Electronic Journal (CALL-EJ)*, 23(3), 166-185.
- Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21(3), 291-311. [https://doi.org/10.1016/s0095-4470\(19\)31340-3](https://doi.org/10.1016/s0095-4470(19)31340-3)

- Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44(6), 1321-1337. [https://doi.org/https://doi.org/10.1044/1092-4388\(2001/103\)](https://doi.org/https://doi.org/10.1044/1092-4388(2001/103))
- Strange, W. (1992). Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In Y. i. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech Perception, Production and Linguistic Structure* (pp. 197-219). Ohmsa.
- Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, 36(2), 131-145. <https://doi.org/10.3758/BF03202673>
- Su, F., Xie, H., & Wang, F. L. (2021). A a comparative review of mobile and non-mobile games or language learning. *SAGE Open*, 11(4), 1-20. <https://doi.org/10.1177/21582440211067247>
- Tanaka, A., & Nakamura, K. (2004). Auditory memory and proficiency of second language speaking: A latent variable analysis approach. *Psychological Reports*, 95(3), 723-734. <https://doi.org/https://doi.org/10.2466/pr0.95.3.723-734>
- Tateishi, M., & Winters, S. (2015). Does ultrasound training lead to improved perception of a non-native sound contrast?: Evidence from Japanese learners of English. *Proceedings of the 2013 Annual Conference of the Canadian Linguistic Association*, 1-15.
- Tejedor-García, C., Escudero-Mancebo, D., González-Ferreras, C., Cámara-Arenas, E., & Cardenoso-Payo, V. (2017). Evaluating the efficiency of synthetic voice for providing corrective feedback in a pronunciation training tool based on minimal pairs. *7th ISCA Workshop on Speech and Language Technology in Education (SLaTE 2017)*, 26-30. <https://doi.org/10.21437/SLaTE.2017-5>
- Teshima, M. (2011). On teaching English pronunciation in Japanese secondary education: How it is and how it should be. *Journal of the Phonetics Society*, 15(1), 31-43.
- Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, 28(3), 744-765. <https://doi.org/10.11139/cj.28.3.744-765>
- Thomson, R. I. (2012a). *English Accent Coach*. <https://www.englishaccentcoach.com/home>
- Thomson, R. I. (2012b). Improving L2 Listeners' Perception of English Vowels: A Computer-Mediated Approach. *Language Learning*, 62(4), 1231-1258. <https://doi.org/10.1111/j.1467-9922.2012.00724.x>
- Thomson, R. I. (2024). *English Accent Coach Asteroids*. <https://apps.apple.com/ca/app/english-accent-coach-asteroids/id1607018351>
- Thomson, R. I., & Derwing, T. M. (2014). The Effectiveness of L2 Pronunciation Instruction: A Narrative Review. *Applied Linguistics*, 36(3), 326-344. <https://doi.org/10.1093/applin/amu076>
- Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference, Dallas, TX, October 2015* (pp. 88-97). Iowa State University.
- Thorin, J. (2020). *Can you hear what you cannot say? The interactions of speech perception and production during non-native phoneme learning* [Doctoral dissertation, Radboud University].
- Thorin, J., Sadakata, M., Desain, P., & McQueen, J. M. (2018). Perception and production in interaction during non-native speech category learning. *Journal of the Acoustical Society of America*, 144(1), 92-103. <https://doi.org/10.1121/1.5044415>
- Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97(2), B25-B34. <https://doi.org/10.1016/j.cognition.2005.01.006>
- Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 47(2), 466-472.

- Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2014). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, 19(1), 122-140. <https://doi.org/10.1017/s1366728914000832>
- Tsalikis, J., DeShields, O. W., & Latour, M. S. (1991). The role of accent on the credibility and effectiveness of the salesperson. *The Journal of Personal Selling and Sales Management*, 11(1), 31-41. <https://www.jstor.org/stable/40471845>
- Tsui, H. M.-L. (2012). *Ultrasound speech training for Japanese adults learning English as a second language* [Master's thesis, The University of British Columbia].
- Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. T. (1994). Discrimination of English /r-/ and /w-y/ by Japanese infants at 6-12 months: Language-specific developmental changes in speech perception abilities. *The Emergence of Human Cognition and Language*, 3, 57-61.
- Uchida, Y., & Sugimoto, J. (2018). A survey of pronunciation instruction by Japanese teachers of English: Phonetic knowledge and teaching practice. *Journal of the Tokyo University of Marine Science and Technology*, 14, 65-75.
- Ueda, R., & Hashimoto, K. (2019). Perceptual training in a classroom setting: Phonemic category formation by Japanese EFL learners. In J. M. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference, Iowa State University, September 2018* (pp. 237-249).
- van de Weijer, J. (1999). *Language input for word discovery* [Doctoral dissertation, Radboud University].
- van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology*, 6, Article 1000. <https://doi.org/10.3389/fpsyg.2015.01000>
- van Lier, L. (1997). Observation from an ecological perspective. *TESOL Quarterly*, 31(4). <https://doi.org/10.2307/3587762>
- Vance, T. J. (2008). *The Sounds of Japanese with Audio CD*. Cambridge University Press.
- Venkatagiri, H. S., & Levis, J. M. (2007). Phonological awareness and speech comprehensibility: An exploratory study. *Language Awareness*, 16(4), 263-277. <https://doi.org/10.2167/la417.0>
- Vihman, M. M., DePaolis, R. A., & Keren-Portnoy, T. (2014). The role of production in infant word learning. *Language Learning*, 64(s2), 121-140. <https://doi.org/10.1111/lang.12058>
- Vlahou, E. L., Protopapas, A., & Seitz, A. R. (2012). Implicit training of nonnative speech stimuli. *Journal of Experimental Psychology: General*, 141(2), 363-381. <https://doi.org/10.1037/a0025014>
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *The Journal of Educational Research*, 84(1), 30-43. <https://www.jstor.org/stable/40539680>
- Wang, X., & Munro, M. J. (1999). The perception of English tense-lax vowel by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 125-128).
- Wang, Y. H., & Young, S. S. C. (2015). Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system. *Journal of Computer Assisted Learning*, 31(6), 493-504. <https://doi.org/10.1111/jcal.12079>
- Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, 36(2), 250-267. <https://doi.org/10.1016/j.wocn.2007.06.004>
- Wee, D. T. J., Grenon, I., Sheppard, C., & Archibald, J. (2019). Identification and discrimination training yield comparable results for contrasting vowels. In *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS)* (pp. 939-943).

- Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, 24(5), 672-683.
<https://doi.org/10.1037/0012-1649.24.5.672>
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49-63.
[https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- Wong, J. W. S. (2014). The effects of high and low variability phonetic training on the perception and production of English vowels /e/-/æ/ by Cantonese ESL learners with high and low L2 proficiency levels. *Proceedings of the 15th Annual Conference of the International Speech Communication Association*, 524-528.
- Wright, J. (2021). *Factors affecting the incidental formation of novel suprasegmental categories* [Doctoral dissertation, University of Oregon].
- Yamada, R. A., Strange, W., Magnuson, J. S., Pruitt, J. S., & Clarke, W. D. (1994). The intelligibility of Japanese speakers' production of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. ICSLP-1994, Yokohama, Japan.
- Yamane, N., Abel, J., Allen, B., Burton, S., Kazama, M., Noguchi, M., Tsuda, A., & Gick, B. (2015). Ultrasound-integrated pronunciation teaching and learning. Ultrafest VII, The University of Hong Kong.
- Zahler, S., & Lord, G. (2022). The role of phonological short-term memory in second language phonology. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 35(1), 173-205. <https://doi.org/10.1075/resla.20003.zah>
- Zamuner, T. S., Strahm, S., Morin-Lessard, E., & Page, M. P. A. (2017). Reverse production effect: Children recognize novel words better when they are heard rather than produced. *Developmental Science*, 21(4), e12636. <https://doi.org/10.1111/desc.12636>
- Zamuner, T. S., & Thiessen, A. (2018). A phonological, lexical, and phonetic analysis of the new words that young children imitate. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 63(4), 609-632. <https://doi.org/10.1017/cnj.2018.10>
- Zamuner, T. S., Yeung, H. H., & Ducos, M. (2017). The many facets of speech production and its complex effects on phonological processing. *British Journal of Psychological Society*, 108(1), 37-39. <https://doi.org/10.1111/bjop.12220>
- Ziegler, J. C., Besson, M., Jacobs, A. M., Nazir, T. A., & Carr, T. H. (1997). Word, pseudoword, and nonword processing: A multitask comparison using event-related brain potentials. *Journal of Cognitive Neuroscience*, 9(6), 758-775.

