

**SENSORIMOTOR PROCESSING IN SPEECH  
EXAMINED IN AUTOMATIC IMITATION TASKS**

Yuchunzi Wu

Department of Speech, Hearing and Phonetic Sciences

University College London

October 2020

A thesis submitted for the degree of

Doctor of Philosophy

## **Declaration**

I, Yuchunzi Wu, confirm that the work presented in this thesis is on my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Yuchunzi Wu

## **Abstract**

The origin of humans' imitative capacity to quickly map observed actions onto their motor repertoire has been the source of much debate in cognitive psychology. Past research has provided a comprehensive account of how sensorimotor associative experience forges and modulates the imitative capacity underlying familiar, visually transparent manual gestures. Yet, little is known about whether the same associative mechanism is also involved in imitation of visually opaque orofacial movements or novel actions that were not part of the observers' motor repertoire. This thesis aims to establish the role of sensorimotor experience in modulating the imitative capacity underlying communicative orofacial movements, namely speech actions, that are either familiar or novel to perceivers.

Chapter 3 first establishes that automatic imitation of speech occurs due to perception-induced motor activation and thus can be used as a behavioural measure to index the imitative capacity underlying speech. Chapter 4 demonstrates that the flexibility observed for the imitative capacity underlying manual gestures extends to the imitative capacity underlying visually perceived speech actions, suggesting that the associative mechanism is also involved in imitation of visually opaque orofacial movements. Chapter 5 further shows that sensorimotor experience with novel speech actions modulates the imitative capacity underlying both novel and familiar speech actions produced using the same articulators. Thus, findings from Chapter 5 suggest that the associative mechanism is also involved in imitation of novel actions and that experience-induced modification probably occurs at the feature level in the perception-production link presumably underlying the imitative capacity.

Results are discussed with respect to previous imitation research and more general action-perception research in cognitive and experimental psychology, sensorimotor interaction studies in speech science, and native versus non-native processing in second language research. Overall, it is concluded that the development of speech imitation follows the same basic associative learning rules as the development of imitation in other effector systems.

## Impact Statement

People tend to imitate their interaction partners' actions in everyday social activity and imitate a wide range of actions they perceive. Researchers have been trying to answer questions like why we imitate, how we are able to imitate, and when we tend to imitate more/less in certain situations. In addition, researchers have also been trying to measure people's imitative behaviours under various circumstances, e.g., relatively naturalistic social situations or tightly controlled laboratory settings. Prior research has mostly examined imitation of manual gestures. The work presented in this thesis investigates imitation of speech actions to thus explore commonalities between different effector systems.

Firstly, automatic imitation effects provide a way to measure the imitative tendency in tightly controlled laboratory settings. Experiment 1 and 2 presented in Chapter 3 are among the first to systematically investigate speech automatic imitation effects in different task designs. Automatic imitation effects are measured using stimulus response compatibility (SRC) tasks, in which participants have to respond to a task-relevant prompt and ignore a task-irrelevant distractor that may interfere with their response. This task paradigm that has been used in the past few decades to investigate more general cognitive functions in experimental psychology. Thus, SRC tasks provide a way to link speech processing to more general cognitive functions. Additionally, Chapter 3 explores the possibility of using SRC tasks to behaviourally examine multimodal speech processing (i.e., audiovisual versus sensorimotor), which may potentially complement neural research of speech. There are not many tasks in speech science that can be used to examine pre-lexical audiovisual and/or sensorimotor interactions, and the existing ones are limited in its generalisability to a wider range of speech sounds or offering possible comparisons across different effector systems. Therefore, experiments presented in Chapter 3 help broaden research in speech science by methodologically offering an approach to examine multimodal speech processing and theoretically linking speech to more general cognitive functions.

More importantly, investigating the origin of speech imitation provides a different scenario than that of hand imitation, thus advancing the current imitation research. Experiments presented in Chapter 4 and 5 are among the first to investigate the origin of communicative orofacial movements, namely speech actions. Compared to simple manual gestures, speech actions are multimodal in nature and are more frequently used to convey communicative intentions. Therefore, findings reported in Chapter 4 and 5 not only extend current imitation research to a different effector system, but also offer a way to investigate imitation through different modalities, i.e., auditory, visual, or audiovisual. Furthermore, due to its communicative nature, investigating the mechanism underlying speech imitation may contribute to our understanding of how and why people imitate when they try to communicate

with others; or, how communicative intentions potentially modulate people's imitative behaviours that are prevalent in everyday social activity and online interactions. It is believed that the results of this thesis can contribute to our pre-existing knowledge in imitation research and in speech processing.

## Table of Contents

Declaration.....	2
Abstract.....	3
Impact Statement .....	4
Table of Contents.....	6
List of Figures.....	9
List of Tables .....	13
List of Abbreviations .....	16
Acknowledgments .....	17
Chapter 1 Introduction.....	19
Chapter 2 Literature review .....	23
2.1 Imitation of manual gestures .....	23
2.1.1 Automatic imitation and spatial compatibility .....	23
2.1.2 Proposed neural substrate underlying imitation .....	25
2.1.3 The origin of the imitative capacity .....	27
2.1.4 Sensorimotor training .....	31
2.2 Imitation of speech actions.....	34
2.2.1 Motor involvement in speech perception .....	34
2.2.2 Theories and models of speech perception .....	38
2.2.3 Sensory feedback in speech production .....	41
2.3 Summary and the thesis outline .....	44
Chapter 3 The perceptual and the motor processes triggered by visual distractors in automatic imitation of speech.....	47
3.1 Introduction .....	47
3.1.1 Stimulus-response and stimulus-stimulus compatibility.....	48
3.1.2 The multimodal nature of speech.....	50
3.1.3 Time courses of compatibility effects .....	51
3.1.4 The current study.....	55
3.2 Experiment 1 .....	56
3.2.1 Method .....	57
3.2.2 Results .....	61
3.2.3 Discussion .....	65

3.3	Experiment 2 .....	67
3.3.1	Method.....	68
3.3.2	Results .....	75
3.3.3	Discussion.....	85
3.4	General discussion .....	89
3.4.1	Time courses of compatibility effects .....	91
3.4.2	Automatic imitation in cross-modal settings.....	92
3.4.3	Audiovisual speech in stimulus-stimulus compatibility effects .....	93
3.4.4	Visual modulations on auditory speech.....	94
3.5	Conclusion.....	95
Chapter 4	Sensorimotor training effects on automatic imitation of native speech.....	98
4.1	Introduction .....	98
4.1.1	Expertise effects .....	99
4.1.2	Sensorimotor training .....	100
4.1.3	The current study.....	102
4.2	Method .....	103
4.3	Results .....	107
4.3.1	Testing sessions .....	107
4.3.2	Training sessions .....	116
4.4	Discussion .....	117
4.4.1	Mirror and counter-mirror training.....	118
4.4.2	Counter-mirror training and visual speech.....	118
4.4.3	Mirror training and the simulation theory of speech perception .....	121
4.4.4	Second language research and imitation research.....	122
4.5	Conclusion.....	123
Chapter 5	Sensorimotor training effects on automatic imitation of non-native speech ..	126
5.1	Introduction .....	126
5.1.1	Imitation learning .....	128
5.1.2	Novel speech actions .....	132
5.1.3	The current study.....	136
5.2	Pilot study.....	137
5.3	Method .....	144
5.4	Results .....	152
5.4.1	Testing sessions .....	152

5.4.2	Testing sessions – syllable effects.....	163
5.4.3	Training sessions.....	169
5.5	Discussion .....	170
5.5.1	Automatic imitation of non-native speech actions.....	171
5.5.2	Sensorimotor training with non-native speech actions .....	176
5.6	Conclusion.....	181
Chapter 6	General discussion .....	184
6.1	Summary of current research findings .....	184
6.2	Limitations, implications, and future directions.....	185
6.2.1	Compatibility effects and speech processing .....	185
6.2.2	The associative account of imitation and its future development.....	187
6.2.3	Speech processing examined in the context of cognitive psychology .....	191
6.3	Conclusion.....	195
	Bibliography .....	197
	Appendices .....	215
Appendix A	Experiment 3 .....	215
Method.....		215
Results .....		216
Discussion.....		223
Appendix B	Further analysis of Experiment 2 and 3 .....	225
Results .....		225
Discussion.....		239
Appendix C	Table C.1 .....	240