Appendices

Appendix A: Stimulus List of the Discrimination and Identification Tasks

	Testing and Training		Practice	
	/l/	/r/		
Pre- and post-tests, Training	lush	rush	sink	pink
	laugh	rough	book	hook
	lied	ride	boy	toy
	lice	rice	test	nest
	leach	reach	night	white
	leaf	reef		
	lope	rope		
	liver	river		
	loom	room		
	lude	rude		
	lace	race		
	lane	rain		
	lent	rent		
	load	road		
	loan	roan		
	lobe	robe		
	lock	rock		
	lot	rot		
	late	rate		
	loyal	royal		
Generalisation to new words	light	right		
	loot	root		
	lake	rake		
	link	rink		
	low	row		

Appendix B: Stimulus List of the PSTM tasks

Task 1									
Practice	にふえ	ni-fu-e							
Practice	かでせ	ka-de-se	しあめ	shi-a-me					
1	げのて	ge-no-te	みたじ	mi-ta-ji					
2	にはえ	ni-ha-e	がてせ	ga-te-se	もかね	mo-ka-ne			
3	ほとき	ho-to-ki	つれむ	tsu-re-mu	こねい	ko-ne-i			
4	ひつら	hi-tsu-ra	すけね	su-ke-ne	れきあ	re-ki-a			
5	てあふ	te-a-fu	へてず	he-te-zu	むみほ	mu-mi-ho			
6	こへに	ko-he-ni	ぬたよ	nu-ta-yo	めちな	me-chi-na			
7	のほえ	no-ho-e	らのの	ra-re-no	ふらま	fu-ra-ma			
8	そてせ	so-te-se	ちうふ	chi-u-fu	まそあ	ma-so-a	ぞわひ	zo-wa-hi	
9	ぬろく	nu-ro-ku	わえふ	wa-e-fu	ろこよ	ro-ko-yo	れのき	re-no-ki	
10	もらま	mo-ra-ma	つふぬ	tsu-fu-nu	ほせね	ho-se-ne	れらと	re-ra-to	
11	れきみ	re-ki-mi	ふよま	fu-yo-ma	みけへ	mi-ke-he	れふか	re-fu-ka	
12	なてす	na-te-su	ぜほけ	ze-ho-ke	れちう	re-chi-u	てなせ	te-na-se	
13	てさや	te-sa-ya	にみか	ni-mi-ka	けかゆ	ke-ka-yu	めてか	me-te-ka	
14	しえめ	shi-e-me	よつろ	yo-tsu-ro	のへえ	no-he-e	まにむ	ma-ni-mu	
15	まこぞ	ma-ko-zo	きでや	ki-de-ya	よみと	yo-mi-to	なおて	na-o-te	
16	るひに	ru-hi-ni	らあろ	ra-a-ro	はひね	ha-hi-ne	まへた	ma-he-ta	すちと su-chi-to
17	せむて	se-mu-te	ぎはと	gi-ha-to	かろぎ	ka-ro-gi	にけえ	ni-ke-e	ゆにや yu-ni-ya

Task 2									
Practice	かぶた	ka-bu-ta							
Practice	じたに	ji-ta-ni	びゆて	bi-yu-te					
1	むたり	mu-ta-ri	こたお	ko-ta-o					
2	ぎるこ	gi-ru-ko	ふたぼ	fu-ta-bo	じゆき	ji-yu-ki			
3	りおに	ri-o-ni	やくら	ya-ku-ra	しあれ	shi-a-re			
4	こむせ	ko-mu-se	たれい	ra-re-i	ゆうせ	yu-u-se			
5	ほとぶ	ho-to-bu	ちえま	chi-e-ma	ふえこ	fu-e-ko			
6	ひふに	hi-fu-ni	しみむ	shi-mi-mu	やとせ	ya-to-se			
7	ぶあれ	bu-a-re	こふい	ko-fu-i	とれで	to-re-de			
8	すとえ	su-to-e	さじめ	sa-ji-me	どふら	do-fu-ta	せたう	se-ta-u	
9	にれく	ni-re-ku	もによ	mo-ni-yo	きため	ki-ta-me	じとす	ji-to-su	
10	うえみ	u-e-mi	こにさ	ko-ni-sa	いがじ	i-ka-ji	こみか	ko-mi-ka	
11	いえふ	i-e-fu	えおし	e-o-shi	しろて	shi-ro-te	にえひ	ni-e-hi	
12	こりぐ	ko-ri-gu	だじり	da-ji-ri	ろこあ	ro-ko-a	とにこ	to-ni-ko	
13	えてし	e-te-shi	ひぶき	hi-bu-ki	じてさ	ji-te-sa	とえぎ	to-e-gi	
14	これぜ	ko-re-ze	りしえ	ri-shi-e	あぞみ	a-zo-mi	ゆたえ	yu-ta-e	
15	むおし	mu-o-shi	ひてび	hi-te-bi	ふたお	fu-ta-o	ぎおた	gi-o-ta	
16	れして	re-shi-te	おむた	o-mu-ta	すおど	du-o-do	こてみ	ko-te-mi	あとす a-to-su
17	うみら	u-mi-ra	いけり	i-ke-ri	おてか	o-te-ka	むふあ	mu-fu-a	のまこ no-ma-ko

Appendix C: Questions on the Linguistic Background Questionnaire

- Gender
- Age
- Is your first language Japanese?
- At what age did you start learning English? (Including English lessons outside of school)
- Have you lived or stayed in a foreign country for more than three months? If so, in which country, for how long, and at what age?
- Do you speak any other languages besides Japanese and English? If so, please name the language(s) and the proficiency.

Appendix D: Picture Description Task (Expected descriptions and keywords)

The pictures used in the task are not shared due to copyright restrictions.

Practice

1. The family is climbing the mountain. (family / mountain)
2. The man made a curry, but it ended up tasting bad. (curry / taste 味がする)
3. The woman was using a smartphone until 1 am last night, so she fell asleep at work today.
(woman / smartphone / work)

Test

1. The student came to class late three days in a row. (student / late / in a row 続けて)
2. The birth rate in Japan is low. (low / birth rate 出生率 / Japan)
3. The teacher needs to turn off the light and lock the door. (lock / need / light)
4. The girl sits in the front row. (girl / front row 最前列)
5. The horse won the race. (race / won)
6. The woman bought a lace curtain. (lace curtain / bought)
7. Cars drive in the right lane. (right / lane / drive)
8. The woman is singing in the rain. (in the rain / singing)
9. Her mother lent her daughter some money. (daughter / lent 貸した)
10. The woman goes to the CD shop to rent a rock CD. (rock / CD / rent 借りる)

Appendix E: Sentence Reading Task

Practice 1	sink	She put the dishes in the sink.
	pink	She bought some pink dishes.
Practice 2	book	There were two books on the desk.
	hook	There were two hooks on the desk.
Practice 3	boy	The boy was cute.
	toy	The toy was cute.
Practice 4	test	The teacher made a test.
	nest	The bird made a nest.
Practice 5	white	I have got a white gown.
	night	I have got a night gown.
Generalisation 1	lead	This path will lead to the university.
	read	The man will read the newspaper today.
Generalisation 2	light	Turn right at the next corner.
	right	Turn on the light, please.
Generalisation 3	low	His voice is very low.
	row	I always sit in the front row.
Generalisation 4	link	I will send you the link to our website.
	rink	I went to a skate rink yesterday.
Generalisation 5	lace	You have to lace up your shoes.
	race	It is a race against time.
Trained words 1	late	He usually comes home late.
	rate	She works at her own rate.
Trained words 2	lied	The boy lied to his mother.
	ride	The girl needs a ride to school.
Trained words 3	liver	They do not eat beef liver.
	river	Do not go close to that river.
Trained words 4	lane	My friend will swim in lane two.
	rain	I got caught in a rain storm on my way home.
Trained words 5	lent	The bank lent money to the company.
	rent	The company pays the rent for the place.
Trained words 6	load	The truck was carrying a heavy load.
	road	The truck was travelling on a busy road.
Trained words 7	lock	Teachers have to lock their office door.
	rock	The teacher likes listening to rock music.
Trained words 8	loyal	He is loyal to his family.
	royal	He is from a royal family.