## List of Figures

Figure 3.1 a Figure 7 from Kornblum et al. (1999). Demonstrations of how SSC (left) and SRC (right) effects change across SOAs. Light and dark grey rectangles were further added to respectively mark the stimulus-identification and the response-production stages proposed in the Dimensional Overlap model. b Illustrations of time courses of SSC (left) and SRC (right) effects reported in Experiment 1 by Kornblum et al. and SSC+SRC (right) effects in Experiment 2 also by Kornblum et al. ....	52
Figure 3.2 a Schematic timeline of a trial presented in the syllable sessions in Experiment 1. The visual prompt <i>ba</i> is positioned extending from the speaker's bottom lip to her top lip. b The Experiment 1 procedure with a symbol and a syllable SRC task session. The order of two sessions was counterbalanced across participants. ....	57
Figure 3.3 a ERs $\pm$ standard errors (SEs) in each experimental condition. Two panels represent the symbol (left) and the syllable (right) sessions where SRC and SSC+SRC effects were expected respectively. b ER compatibility effects (i.e., incompatible - compatible) $\pm$ SEs at each SOA for each prompt type.....	61
Figure 3.4 a Mean RTs $\pm$ SEs in each experimental condition. Two panels represent the symbol (left) and the syllable (right) sessions where SRC and SSC+SRC effects were expected respectively. b RT compatibility effects $\pm$ SEs at each SOA for each prompt type.....	63
Figure 3.5 Mean RTs $\pm$ SEs for each prompt type in each compatibility condition. ....	64
Figure 3.6 The study design of Experiment 2 with two response modalities and two prompt types that factorially comprised four tasks. In each task, participants had to either manually press a key or vocally produce a syllable in response to either a pure tone or a spoken syllable prompt; meanwhile, participants had to ignore the speaker who was mouthing either /ba/ or /da/ in distractor videos. Shaded cells indicate expected S-S and/or S-R overlap in each task. ....	67
Figure 3.7 a Schematic timeline of a trial in Experiment 2. b Four key-tone-syllable assignments counterbalanced across participants. c AOI used to sort eye-tracking data. The top and the bottom white squares define the eye and the mouth areas respectively, and the area outside these two squares were defined as the other area. The size of each square was $5.22^\circ \times 5.22^\circ$ of visual angle. ....	69
Figure 3.8 ER compatibility effects $\pm$ SEs in each experimental condition. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where SSC+SRC effects were expected.....	75
Figure 3.9 ER compatibility effects $\pm$ SEs for each prompt type for each response modality. ....	76
Figure 3.10 ER compatibility effects $\pm$ SEs at each SOA for each prompt type. ....	77

Figure 3.11 Standardised RT compatibility effects $\pm$ SEs in each experimental condition. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where SSC+SRC effects were expected. ....	78
Figure 3.12 Standardised RT compatibility effects $\pm$ SEs for each prompt type for each response modality.....	79
Figure 3.13 Standardised RT compatibility effects $\pm$ SEs at each SOA for each prompt type.....	80
Figure 3.14 Standardised RT compatibility effects $\pm$ SEs at each SOA for each response modality. ....	81
Figure 3.15 Mean sample sizes $\pm$ SEs for each AOI in each task. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where c SSC+SRC effects were expected. ....	84
Figure 4.1 a Schematic timeline of a trial presented in either the pre-training or post-training session. b The experimental procedure with pre-training, training, and post-training sessions. Participants were randomly assigned to either mirror or counter-mirror group.....	104
Figure 4.2 a ERs $\pm$ SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training (bottom-right) sessions in the counter-mirror group. b ER compatibility effects $\pm$ SEs for the pre- and post-training sessions at each SOA in each group. ....	108
Figure 4.3 ERs $\pm$ SEs for each testing session in each group. ....	109
Figure 4.4 ERs $\pm$ SEs at each SOA in each testing session. ....	110
Figure 4.5 a Mean RTs $\pm$ SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training (bottom-right) sessions in the counter-mirror group. b RT compatibility effects $\pm$ SEs for the pre- and post-training sessions at each SOA in each group. ....	111
Figure 4.6 Mean RTs $\pm$ SEs at each SOA for each testing session. ....	113
Figure 4.7 Mean RTs $\pm$ SEs at each SOA for each compatibility condition. ....	114
Figure 4.8 Mean RTs $\pm$ SEs for each compatibility condition in each testing session in each group. ....	115
Figure 4.9 ERs $\pm$ SEs for each block in each training group. ....	116
Figure 4.10 Mean RTs $\pm$ SEs for each block in each training group.....	116
Figure 5.1 Waveforms and spectrograms of the speaker's recordings of /Ba/ in a, /La/ in b, /ba/ in c, and /la/ in d. Yellow rectangles make the consonant part of each syllable, and blue rectangles make the vowel part of each syllable. ....	138

Figure 5.2 Mean RTs $\pm$ SEs in each experimental condition. Four panels represent the native (top-left) and non-native (top-right) sessions with manually measured response times and the native (bottom-left) and non-native (bottom-right) sessions with response times measured using the voice-key. ....	139
Figure 5.3 Mean RTs $\pm$ SEs in each experimental condition. Four panels represent the native (top-left) and non-native (top-right) sessions with manually measured response times and the native (bottom-left) and non-native (bottom-right) sessions with response times measured using the voice-key. ....	141
Figure 5.4 Mean RTs $\pm$ SEs in each experimental condition. Four panels represent the native (top-left) and non-native (top-right) sessions with manually measured response times and the native (bottom-left) and non-native (bottom-right) sessions with response times measured using the voice-key. ....	143
Figure 5.5 a Schematic timeline of a trial presented in either the pre-training or post-training session. b The experimental procedure with pre-training, training, and post-training sessions. Participants were randomly assigned to either mirror or counter-mirror group. Both pre- and post-training sessions consist of one native and one non-native SRC task, and the order of two tasks was counterbalanced across participants. ....	145
Figure 5.6 a ERs $\pm$ SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training (bottom-right) sessions in the counter-mirror group. b ER compatibility effects $\pm$ SEs at each SOA in each testing session for each training group. ....	153
Figure 5.7 ERs $\pm$ SEs at each SOA for each compatibility condition. ....	154
Figure 5.8 a Mean RTs $\pm$ SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training (bottom-right) sessions in the counter-mirror group. b RT compatibility effects $\pm$ SEs for at each SOA in each testing session for each training group. ....	156
Figure 5.9 Mean RTs $\pm$ SEs for each compatibility condition in each nativeness session. ....	159
Figure 5.10 Mean RTs $\pm$ SEs at each SOA for each testing session. ....	159
Figure 5.11 Mean RTs $\pm$ SEs at each SOA for each compatibility condition. ....	160
Figure 5.12 Mean RTs $\pm$ SEs at each SOA in each testing session for each nativeness session. ....	161
Figure 5.13 Mean RTs $\pm$ SEs for each compatibility condition in each testing session for each group. ....	162
Figure 5.14 a ERs $\pm$ SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training (bottom-right) sessions in the counter-mirror group. b ER compatibility effects $\pm$ SEs for each syllable response in each testing session for each training group. ....	163
Figure 5.15 ERs $\pm$ SEs for each compatibility condition for each response syllable. ....	165

Figure 5.16 a Mean RTs  $\pm$  SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training (bottom-right) sessions in the counter-mirror group. b RT compatibility effects  $\pm$  SEs for each syllable response in each testing session for each training group. .... 166

Figure 5.17 Mean RTs  $\pm$  SEs for each compatibility condition for each response syllable. .... 168

Figure 5.18 ERs  $\pm$  SEs for each block in each group. .... 169

Figure 5.19 Mean RTs  $\pm$  SEs for each block in each group. .... 170

## List of Tables

Table 3.1 Task paradigms used by Kerzel & Bekkering (2000) and Kerzel (2002) to investigate SSC and SRC effects.....	48
Table 3.2 Three-way ANOVA summary for ERs, as a function of prompt type, compatibility, and SOA.....	61
Table 3.3 ERs, SEs, and 95% confidence intervals (CIs) for each level within prompt type and compatibility.....	62
Table 3.4 Three-way ANOVA summary for log-transformed RTs, as a function of prompt type, compatibility, and SOA. ....	63
Table 3.5 Mean log-transformed RTs, back-transformed RT, SEs, and 95% CIs for each level within prompt type, compatibility, and SOA.....	64
Table 3.6 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between prompt type and compatibility.....	65
Table 3.7 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between prompt type, compatibility, and SOA. ....	65
Table 3.8 Three-way ANOVA summary for ER compatibility effects, as a function of prompt type, response modality, and SOA. ....	75
Table 3.9 ER compatibility effects, SEs, and 95% CIs for each level within response modality.	76
Table 3.10 ER compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and response modality.....	76
Table 3.11 ER compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and SOA. ....	77
Table 3.12 Three-way ANOVA summary for standardised RT compatibility effects, as a function of prompt type, response modality, and SOA.....	78
Table 3.13 Standardised RT compatibility effects, SEs, and 95% CIs for each level within response modality and SOA.....	79
Table 3.14 Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and response modality. ....	80
Table 3.15 Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and SOA.....	81
Table 3.16 Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between response modality and SOA. ....	82
Table 3.17 Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type, response modality, and SOA. ....	83
Table 3.18 Three-way ANOVA summary for eye-tracking same sizes, as a function of prompt type, response modality, and AOI. ....	84
Table 3.19 Mean sample sizes, SEs, and 95% CIs for each level within AOI.....	85

Table 4.1 Four-way ANOVA summary for ERs, as a function of training, test, compatibility, and SOA.....	108
Table 4.2 ERs, SEs, and 95% CIs for each level within test, compatibility, and SOA. ....	109
Table 4.3 ER, SEs, and 95% CIs for each level within the interaction between training and test. ....	109
Table 4.4 ERs, SEs, and 95% CIs for each level within the interaction between test and SOA. ....	110
Table 4.5 Four-way ANOVA summary for log-transformed RTs, as a function of training, test, compatibility, and SOA. ....	112
Table 4.6 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within test, compatibility, and SOA. ....	113
Table 4.7 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between test and SOA.....	113
Table 4.8 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between compatibility and SOA.....	114
Table 4.9 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level in the three-way interaction between training, test, and compatibility.....	115
Table 4.10 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each block. ....	117
Table 5.1 Consonant and vowel durations in ms of the speaker’s articulation of each syllable in both video (V) and audio (A) recordings.....	147
Table 5.2 Five-way ANOVA summary for ERs, as a function of training, test, nativeness, compatibility, and SOA. ....	153
Table 5.3 ERs, SEs, and 95% CIs for each level within nativeness, compatibility, and SOA....	154
Table 5.4 ERs, SEs, and 95% CIs for each level within the interaction between compatibility and SOA.....	155
Table 5.5 Five-way ANOVA summary for log-transformed RTs, as a function of training, test, nativeness, compatibility, and SOA. ....	157
Table 5.6 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within test, nativeness, compatibility, and SOA. ....	158
Table 5.7 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between nativeness and compatibility. ....	159
Table 5.8 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between test and SOA.....	160
Table 5.9 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between compatibility and SOA.....	160
Table 5.10 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between test, nativeness, and SOA. ....	161

Table 5.11 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between training, test, and compatibility. ....	162
Table 5.12 Four-way ANOVA summary for ERs, as a function of training, test, compatibility, and syllable. ....	164
Table 5.13 ERs, SEs, and 95% CIs for each level within compatibility and syllable. ....	164
Table 5.14 ERs, SEs, and 95% CIs for each level within the interaction between compatibility and syllable. ....	165
Table 5.15 Four-way ANOVA summary for log-transformed RTs, as a function of training, test, compatibility, and syllable. ....	167
Table 5.16 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within test, compatibility, and syllable. ....	168
Table 5.17 Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between compatibility and syllable. ....	169

## List of Abbreviations

AOI	Area of Interest
ANOVA	Analysis of Variance
CI	Confidence Interval
ER	Error Rate
DIVA	Directions Into Velocities of Articulators
fMRI	Functional Magnetic Resonance Imaging
IFC	Inferior Frontal Cortex
IPA	International Phonetic Alphabet
L2	Second Language
MEP	Motor-Evoked Potential
mPFC	Medial Prefrontal Cortex
RT	Response Time
S-R overlap	Stimulus-Response overlap
S-S overlap	Stimulus-Stimulus overlap
SD	Standard Deviation
SE	Standard Error
SOA	Stimulus Onset Asynchrony
SRC	Stimulus-Response Compatibility
SSC	Stimulus-Stimulus Compatibility
TBS	Theta-Burst Stimulation
tDCS	Transcranial Direct Current Stimulation
TMS	Transcranial Magnetic Stimulation
TPJ	Temporoparietal Junction
VOT	Voice Onset Time

## Acknowledgments

The work presented in this thesis was funded by the China Scholarship Council and by the BIAL Foundation under Grant No. 267/14 to Dr Patti Adank.

Over the past five years, I was fortunate to meet many wonderful people who I would like to thank for making completion of this thesis possible.

First, I would like to thank Patti for all the guidance and support; for teaching me how to do research and how to write a good scientific paper (though I'm still far from being capable of doing it); for always challenging and pushing me to do better; for being extremely patient with my super slow writing and never complaining about it; for giving me the opportunity to learn to teach and supervise students and to communicate with them; for sharing your life experience with me, your happy moments and your struggles as a researcher. I had a great time working with you. Thank you, Patti.

Thanks to Andrew, Helen, Dan, Gwijde, Julie, Max, Tony, and Han for all the advice and help and all the conversations that were either serious or maybe scientifically nonessential yet personally important. This PhD journey was uneasy and emotional, and I am grateful to be able to share all my academic (and also otherwise) ups and downs with you.

Thanks to Bronwen for your advice and support and for the cheerful conversations we had in your office, in the corridor, and in the kitchen. Thanks to Anqi, Rachel, Xiao, Julie, Jieun, Chengxia, Shiran (and Maddie!), Katherina, Anna, Clara, Ana, Elise, Bryony, Gwen, and Outi for making Chandler House a great place to work.

Thanks to all of the participants who took part in most boring studies ever to make this research possible. To Mairéad and Clare for accepting this viva invitation and giving your commitment during this unusual period and for taking the time to read this thesis.

Thanks to Waitrose in Brunswick Centre for providing me with a free coffee every day and offering a place to wander after a day's work. To Ginger for your meow greetings every time we came across each other. To Mafang whose radio programmes kept me company through innumerable sleepless nights. To BFI Southbank, Prince Charles Cinema, and Curzon Bloomsbury for letting me enjoy my favourite old films on big screenings and offering a place for me to withdraw from the crowd and yet enjoy the fullness of life.

Thanks to Yuwei, Shelly, Wentian, Lake, Han, Chengsheng, and Anqi (unofficially acknowledged!) for making home home. I enjoyed every meal we cooked/ordered, every bottle of wine/whisky/beer/sake we opened, every chat we had, and every walk we took in Bloomsbury and also other areas in London. To Alice, Amber, Jia, Fish, Tuzi, and Tyia for constantly checking whether I was still staying sane. I look forward to seeing you again somewhere on this earth in the future. To Carol for frequently reaching out to me and listening to my existential murmuring. I will miss every café/desert place where we spent hours chatting. To Mary and Menglu for spoiling me with gifts and letters from afar. I miss travelling with you and hope to see you soon.

Last but most importantly, thanks to Yidi and Jing and their families for supporting my family when they were going through difficult times and when I was far away from home. I am eternally grateful for everything you have done for my family.

This thesis is dedicated to my mom, my dad, and my grandma Laopan who support me in all of my adventures with unwavering love. I love you all.

## Chapter 1 Introduction

In 1872, Charles Darwin stated in his book *The Expression of the Emotions in Man and Animals* that there exists a strong tendency in humans to imitate actions they observe and that such imitative behaviours occur independently of the observers' conscious will. In this book, Darwin discussed his observations that members of the audience tended to clear their throats when hearing a public singer suddenly become a little hoarse and that spectators tended to move their feet when watching a performer make his spring at leaping matches. Both observations, according to Darwin, were examples showing people's imitative tendency, though he mentioned that the first example could also be due to habit.

While Darwin merely made observations of people's imitative tendency, more recent experimental research has demonstrated that people tend to imitate a wide range of actions they observe in both relatively social situations and tightly controlled laboratory settings, and the imitated actions could be facial expressions, manual gestures, bodily movements, or speech patterns (Dijksterhuis & Bargh, 2001; Heyes, 2011). During social interactions, people tend to copy their conversation partners' incidental gestures, such as face-rubbing and foot-shaking, and they tend to smile more if their partners do so, a phenomenon known as *motor mimicry* (Chartrand & Bargh, 1999). Additionally, unacquainted interlocutors' pronunciations are judged to sound more similar over the course of their conversation and after it, as compared to their pronunciations recorded before the conversation, a phenomenon called *phonetic convergence* (Pardo, 2006). Furthermore, in tightly controlled laboratory settings, people tend to respond faster when observing the same action as they are prompted to perform than when observing a different action (Brass, Bekkering, Wohlschläger, & Prinz, 2000; Kerzel & Bekkering, 2000; Stürmer, Aschersleben, & Prinz, 2000). Effects of this kind are known as *automatic imitation* which, unlike motor mimicry or phonetic convergence, is suggested to represent a kind of covert imitation during which motor representations of the observed actions are activated but the actual actions are not overtly executed by the observers (Heyes, 2011). Nevertheless, all these three types of imitative behaviours occur automatically and unintentionally to the actions that are already part of the observers' motor repertoire.

In cognitive psychology, automatic and unintentional imitative behaviours are categorized as *simple imitation*, as opposed to *complex imitation* during which observers intentionally copy a novel sequence of movements (i.e., imitation learning), a sequence that they have never performed before (Heyes, 2013; Tomasello, 1996). However, despite the distinctions between these two types of imitation, it is suggested that the psychological mechanisms underlying simple imitation may also mediate complex imitation and thus studies of the former may contribute to the understanding of the latter (Heyes, 2005). Particularly, imitation requires a translational process between perceptual input and behavioural output

(Chartrand & van Baaren, 2009; Dijksterhuis & Bargh, 2001; Heyes, 2001). The advantage of using simple imitation tasks to study the perceptuomotor translational process is that they help isolate the perceptuomotor process by reducing demands on working memory and other executive processes possibly involved in intentional imitation tasks (Catmur, Walsh, & Heyes, 2009).

The focus of this thesis is on automatic imitation of speech and its underlying mechanisms. Methodologically, automatic imitation is examined using the *stimulus-response compatibility* (SRC) paradigm (see Section 2.1.1 for detailed descriptions), a paradigm that has been widely employed in both experimental and cognitive psychology for decades (e.g., Eriksen & Eriksen, 1974; Fitts & Seeger, 1953; Simon & Rudell, 1967; Stroop, 1935). Particularly, in the last two decades, research on automatic imitation of manual gestures has been productive, and various theories and models have been developed to account for this imitative behaviour and its underlying mechanisms (see Campbell & Cunnington, 2017; Cook, Bird, Catmur, Press, & Heyes, 2014; Heyes, 2011 for reviews). Behaviourally, automatic imitation demonstrates an imitative capacity in humans to quickly map an observed action onto their motor repertoire, indicating the existence of an observation-execution matching link between motor and sensory representations of the same action. The origin of the imitative capacity has been the source of much debate in cognitive psychology. Past studies have examined the flexibility of the observation-execution link underlying people's imitative capacity and suggested that the link is experience-dependent (see Catmur, 2013 for a review). Specifically, it is proposed that the matching properties of the observation-execution link are a product of associative sensorimotor learning during which people experience highly correlated sensory and motor events (Heyes, 2005, 2010; see Section 2.1.3 for detailed descriptions), and sources of this sensorimotor experience are suggested to be abundant in social life (Ray & Heyes, 2011).