Appendix F: Questions on the Exit Questionnaire

- I want to improve my English skills. (scale of 1-10)
- I want to improve my English pronunciation. (scale of 1-10)
- The training task was effective in improving my English pronunciation. (scale of 1-10)
- →The reasons. (open-ended answer)

- I want to continue using the training task to improve my English pronunciation. (scale of 1-10)
- →The reasons. (open-ended answer)
- Your thoughts/comments on the training. (open-ended answer).

Appendix G: [Experiment 3] Model Output of a Logistic Mixed Effects Model for the Discrimination Task

<i>Predictors</i>	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	3.54 ***	2.39 – 5.24	< 0.001
time [Post-test]	1.63 **	1.16 – 2.29	0.005
group [MIX]	0.88	0.55 – 1.42	0.606
group [PROD-ONLY]	1.20	0.74 – 1.96	0.454
group [PERC-PROD]	1.23	0.75 – 2.01	0.421
stimuli [TWUS]	0.87	0.63 – 1.19	0.376
stimuli [UWTS]	0.73	0.46 – 1.15	0.178
stimuli [UWUS]	0.57 *	0.36 – 0.90	0.015
time [Post-test] × group [MIX]	1.37	0.85 – 2.21	0.201
time [Post-test] × group [PROD-ONLY]	0.70	0.43 – 1.13	0.146
time [Post-test] × group [PERC-PROD]	1.01	0.61 – 1.67	0.973
time [Post-test] × stimuli [TWUS]	0.65	0.41 – 1.03	0.070
time [Post-test] × stimuli [UWTS]	0.72	0.44 – 1.18	0.195
time [Post-test] × stimuli [UWUS]	0.72	0.44 – 1.18	0.190
group [MIX] × stimuli [TWUS]	1.27	0.82 – 1.97	0.283
group [PROD-ONLY] × stimuli [TWUS]	0.75	0.48 – 1.17	0.203
group [PERC-PROD] × stimuli [TWUS]	0.77	0.49 – 1.22	0.265
group [MIX] × stimuli [UWTS]	1.48	0.91 – 2.40	0.111
group [PROD-ONLY] × stimuli [UWTS]	0.80	0.49 – 1.30	0.360

group [PERC-PROD] × stimuli [UWTS]	1.27	0.76 – 2.11	0.365
group [MIX] × stimuli [UWUS]	1.46	0.91 – 2.34	0.115
group [PROD-ONLY] × stimuli [UWUS]	0.97	0.60 – 1.57	0.902
group [PERC-PROD] × stimuli [UWUS]	0.82	0.50 – 1.34	0.431
(time [Post-test] × group [MIX]) × stimuli [TWUS]	0.64	0.33 – 1.22	0.174
(time [Post-test] × group [PROD-ONLY]) × stimuli [TWUS]	2.13 *	1.10 – 4.09	0.024
(time [Post-test] × group [PERC-PROD]) × stimuli [TWUS]	1.41	0.72 – 2.79	0.319
(time [Post-test] × group [MIX]) × stimuli [UWTS]	0.84	0.41 – 1.72	0.628
(time [Post-test] × group [PROD-ONLY]) × stimuli [UWTS]	2.25 *	1.10 – 4.61	0.027
(time [Post-test] × group [PERC-PROD]) × stimuli [UWTS]	0.86	0.41 – 1.83	0.705
(time [Post-test] × group [MIX]) × stimuli [UWUS]	0.70	0.35 – 1.40	0.315
(time [Post-test] × group [PROD-ONLY]) × stimuli [UWUS]	1.34	0.67 – 2.69	0.406
(time [Post-test] × group [PERC-PROD]) × stimuli [UWUS]	1.13	0.55 – 2.31	0.748