The primary aims of this thesis are to establish the role of sensorimotor experience in modulating the imitative capacity underlying speech actions that are either familiar or novel to perceivers and to therefore explore commonalities between different effector systems. Speech actions and manual gestures are qualitatively different in terms of effectors (articulators versus hands) used for production and dominant sensory modalities (auditory versus visual) involved in perception. However, despite these differences, people show tendencies to imitate both types of movements during perception and frequently use both in daily communication. These similarities suggest the existence of a perception-production matching link between sensory and motor representations of the same speech actions and potentially abundant sources of sensorimotor experience supporting the development of this link. It is therefore reasonable to ask whether the imitative capacity underlying speech actions is governed by the same mechanism that mediates the imitative capacity underlying manual

gestures and, to be more specific, whether the flexibility observed for the observation-execution link underlying manual gestures extends to the perception-production link underlying speech actions.

Beyond the primary interest of speech imitation and the importance of discovering commonalities between different effector systems, there are advantages to studying imitation in the speech domain. Speech actions (e.g., consonant-vowel syllable articulations) consist of sub-movements executed by different muscle groups and require precise temporal control (Borden, Harris, & Raphael, 1994). Thus, the distinction between two speech sounds could result from different articulators performing the same sub-movement or different sub-movements performed by the same articulator. Thus, each speech action could result from a unique effector-movement combination (see Section 5.1.2 for detailed discussion). The fact that not all languages share the same set of speech sounds makes some combinations (i.e., non-native speech sounds) more difficult for non-native speakers to produce and/or perceive and accordingly, novel speech actions can be defined as unfamiliar effector-movement combinations. In contrast, novel manual gestures have been commonly defined as unfamiliar sequences of simple manual movements (Heyes, 2005; Wiggett, Hudson, Clifford, Tipper, & Downing, 2012), the imitation learning of which may also depend on individuals' working memory capacity. The same mechanism has been proposed to be involved in imitation of both familiar and novel actions (Heyes, 2005), and speech may provide an avenue to investigate the role of sensorimotor experience in the development of the imitative capacity underlying novel actions by minimizing demand on working memory. One thing to clarify is that how speech actions differ from one another could be more complicated than what has been described above; but the question has been simplified to emphasize the point that, in the speech domain, novel actions can be defined as unfamiliar effector-movement combinations rather than unfamiliar sequences of movements used in the hand literature.

This thesis consists of six chapters. Chapter 1 provides a brief introduction and an outline of the thesis. Chapter 2 presents literature reviews of imitation research in cognitive psychology and sensorimotor interaction research in speech science and then summarises research aims and objectives investigated in this thesis. Chapter 3 evaluates automatic imitation effects measured using different speech SRC tasks in a within-subjects design, since these tasks have only been used separately in previous studies to investigate automatic imitation of speech. Chapter 4 examines the role of sensorimotor experience in modulating the imitative capacity underlying familiar, native speech actions that are perceived visually. Chapter 5 explores the role of sensorimotor experience in modulating the imitative capacity underlying novel, non-native speech actions that native British English speakers find difficult to produce. Chapter 6 concludes with discussions of the empirical work presented in this thesis

in the context of previous research and discusses the limitations of the presented work and possible directions for further research.

All experiments presented in this thesis were approved by the University Research Ethics committee (UCL #0599/001).

## Chapter 2 Literature review

This chapter includes three main sections. The first section provides a literature review of imitation research in cognitive psychology, and the second section provides a literature review of sensorimotor interaction research in speech science. The last section summarises research aims and objectives examined in this thesis.

### 2.1 Imitation of manual gestures

In this section, I will first give an overview of automatic imitation studies based on manual gestures, discuss the proposed neural substrate underlying imitation, outline several theoretical accounts of the origin of humans' imitative capacity, and lastly explain sensorimotor training studies that examine the flexibility of the observation-execution link underlying manual gestures.

#### 2.1.1 *Automatic imitation and spatial compatibility*

Automatic imitation is measured behaviourally using SRC tasks, in which the speed and/or accuracy of participants' performance is modulated by the relationship between task-irrelevant distractors and the participants' responses (see Heyes, 2011 for a review). In an early study, Stürmer et al. (2000, Experiment 1) instructed participants to either open or close their hand (i.e., spread apart or grasp) when seeing the hand in a distractor video turned red or blue (i.e., a colour prompt). Meanwhile, the hand in the distractor video was performing either the same (i.e., compatible) or a different (i.e., incompatible) movement, and participants were asked to ignore this task-irrelevant stimulus movement but to respond to the task-relevant stimulus colour. Compatible distractor movements elicited faster responses than incompatible distractor movements, despite the fact that participants were asked to ignore these movements. Automatic imitation is defined as the response time difference between the two compatibility conditions (Heyes, 2011). However, in Stürmer et al., it was unclear whether observing the same action facilitated participants' responses and/or observing a different action delayed their responses. In another study by Brass et al. (2000, Experiment 1), participants were instructed to respond to a number prompt ("1" or "2") by lifting their index or middle finger while ignoring the hand that was simultaneously lifting either the same (i.e., compatible) or the other (i.e., incompatible) finger in a distractor video. A baseline condition was included in which the hand remained static in the distractor video while participants were responding to the prompt. Responses in the baseline condition were slower than those in the compatible condition and faster than those in the incompatible condition. Thus, Brass et al. further

confirmed that automatic imitation results from both facilitation by the observation of a compatible action and interference by the observation of an incompatible action.

Automatic imitation has been categorised as an imitative SRC effect in which performance is modulated by the relative correspondence between movements observed from distractor stimuli and movements executed in responses (Catmur & Heyes, 2011; Catmur et al., 2009). In cognitive psychology, several dual-route models have been proposed to explain SRC effects (e.g., De Jong, Liang, & Lauber, 1994; Hommel, 1993, 1997; Kornblum, Hasbroucq, & Osman, 1990; Kornblum, Stevens, Whipple, & Requin, 1999). The most general is the Dimensional Overlap model that identifies a taxonomy of compatibility effects (Kornblum et al., 1990, 1999). According to the Dimensional Overlap model, when both prompts and distractors (i.e., task-relevant and irrelevant stimuli) are presented in a task, the presentation of a prompt activates the assigned response through a controlled process based on the instructions; meanwhile, the presentation of a distractor automatically activates a corresponding response, if there is *dimensional overlap* (i.e., similarity in mental representations) between distractors and responses. This automatic processing pathway is proposed to consist of excitatory links between corresponding distractors and responses, and thus activation of a distractor automatically leads to activation of the corresponding response. Accordingly, in an SRC effect, performance is facilitated when responses activated through the controlled and the automatic processes are compatible and is delayed when the two activated responses are incompatible.

However, automatic imitation effects observed by Brass et al. (2000) and Stürmer et al. (2000) were likely to be confounded with spatial SRC effects in which performance is modulated by the relative left-right or top-down spatial alignment between distractor movements and response movements. To test whether automatic imitation can be reduced to a spatial SRC effect, Brass, Bekkering, & Prinz (2001, Experiment 3) instructed participants to tap or lift their right index finger while ignoring the hand that was simultaneously lifting or tapping its index finger in a distractor video. In one condition, the video images were flipped upside down. As a result, the finger that was being lifted was seen as moving downward and the finger tapped was seen as moving upward. Consequently, in this condition, imitative compatible trials were always spatial incompatible, and imitative incompatible trials were spatial compatible. Faster responses were observed for imitative compatible trials than for imitative incompatible trials in both normal and flipped conditions, suggesting that, different from spatial SRC effects, automatic imitation is movement-specific. To control for left-right spatial compatibility, studies also employed both left and right hands to perform finger movements in distractor videos while asking participants to use their right hand only to respond (e.g., index/middle finger tapping in Bertenthal et al., 2006; index/little finger abduction in Catmur et al., 2009). For instance, Catmur et al. presented hand stimuli in the

first person perspective, so the left-right position of two fingers in the right-hand video was the same (i.e., spatial compatible) as that of participants', but the left-right position of two fingers in the left-hand video was reversed (i.e., spatial incompatible). Automatic imitation occurred regardless of spatial compatibility, suggesting that automatic imitation is also effector-specific. Additionally, to isolate automatic imitation of hand opening/closing movements from spatial SRC effects, Heyes, Bird, Johnson, & Haggard (2005) presented distractor movements in a plane orthogonal to that of response movements and automatic imitation was also observed.

Therefore, it is reasonable to conclude that automatic imitation represents a type of imitative SRC effects, which occurs because observing actions in distractor videos automatically activates corresponding motor patterns that then interact with responses cued by prompts. Responses are facilitated when the observation of actions shown in compatible distractors activates correct motor patterns that will be used in following responses; conversely, responses are delayed when the observation of actions shown in incompatible distractors activates wrong motor patterns that have to be suppressed before correct responses are performed. The size of an automatic imitation effect thus indicates the degree of observation-induced motor activation with a larger effect indicating greater motor activation.

### ***2.1.2 Proposed neural substrate underlying imitation***

The discovery of mirror neurons in the brain of macaque monkeys (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992) has renewed interests in imitation and supported the hypothesis that automatic imitation involves observation-induced motor activation. Mirror neurons fire when the monkey is executing an action (e.g., grasping an object) and when the monkey is observing the same action, suggesting a direct neural link between the observation and execution of an action. Mirror neurons were first found in area F5 in the ventral premotor cortex (Bonini et al., 2010; Bonini, Maranesi, Livi, Fogassi, & Rizzolatti, 2014; di Pellegrino et al., 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996) and then in other areas that are anatomically connected with F5. These other areas include the dorsal premotor cortex (Tkach, Reimer, & Hatsopoulos, 2007), the parietal cortex (Bonini et al., 2010; Fogassi et al., 2005; Pani, Theys, Romero, & Janssen, 2014), and the primary motor cortex (Dushanova & Donoghue, 2010; Tkach et al., 2007; Vigneswaran, Philipp, Lemon, & Kraskov, 2013). Particularly, the rostral part of the inferior parietal lobule and the ventral premotor cortex (area F4 & F5) have been suggested to form the basic cortical mirror circuit (also called the parieto-frontal mirror circuit by Rizzolatti & Sinigaglia [2010]), supporting sensorimotor transformation during action observation (Ferrari, Gerbella, Coudé, & Rozzi, 2017; Rizzolatti & Craighero, 2004). Furthermore, Ferrari, Gallese, Rizzolatti, & Fogassi (2003) found mirror

neurons that respond to execution and observation of digestive and communicative mouth movements. Keysers et al. (2003) and Kohler et al. (2002) further reported auditory mirror neurons that respond when a monkey performs an action executed by its hand or mouth and when it hears another individual perform the same action. Thus, mirror neurons seem to underlie actions produced by different effectors and perceived through different modalities. Using direct neural recordings in humans, two studies have shown that humans also possess mirror neurons (Mukamel, Ekstrom, Kaplan, Jacoboni, & Fried, 2010; Perry et al., 2017).

While only few studies have shown direct evidence of the existence of mirror neurons in humans, other studies have employed non-invasive techniques to indirectly demonstrate the presence of a human mirror neuron system that shows ‘mirror’ activation when participants are executing actions and when they are passively observing the same actions. Functional magnetic resonance imaging (fMRI) studies have identified several regions that show mirror responses, and these regions include the inferior frontal gyrus, the premotor cortex, the anterior intraparietal sulcus, and both inferior and superior parietal lobule (e.g., Buccino et al., 2004; Gazzola, Rizzolatti, Wicker, & Keysers, 2007; Jacoboni et al., 1999; Vogt et al., 2007; see Molenberghs, Cunnington, & Mattingley, 2012 for a review). Particularly, the inferior frontal gyrus is suggested to be the human homologue of monkey area F5, and the premotor cortex is the human homologue of monkey area F4 that is anatomically connected with F5 (Rizzolatti & Craighero, 2004). These homologues indicate the existence of the basic parieto-frontal mirror circuits in both humans and monkeys (Rizzolatti & Sinigaglia, 2010).

The combined results from fMRI studies indicate a neural network through which action observation leads to activation in the motor system involved in action execution. More recently, a few studies have employed fMRI adaptation paradigms to further determine the presence of neural populations selective for particular actions but invariant in responses across the observation and execution of these actions (e.g., Chong, Cunnington, Williams, Kanwisher, & Mattingley, 2008; Kilner, Neal, Weiskopf, Friston, & Frith, 2009; Lingnau, Gesierich, & Caramazza, 2009; Press, Weiskopf, & Kilner, 2012). Functional MRI adaptation refers to the observation that repeated presentations of a sensory stimulus decrease firing rate in neurons encoding that stimulus feature and consequently reduce blood-oxygen-level dependent responses relative to presentations of a novel stimulus (Grill-Spector, Henson, & Martin, 2006; Krekelberg, Boynton, & van Wezel, 2006). The inferior frontal gyrus and inferior parietal lobe have been found to show adaptation effects during the observation of previously executed actions but not during the observation of novel actions (Chong et al., 2008; Kilner et al., 2009). Findings from these fMRI adaptation studies have further suggested the presence of neural populations selective for particular actions, regardless whether these actions are observed or executed, and thus indicated the matching properties of the human mirror neuron system.

The selective characteristic of observation-induced motor activation has also been exhibited in studies using single-pulse transcranial magnetic stimulation (TMS). Specifically, these TMS studies demonstrate that passive observation of an action selectively enhances the excitability of the motor representations of muscles involved in executing the same action, indicating the matching properties of the observation-execution link at the neurophysiological level (Catmur, Walsh, & Heyes, 2007; Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Strafella & Paus, 2000). Single-pulse TMS applied to the primary motor cortex elicits motor evoked potentials (MEPs) in the corresponding peripheral muscles, with a larger MEP indicating greater motor excitability (Hallett, 2007). In Catmur et al., for instance, observing index-finger movements elicited higher MEPs in the index-finger muscle than observing little-finger movements and the other way around for MEPs measured from the little-finger muscle. These TMS studies have revealed effector-specific sensorimotor mapping at the neurophysiological level, and automatic imitation is proposed to reflect the behavioural consequences of this observation-induced motor facilitation (Heyes et al., 2005).

Furthermore, automatic imitation is proposed to be mediated by the human mirror neuron system and studies have used automatic imitation to index the mirror system functioning (e.g., Longo, Kosobud, & Bertenthal, 2008; van Schie, van Waterschoot, & Bekkering, 2008). Crucial evidence for a causal relation between the mirror neuron system and automatic imitation comes from studies using repetitive or double-pulse TMS to temporarily disrupt the functioning of the stimulated areas (Catmur et al., 2009; Mengotti, Ticini, Waszak, Schütz-Bosbach, & Rumiati, 2013; Newman-Norlund, Ondobaka, Van Schie, van Elswijk, & Bekkering, 2010). Catmur et al., for example, demonstrated that automatic imitation of finger movements was impaired after applying repetitive TMS to the inferior frontal gyrus but did not change after applying repetitive TMS to a control site or in the other control condition with no TMS applied.

Together, research has demonstrated that people tend to covertly imitate actions they observe, which then affects their subsequent performance. This imitative behaviour is suggested to occur because action observation leads to activation in the corresponding motor patterns used to perform the same action, and the neural network possibly mediating this imitative behaviour consists of brain regions that are homologous with the key areas where monkey mirror neurons are, responding to both action observation and execution.

### ***2.1.3 The origin of the imitative capacity***

Imitation has been suggested to play a key role in the development of a range of behaviours and cognitive functions and also in supporting social interactions (Byrne & Russon, 1998; Chartrand & Bargh, 1999; Frith & Frith, 2012; Wang & Hamilton, 2012). However, there has

been no consensus on the question of how humans acquire the imitative capacity, and various accounts have been proposed. The Active Intermodal Mapping model (Meltzoff, 2002; Meltzoff & Decety, 2003; Meltzoff & Moore, 1997) proposes that humans' imitative capacity is innate and fundamental to social cognition. Developmental studies showing that infants can imitate a range of orofacial movements have provided strong support to this proposition (Meltzoff & Moore, 1977, 1983, 1989). In their 1983 study, Meltzoff & Moore tested 40 newborn infants with a mean age of 32 hours old (the oldest was 72 hours old and the youngest 42 minutes old) and found that newborn infants were able to imitate orofacial movements without much experience of seeing or doing these movements. This finding led the authors to propose that observation and execution of actions, particularly of orofacial movements, are intrinsically linked and coded within a "supramodal" framework where observed and executed actions are compared.

A similar idea is shared by the Action Understanding theory first proposed to explain the functions of the mirror neuron mechanism (Ferrari, Bonini, & Fogassi, 2009; Rizzolatti & Arbib, 1998; Rizzolatti & Craighero, 2004; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). The mirror neuron mechanism is suggested to constitute an observation-execution matching system whose matching properties are present at birth and play an essential role in action understanding and imitation. The meaning/goal of an action is recognised because observing an action triggers a neural activity similar to that generated internally when the observer is to perform the same action. The triggered neural activity includes not only the movement representation of the observed action but also the associated movement consequences; thus, the observer knows what to expect following the action. The mirror mechanism is also suggested to be responsible for both neonatal and automatic imitation which are crucial for promoting interactions between infants and caregivers and supporting early development of sensorimotor skills and social interaction abilities. Hence, the matching properties are proposed to be genetically predisposed because individuals better at imitating and understanding others' actions were probably more reproductively successful. Notwithstanding, both the Active Intermodal Mapping model and the Action Understanding theory, both of which will be considered as belonging to a *genetic account* in this thesis, also value the role of experience in the development of the imitative capacity and suggest that motor and/or sensory experience facilitates imitation (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005; Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006; Meltzoff & Decety, 2003). Thus, according to the genetic account, sensory and/or motor experience plays a facilitative or tuning role in the development of the imitative capacity (Gottlieb, 1976).