Random Effects

σ^2	3.29
τ_{00} participant	0.25
τ_{00} word	0.29
ICC	0.14
$N_{\text{participant}}$	60
N_{word}	50
Observations	12000
Marginal R^2 / Conditional R^2	0.026 / 0.164

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Appendix H: [Experiment 3] Model Output of a Logistic Mixed Effects Model for the Identification Task

<i>Predictors</i>	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	1.38	0.98 – 1.94	0.063
time [Post-test]	1.17	0.89 – 1.55	0.256

group [MIX]	1.08	0.71 – 1.64	0.713
group [PROD-ONLY]	1.18	0.77 – 1.80	0.442
group [PERC-PROD]	1.09	0.71 – 1.68	0.693
stimuli [TWUS]	0.85	0.64 – 1.11	0.234
stimuli [UWTS]	0.88	0.59 – 1.30	0.512
stimuli [UWUS]	1.00	0.67 – 1.49	0.986
time [Post-test] × group [MIX]	1.39	0.94 – 2.06	0.100
time [Post-test] × group [PROD-ONLY]	1.25	0.84 – 1.86	0.271
time [Post-test] × group [PERC-PROD]	1.15	0.77 – 1.71	0.505
time [Post-test] × stimuli [TWUS]	1.12	0.76 – 1.66	0.561
time [Post-test] × stimuli [UWTS]	1.23	0.79 – 1.90	0.354
time [Post-test] × stimuli [UWUS]	0.82	0.53 – 1.26	0.357
group [MIX] × stimuli [TWUS]	1.02	0.70 – 1.50	0.902
group [PROD-ONLY] × stimuli [TWUS]	1.05	0.71 – 1.55	0.823
group [PERC-PROD] × stimuli [TWUS]	1.45	0.97 – 2.16	0.066
group [MIX] × stimuli [UWTS]	1.45	0.94 – 2.22	0.094
group [PROD-ONLY] × stimuli [UWTS]	1.21	0.78 – 1.87	0.389
group [PERC-PROD] × stimuli [UWTS]	1.07	0.69 – 1.66	0.772
group [MIX] × stimuli [UWUS]	0.93	0.60 – 1.42	0.731
group [PROD-ONLY] × stimuli [UWUS]	0.77	0.50 – 1.19	0.234
group [PERC-PROD] × stimuli [UWUS]	1.03	0.66 – 1.60	0.897
(time [Post-test] × group [MIX]) × stimuli [TWUS]	0.79	0.46 – 1.38	0.413
(time [Post-test] × group [PROD-ONLY]) × stimuli [TWUS]	0.80	0.46 – 1.40	0.436
(time [Post-test] × group [PERC-PROD]) × stimuli [TWUS]	0.65	0.37 – 1.16	0.144
(time [Post-test] × group [MIX]) × stimuli [UWTS]	0.50 *	0.27 – 0.92	0.027
(time [Post-test] × group [PROD-ONLY]) × stimuli [UWTS]	0.70	0.37 – 1.31	0.266
(time [Post-test] × group [PERC-PROD]) × stimuli [UWTS]	0.78	0.42 – 1.46	0.439
(time [Post-test] × group [MIX]) × stimuli [UWUS]	0.86	0.47 – 1.59	0.640

(time [Post-test] × group [PROD-ONLY]) × stimuli [UWUS]	1.07	0.58 – 1.99	0.825
(time [Post-test] × group [PERC-PROD]) × stimuli [UWUS]	0.87	0.46 – 1.63	0.661

Random Effects

σ^2	3.29
τ_{00} participant	0.20
τ_{00} word	0.21
ICC	0.11
N _{participant}	60
N _{word}	50
Observations	12000
Marginal R ² / Conditional R ²	0.010 / 0.119
* <i>p</i> <0.05 ** <i>p</i> <0.01 *** <i>p</i> <0.001	

Appendix I: [Experiment 3] Model Output of a Logistic Mixed Effects Model for Identification Rating Scores