Alternatively, the Associative Sequence Learning model (Catmur et al., 2009; Heyes, 2001, 2005, 2010; Ray & Heyes, 2011) postulates an *associative account* and proposes that the imitative capacity is essentially a product of associative sensorimotor experience that

consists of concurrent observation and execution of the same actions. The associative mechanism is suggested to be a domain-general process that also produces Pavlovian and instrumental conditioning (R. Cook, Press, Dickinson, & Heyes, 2010). Therefore, the associative account further proposes that the imitative capacity depends on the domain-general mechanism that allows for a positive contribution to the development of imitation and other cognitive functions but not necessarily develop to do so. For example, people can see their own hands while performing different movements, thus allowing for the opportunity to establish and reinforce *excitatory links* between sensory and motor representations of the same manual gestures. Consequently, through these excitatory links, seeing a manual gesture automatically activates the motor representations of that same gesture. More importantly, however, due to its experience-dependence, the imitative capacity is hypothesised to be flexible and subject to change through sensorimotor experience that associates observation and execution of different actions. Hence, according to the associative account, sensorimotor experience plays an inductive or forging role in the development of humans' imitative capacity, meaning that, without such experience, there would be no observation-execution matching link supporting imitation (Gottlieb, 1976).

As mentioned above, research on neonatal orofacial imitation has provided strong support for the genetic account. This is important because in general infants have little associative experience of observing and executing the same actions. Furthermore, infants cannot see their own faces when performing orofacial movements, thus giving them even less chance to associate observation and execution of the same orofacial movements. Following the genetic account, since infants' sensorimotor experience is limited, so the imitative capacity they demonstrate is more likely to be innate rather than learned. However, findings of neonatal imitation have been questioned in recent studies (Oostenbroek et al., 2016; Ray & Heyes, 2011). For example, Oostenbroek et al. conducted a longitudinal study involving more than 100 infants and found that infants were just as likely to produce the same gestures in the control condition as they were in the imitation condition. In imitation literature, manual and orofacial movements have been considered as fundamentally different processes. This is because manual and orofacial actions differ in the fact that people can directly observe their own manual movements that are visually transparent, but they cannot observe their own orofacial movements which are visually opaque, unless observed through mirrors or other optical surfaces (Heyes, 2005; Meltzoff, 2002). To solve the dichotomy between orofacial and manual movements, Casile, Caggiano, & Ferrari (2011) have suggested different developmental trajectories for the mirror systems underlying two types of movements, with the orofacial mirror system being innate and the hand being learned.

The associative account strongly emphasises the role of social environments, instead of genetic constraints, in determining the development of humans' capacity to imitate both

orofacial and manual actions (Heyes, 2005, 2018; Ray & Heyes, 2011). In addition to direct self-observation, the observation-execution matching link could also be built indirectly through exposure to a *common stimulus* that occurs when perceiving an action done by others and when doing that same action by oneself. In this case, language has provided an important tool particularly for children to learn to associate the same actions they do (but cannot see) and they see other people do. For example, a child may hear the word “smile” when the child is smiling and when the child sees other people smile; gradually, the child will learn to associate the experience of smiling and seeing others smile. Heard speech sounds are also important cues for children to associate their own articulations with visual perception of others' articulations. Furthermore, human infants mostly live in an imitative environment where children's actions and articulations are imitated by their caregivers more than the other way around, suggesting that infants receive plenty of visual feedback of their own actions during development (e.g., Kokkinaki & Kugiumutzakis, 2000; Masur & Rodemaker, 1999). In addition, sensorimotor experience could also be acquired through synchronous actions that occur when two or more individuals react in the same way to an event, e.g., when both the caregiver and the infant open their mouths in response to the spoon during spoon-feeding (O’Toole & Dubin, 1968) or when participating in group activities. In other words, there are plenty of sources for imitative sensorimotor experience in humans’ social life.

Following the associative hypothesis, it seems possible that the same learning mechanism contributes to the imitative capacity underlying both manual and orofacial movements since imitative sensorimotor experience is suggested to be abundant in social life. However, even within the associative scenario, whether or not people can directly observe self-generated movements could still lead to some differences affecting the development of the perception-production links underlying manual versus orofacial movements. Specifically, not being able to directly observe self-generated movements suggests that most associative sensorimotor experience with such movements is acquired through the observation of others’ actions, and this has several consequences. Firstly, the contingency and contiguity (i.e., the predictive relationship) between observed and executed movements have been suggested to play a key role in associative learning (R. Cook et al., 2010). For visually transparent manual movements, direct observation of self-generated movements guarantees such contingency and contiguity. Yet, for visually opaque orofacial movements, sensorimotor experience that relies on the observation of others’ actions reduces the contingency and contiguity and may thus affect the development of the perception-production link underlying these movements. In addition, it is likely that most associative sensorimotor experience with orofacial movements is acquired through social interactions. Accordingly, compared to manual gestures, a relatively more ecological learning environment (e.g., directly interacting with another person rather than responding to stimuli presented on the computer screen) may facilitate the establishment

of the perception-production link underlying orofacial movements. Therefore, to test the associative hypothesis, it is essential to examine both whether and to what extent different types of associative sensorimotor experience modulate the imitative capacity underlying manual versus orofacial movements.

#### **2.1.4 *Sensorimotor training***

The dispute between the genetic and the associative accounts of the origin of humans' imitative capacity echoes one of the classic psychology debates between nature versus nurture. However, a closer look at their arguments suggests that the disagreement is more about whether experience plays a tuning or forging role in the development of humans' imitative capacity and what type of experience is essential. Using automatic imitation as a behavioural measure, a series of training studies has been conducted to test the flexibility of the observation-execution link underlying manual gestures (see Catmur, 2013 for a review). These training studies are based on the assumptions that, if humans' imitative capacity is a product of associative sensorimotor learning, a short period of sensorimotor training that associates observation and execution of different actions should lead to change in automatic imitation. Findings of such would suggest that the observation-execution link remains flexible in adults and is subject to change through training. However, if humans are born with the capacity to imitate, then the mechanism should be resistant to the change that might weaken its imitative capacity and automatic imitation should not change after training.

Heyes et al. (2005) were the first to demonstrate that automatic imitation of manual gestures could be eliminated following a short period of counter-mirror training that associated observation and execution of different gestures. In the study, half of the participants received counter-mirror training in which they were instructed to open their hands in response to a closing hand video and vice versa. The other half received mirror training in which they had to open their hands in response to an opening hand video and likewise for the closing hand movement. Automatic imitation for hand opening/closing movements was tested 24 hours after training. Heyes et al. found that automatic imitation effects were impaired following counter-mirror training and preserved following mirror training. The authors suggested that the difference between two groups should not be ascribed to either sensory or motor experience alone, because both groups received the same amount of exposure to the movement stimuli and the same amount of practice of the target movements. Rather, it was the relationship between the movements observed and executed in each trial during training that modulated automatic imitation.

The associative account also proposes that sensorimotor experience of observing and executing different actions results in *inhibitory links* between sensory and motor representations of the same actions. Consequently, observing an action no longer activates motor representations of that same action and so no facilitation to performance in the compatible condition, relative to the incompatible condition, in the automatic imitation effect. Heyes et al. (2005) observed an increase in response times in the compatible condition following counter-mirror training and they ascribed this increase to inhibitory links that were established during counter-mirror training. Consequently, the response time difference between two compatibility conditions decreased and so did the automatic imitation effect. Follow-up studies have also reported reduced automatic imitation for hand/foot lifting movements following counter-mirror training (Gillmeister, Catmur, Liepelt, Brass, & Heyes, 2008; Wiggett, Hudson, Tipper, & Downing, 2011) and reduced automatic imitation for goal-directed hand movements (i.e., grasping an object) following counter-mirror training (Catmur & Heyes, 2017). Findings that counter-mirror training reduces automatic imitation seem to favour the associative account over the genetic account. However, further research is needed for sensorimotor training studies so far have been exclusively based on visually transparent manual gestures, whereas one of the problems that remain unclear concerns the flexibility of the observation-execution link underlying orofacial movements that are visually opaque. That is, it remains to be tested whether different types of sensorimotor training modulate automatic imitation of visually opaque orofacial movements. Chapter 4 in this thesis examines whether the flexibility observed for the observation-execution link underlying manual gestures extends to the perception-production link underlying visual speech. To address this issue, a between-subjects training study was conducted to test whether mirror and counter-mirror training lead to divergent changes in automatic imitation of visually perceived speech actions.

Furthermore, the associative account also proposes that the same associative mechanism is involved in imitation of both familiar and novel movements that were or were not part of the observer's motor repertoire (Heyes, 2005). Wiggett et al. (2012) examined brain responses to the observation of either trained or untrained novel sequences of hand movements following different types of training (visual, motor, or visuomotor). Twelve novel sequences of hand movements were employed in the study. During visuomotor training, participants performed three movement sequences while viewing their own movements in real time. During motor training, participants performed other three different movement sequences without being able to view their own movements. During visual training, participants passively viewed other three different movement sequences from a first-person perspective. The final three movement sequences remained novel for they had never been performed and/or viewed during training. Participants received all three types of training in two consecutive days before they passively watched all twelve movement sequences from a third-person

perspective while being scanned. Watching movement sequences trained in the visuomotor condition led to greater activity in the brain areas implicated in action observation and execution, relative to watching movement sequences trained in either motor or visual condition. In other words, Wiggett et al. found that the observation of visuomotor-trained movement sequences resulted in greater activation in the human mirror neuron system, as compared to the observation of either motorically or visually trained movement sequences.

Findings from Wiggett et al. (2012) seemed to support the associative account by demonstrating greater facilitation in the emergence of mirror responses to the observation of novel movement sequences following sensorimotor learning, as compared to either sensory or motor learning alone. However, a closer look would suggest that their findings could not unambiguously support either the associative or the genetic account because their results can be explained in different ways. Following the associative account, the observation of visuomotor-trained movement sequences elicited greater mirror activation because repetitive co-activation of sensory and motor representations of the same movement sequences during visuomotor training led to the formation of the perception-production matching link underlying these movement sequences; yet, such co-activation was absent during either visual or motor training. Alternatively, it was also possible that visuomotor training led to greater mirror activation than visual training because participants had additional motor practice with the trained movement sequences in the former, though they received the same amount of sensory exposure to the movement stimuli in both conditions; similarly, visuomotor training led to greater mirror activation than motor training because participants had additional sensory exposure to the movement stimuli in the former, though they received the same amount of motor practice with the trained movement sequences in both conditions. Thus, compared to either visual or motor training, there was always additional motor output or visual input during visuomotor training, which probably facilitated the further emergence of mirror responses to the observation of novel movement sequences. Moreover, because Wiggett et al. employed novel movements that consist of familiar elements arranged in a new sequence, mirror activation by the observation of newly learned movement sequences could be confounded with mirror activation by the observation of familiar sequence components. Also, the behavioural influence of sensorimotor training on the imitative capacity underlying novel actions has remained unclear. Therefore, whether associative sensorimotor experience plays a key role in modulating people's imitative capacity underlying novel actions requires further investigation. Chapter 5 in this thesis explores whether sensorimotor training modulates the perception-production link underlying novel speech actions. To address this issue, a between-subjects training study was conducted to examine whether mirror and counter-mirror training lead to divergent changes in automatic imitation of non-native speech actions that native British English speakers find difficult to produce.

Furthermore, as discussed in Section 2.1.1, automatic imitation effects are effector- and movement-specific. More importantly, as discussed in Section 5.1.2 below, automatic imitation effects have been found to occur when perceived and produced speech actions are different but share more features, relative to the condition when the two speech actions share fewer features. Questions then arise as to at what level perception-induced motor activation occurs and at what level training-induced modification occurs in the perception-production link. Chapter 5 in this thesis also explores whether sensorimotor training with novel speech actions modulates the perception-production link underlying familiar speech actions that share some similarities with the trained actions. To address this issue, the training study also examined whether mirror and counter-mirror training with non-native speech actions lead to divergent changes in automatic imitation of native speech actions that are produced using the same articulators as their non-native counterparts. If training-induced modification occurs at the feature level, training effects were expected to occur in automatic imitation of both non-native and native speech actions. Conversely, if training-induced modification occurs at the individual action level, training effects were expected to occur in automatic imitation of non-native speech actions only but not in automatic imitation of native speech actions.

## **2.2 Imitation of speech actions**

The link between speech perception and production systems has been indicated in a wide range of studies that demonstrate sensorimotor interaction between the two systems (see Hickok, Houde, & Rong, 2011; Ménard, 2015 for reviews), and the findings can roughly be summarised as follows: 1) the perception system is involved during speech production, e.g., receiving sensory feedback when speaking, and 2) the production system is involved during speech perception, e.g., activating the articulatory motor system when seeing and/or hearing others speak. It has been widely accepted that sensory feedback is essential for speech acquisition and speech motor control. However, whether articulatory motor involvement critically supports speech perception remains debatable, though people consistently show a tendency to either covertly or overtly imitate speech they perceive. In this section, I will first review findings of motor involvement in speech perception, which includes automatic imitation of speech, then discuss various theories and models of speech processing, and finally examine the role of sensory feedback in speech production and the flexibility of the perception-production link indicated in production research.

### ***2.2.1 Motor involvement in speech perception***

Behaviourally, there are at least two approaches to show motor involvement in speech perception. One approach is to show how speaking affects the perception of others' speech,

but this involves acoustic signals both from one's own speech and from others. To avoid auditory feedback from one's own speech, Sams, Möttönen, & Sihvonen (2005) employed covert articulation to examine motor effects on auditory speech. An articulatory McGurk effect was observed in which participants were asked to silently mouth /ka/ while receiving an auditory /pa/ and they reported hearing /ta/, demonstrating motor modulations on the processing of auditory speech. However, the paradigm that Sams et al. adapted from was initially used to examine how ambiguous visual input alters auditory perception (i.e., the McGurk effect; McGurk & MacDonald, 1976), but it does not go the other way around, and the McGurk effect only occurs to a limited set of speech sounds. Therefore, this paradigm is limited for it is difficult to extend it to other speech sounds or to the investigation of motor effects on visual speech.

The other approach is to demonstrate how seeing and/or hearing others speak affects one's own production, as in the case of automatic imitation. It may seem to be an investigation of a production problem, but the rationale is that we are examining a perceptual process by measuring and comparing participants' articulatory performance after they receive different or no speech stimuli. Part of the reason why things become relatively complicated here pertains to the nature of speech perception and production research, or action perception and production research in general. As Casserly & Pisoni (2010) have pointed out, speech production research mainly involves direct observation of overt behaviours, measuring physical movements of articulators and acoustic signals as the movements occur; in contrast, perception research focuses on examinations of hidden cognitive processes, using response time latencies or recall accuracies to draw conclusions about the underlying processing. This contrast poses difficulties for investigation of motor involvement in speech perception because one has to show 'behaviourally loud' motor involvement in a 'behaviourally quiet' perceptual process. However, despite the complications, there are advantages of using this second approach. As mentioned above, automatic imitation has been widely used in cognitive psychology to investigate the observation-execution link underlying manual gestures. More importantly, as discussed below, studies have reported automatic imitation effects for both vowels and consonants and for auditory, visual, and audiovisual speech. Hence, compared to the first approach, the SRC paradigm used to measure automatic imitation shows a greater generalisability to more speech sounds and allows for wider comparisons across effector systems and across sensory modalities. Notably, however, automatic imitation effects could only be used to indicate motor involvement (i.e., being active) during speech perception, but not necessarily motor modulations on speech perception that are demonstrated in the first approach.

In an early automatic imitation study by Kerzel & Bekkering (Experiment 1, 2000), participants were instructed to respond to a printed syllable prompt (e.g., "Ba") by saying that

syllable (e.g., /ba/) and to ignore the speaker who was simultaneously mouthing either a compatible (e.g., /ba/) or an incompatible (e.g., /da/) syllable in a silent distractor video. Compatible distractors elicited faster responses than incompatible distractors, and the observed compatibility effect was used to indicate the degree of automatic imitation. However, in this experiment, the observed effect could be confounded with or reduced to a *stimulus-stimulus compatibility* (SSC; see in-depth discussions in Section 3.1.1) effect emerging due to associations between distractors and syllable prompts. That is, the performance could be modulated by the relative perceptual/phonological correspondence between distractors and syllable prompts, in addition to or rather than the relative perceptuomotor correspondence between distractors and responses. To further confirm motor activation by the perception of speech distractors, Kerzel & Bekkering in Experiment 2 instructed participants to respond to symbol prompts (i.e., “##” and “&&”) that are not commonly associated with either syllable /ba/ or /da/ and observed a compatibility effect. Since there were associations between distractors and responses only, the SRC effect (i.e., automatic imitation) was suggested to occur because the perception of speech distractors triggered a motor process, which then interfered with response production.

However, what remains unclear is whether using linguistic prompts (e.g., printed syllables) in speech SRC tasks leads to the same automatic imitation effect as using non-linguistic prompts (e.g., printed syllables); that is, whether the associations between distractors and linguistic prompts leads to an SSC effect that then interacts with the SRC effect resulting from the associations between distractors and responses. It is important to determine whether speech SRC tasks using linguistic prompts measure the motor process triggered by the perception of speech distractors, because studies have been using syllable prompts to investigate automatic imitation of speech (e.g., Adank, Nuttall, Bekkering, & Maegherman, 2018; Jarick & Jones, 2009). Additionally, as discussed in Section 3.1.2 below, it is also important to differentiate speech SRC effects from SSC effects and thus to determine that automatic imitation of speech results from relative perceptuomotor correspondence, rather than higher-level conceptual correspondence, between distractors and responses. As explained in Section 3.1.3 below, SSC and SRC effects are proposed to occur at different processing stages when distractors are associated with prompts or responses respectively and thus the two types of compatibility effects behaviourally exhibit different time courses. Chapter 3 examines whether speech SRC tasks using linguistic versus non-linguistic prompts measure the same motor process triggered by the perception of speech distractors. To address this issue, Chapter 3 reports on two experiments that compared time courses of compatibility effects measured using speech SRC tasks where distractors were associated with responses and/or prompts.