<i>Predictors</i>	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	1.37	0.84 – 2.23	0.202
group [MIX]	1.28	0.69 – 2.39	0.434
group [PROD-ONLY]	1.52	0.80 – 2.88	0.199
group [PERC-PROD]	1.45	0.76 – 2.78	0.260
time [Post-test]	1.01	0.66 – 1.54	0.964
type [Untrained]	0.88	0.57 – 1.35	0.563
task [Sentence]	1.22	0.85 – 1.77	0.284
task [Picture]	1.01	0.70 – 1.46	0.969
group [MIX] × time [Post-test]	1.86 *	1.11 – 3.13	0.018
group [PROD-ONLY] × time [Post-test]	1.79 *	1.05 – 3.04	0.031
group [PERC-PROD] × time [Post-test]	1.88 *	1.10 – 3.22	0.021
group [MIX] × type [Untrained]	1.20	0.71 – 2.02	0.494
group [PROD-ONLY] × type [Untrained]	1.19	0.70 – 2.03	0.519
group [PERC-PROD] × type [Untrained]	1.14	0.66 – 1.95	0.638

time [Post-test] × type [Untrained]	1.56	0.91 – 2.66	0.106
group [MIX] × task [Sentence]	0.57 *	0.37 – 0.90	0.016
group [PROD-ONLY] × task [Sentence]	0.42 ***	0.27 – 0.68	<0.001
group [PERC-PROD] × task [Sentence]	0.53 **	0.33 – 0.85	0.008
group [MIX] × task [Picture]	0.59 *	0.37 – 0.92	0.021
group [PROD-ONLY] × task [Picture]	0.52 **	0.33 – 0.83	0.006
group [PERC-PROD] × task [Picture]	0.55 *	0.35 – 0.89	0.014
time [Post-test] × task [Sentence]	0.70 *	0.50 – 0.96	0.028
time [Post-test] × task [Picture]	0.80	0.58 – 1.11	0.189
(group [MIX] × time [Post-test]) × type [Untrained]	0.46 *	0.22 – 0.97	0.040
(group [PROD-ONLY] × time [Post-test]) × type [Untrained]	0.61	0.29 – 1.30	0.200
(group [PERC-PROD] × time [Post-test]) × type [Untrained]	0.62	0.29 – 1.33	0.220

Random Effects

σ^2	3.29
τ_{00} participant	0.36
τ_{00} word	0.03
ICC	0.11
N participant	59
N word	12
Observations	4072
Marginal R ² / Conditional R ²	0.033 / 0.135

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Appendix J: [Experiment 3] Model Output of a Linear Mixed Effects Model for Goodness Rating Scores

<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	5.81 ***	5.23 – 6.39	<0.001
group [MIX]	0.25	-0.41 – 0.92	0.453
group [PROD-ONLY]	0.40	-0.27 – 1.07	0.245
group [PERC-PROD]	0.73 *	0.05 – 1.42	0.036

time [Post-test]	0.29	-0.08 – 0.67	0.125
task [Sentence]	0.06	-0.36 – 0.48	0.786
task [Picture]	0.16	-0.26 – 0.59	0.455
group [MIX] × time [Post-test]	0.11	-0.31 – 0.53	0.611
group [PROD-ONLY] × time [Post-test]	0.44 *	0.01 – 0.87	0.045
group [PERC-PROD] × time [Post-test]	0.37	-0.06 – 0.80	0.094
group [MIX] × task [Sentence]	-0.51	-1.02 – 0.00	0.051
group [PROD-ONLY] × task [Sentence]	-0.74 **	-1.27 – -0.22	0.005
group [PERC-PROD] × task [Sentence]	-0.99 ***	-1.51 – -0.46	<0.001
group [MIX] × task [Picture]	-0.66 *	-1.18 – -0.15	0.012
group [PROD-ONLY] × task [Picture]	-0.82 **	-1.34 – -0.29	0.002
group [PERC-PROD] × task [Picture]	-0.97 ***	-1.50 – -0.43	<0.001
time [Post-test] × task [Sentence]	-0.27	-0.63 – 0.10	0.152
time [Post-test] × task [Picture]	-0.39 *	-0.76 – -0.03	0.034

Random Effects

σ^2	5.88
τ_{00} participant	0.50
τ_{00} word	0.27
ICC	0.12
N participant	59
N word	12
Observations	4072
Marginal R ² / Conditional R ²	0.026 / 0.140

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$