Additionally, using moving lines to mimic lip movements during syllable articulation in distractor videos in Experiment 4, Kerzel & Bekkering (2000) observed no compatibility

effect and thus concluded that speech automatic imitation cannot be reduced to a spatial SRC effect. Rather, automatic imitation of speech exhibits an imitative SRC effect, in which participants' performance is affected by the relative correspondence between actions perceived from distractor stimuli and actions produced in responses. Later studies have also reported automatic imitation effects for vowels (Adank et al., 2018; Ghaffarvand Mokari, Gafos, & Williams, 2020) and for auditory, visual, and audiovisual speech (Adank et al., 2018; Galantucci, Fowler, & Goldstein, 2009; Jarick & Jones, 2009), suggesting the existence of both audiomotor and visuomotor links underlying a variety of speech actions.

Further evidence is provided by neuroimaging and brain stimulation research that has shown that perceiving speech activates the articulatory motor system used to produce speech. Studies applying single-pulse TMS to the primary motor cortex have found effector-specific motor facilitation during speech perception. Applying single-pulse TMS to the primary motor lip and hand areas, studies have observed enhanced lip MEPs when participants were passively perceiving visual or auditory speech, as compared to the condition when they were perceiving non-speech stimuli, and such speech-related motor facilitation was not observed for hand MEPs (Murakami, Restle, & Ziemann, 2011; Watkins, Strafella, & Paus, 2003). More importantly, Fadiga, Craighero, Buccino, & Rizzolatti (2002) found that passively hearing tongue-related speech sounds facilitated tongue MEPs relative to hearing lip-related speech sounds, and it was the other way around for lip MEPs as observed by Nuttall, Kennedy-Higgins, Hogan, Devlin, & Adank (2016). These studies have exhibited effector-specific sensorimotor mapping at the neurophysiological level for speech actions, and following Heyes et al. (2005), automatic imitation of speech is likely to result from this observation-induced motor facilitation.

Additionally, using fMRI, Wilson, Saygin, Sereno, & Iacoboni (2004) observed enhanced activation in the primary motor and premotor cortex when participants were passively listening to or articulating meaningless monosyllables, demonstrating sensorimotor response properties of the articulatory motor system. Further specificity of perception-induced motor activation has been indicated in the findings that passively listening to syllables involving lip articulation led to greater activation in the primary motor lip area than listening to syllables involving tongue articulation and vice versa (Pulvermüller et al., 2006). Moreover, in a series of fMRI studies, Hickok and colleagues have observed sensorimotor responses in the left inferior frontal gyrus and premotor cortex, and also in the area Spt, a left lateralized region in the posterior Sylvian fissure at the parietal-temporal boundary (e.g., Buchsbaum, Hickok, & Humphries, 2001; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Okada & Hickok, 2006, 2009; Venezia et al., 2016). All these three regions were repeatedly found to respond during passive perception of a variety of speech stimuli, either auditory or visual, and during covert speech articulation. In recent neural models of speech (see more details in

Section 2.2.2 below), these three regions are proposed to be important nodes in the neural network supporting sensorimotor integration during speech processing (Hickok & Poeppel, 2007). Importantly, the Spt was also found to respond when participants were hearing or covertly humming novel piano melodies (Hickok et al., 2003), but not when they were covertly playing these melodies manually (i.e., imagine playing the melodies on a piano; Pa & Hickok, 2008). Instead, the left anterior intraparietal sulcus was found to respond in the covert playing condition (Pa & Hickok, 2008), suggesting that the sensorimotor response properties of the Spt are vocal-tract-specific whereas those of the anterior intraparietal sulcus are hand-specific.

One study using theta-burst TMS (TBS) examined whether the left inferior frontal gyrus is involved in speech imitation (Restle, Murakami, & Ziemann, 2012). Different TBS protocols (continuous TBS [cTBS], intermittent TBS [iTBS], and intermediate TMS [imTBS]) have been shown to respectively reduce, enhance, and cause no change in the excitability of the primary motor cortex (Huang, Edwards, Rounis, Bhatia, & Rothwell, 2005). Applying these TBS protocols in separate sessions, Restle et al. examined the role of the left inferior frontal gyrus in modulating German speakers' ability to repeat Japanese sentences that they did not understand. Improvement was observed after iTBS, relatively less improvement after imTBS, and no improvement after cTBS, indicating the role of the inferior frontal gyrus in supporting speech imitation. However, such intentional imitation of foreign sentences might involve other cognitive processes such as working memory. Hence, rather than sensorimotor mapping, it could be other cognitive processes being modulated by different TBS protocols, which then led to different behavioural outcomes.

Similar to what has been observed for automatic imitation of manual gestures and its underlying mechanisms, research in speech science has also demonstrated that people tend to covertly imitate speech actions they perceive, which then affects their own articulation. Additionally, this imitative behaviour is suggested to result from perception-induced motor activation and is possibly mediated by the neural network that responds during the perception and production of actions involving the vocal tract.

### ***2.2.2 Theories and models of speech perception***

In speech science, articulatory motor involvement during speech perception has remained controversial because it raises questions about the nature of the objects of speech perception, which has been the focus of debate for a long time. Here, I will outline several theoretical accounts of speech perception, explain how the objects of speech perception are conceived in different perceptual and motor accounts, and then end with more recent neural models of speech processing that unite both perceptual and motor accounts.

The fact that communication is feasible through phone calls suggests the dominance of the auditory modality in speech perception. Linking and comparing speech perception with general auditory processing in both humans and other species, auditory theories (e.g., Diehl & Kluender, 1989; Diehl, Lotto, & Holt, 2004; Holt & Lotto, 2008) propose that the objects of speech perception are auditory/acoustic in nature. Specifically, it is proposed that speech recognition occurs in the auditory system as this system has evolved to handle various types of environmental sounds and that the recognised auditory representations are then used for higher level processing (e.g., lexical). However, the McGurk effect, which occurs when an auditory /ba/ is dubbed onto a visual /ga/ and people report hearing /da/ (McGurk & MacDonald, 1976), demonstrates strong visual modulations on the processing of auditory speech and therefore provides compelling evidence against a pure auditory account.

Indeed, visual modulations on auditory speech perception are not uncommon as we all have experienced the intelligibility benefits of audiovisual speech in noisy environments. Laboratory studies have also investigated how seeing corresponding orofacial movements improves comprehension of auditory speech masked in noise (e.g., Sumbly & Pollack, 1954). Considering the multimodal nature of perceived speech, the Fuzzy Logical Model of Perception (Massaro, 1998) proposes that visual inputs are integrated with auditory signals to achieve the least ambiguous result that matches an abstract prototype stored in memory and that the combination of imperfect auditory and visual speech signals helps yield better perception. Alternatively, however, when there is mismatch between auditory and visual signals, as in the case of the McGurk effect, an illusory percept would consequently arise. In contrast to the auditory theories, the Fuzzy Logical Model of Perception provides a multisensory account, which however cannot sufficiently explain the articulatory McGurk effect observed by Sams et al. (2005) that was discussed in the previous section. Questions then arise as to how and why the motor system supporting speech production is involved in speech perception.

The Motor Theory of Speech Perception (Galantucci, Fowler, & Turvey, 2006; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985) proposes that the production system is critically involved during speech perception because the objects of speech perception are the speaker's intended vocal tract gestures which are represented in the brain as invariant motor commands used to control articulator movements. The Motor Theory was initially proposed to solve the coarticulation problem in which the acoustic patterns of a speech sound are influenced by the sound before or after it. For example, the acoustic patterns of /d/ are different in the context of /di/ versus /du/ and this is because the vowel production occurs during the production of the preceding consonant. In other words, there is no one-to-one relation between acoustic patterns and perceived speech sounds for /d/. To solve this problem, the Motor Theory suggests that perceptual invariance of consonants in

coarticulation is achieved through the use of invariant articulatory gestures. For example, having the tongue tip in contact with the alveolar ridge to stop and then quickly release airflow, the gestures used to produce /d/ remain the same in either /di/ or /du/. However, how acoustic signals are converted into articulatory gestures through the proposed “innately specified” (Liberman & Mattingly, 1985, p. 3) perception-production link remains vague.

Furthermore, the Motor Theory’s hypothesis has been challenged by the findings that people use different tongue shapes to produce American /r/ so as to reduce acoustic variability in different phonetic contexts (e.g., Guenther et al., 1999). Hence, instead of one invariant gesture, multiple articulatory gestures are sometimes recruited for one speech sound. Additionally, findings challenging the Motor Theory in its strong form have been from studies showing the ability to perceive speech sounds in patients who have never acquired the ability to speak due to congenital disease (e.g., Bishop, Brown, & Robson, 1990) or in birds specifically trained to discriminate speech sounds (e.g., Lotto, Kluender, & Holt, 1997). Importantly, though findings from patient studies in particular indicate that the motor system supporting speech production is not critically required for or involved in speech perception, thus rejecting the Motor Theory in its strong form; these findings do not necessarily contradict the findings of motor involvement during speech perception.

In fact, as pointed out by Galantucci et al. (2006) and Holt & Lotto (2008), the perceptual (auditory or multisensory) account and the motor account do not have to be mutually exclusive and could together contribute to speech perception. Recent Dual-Route Models of speech processing include a ventral stream where sensory signals are mapped onto conceptual and semantic representations and a dorsal stream where sensory signals are mapped onto articulatory motor representations (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). In the model by Hickok & Poeppel, important nodes in the left lateralized dorsal stream include inferior frontal gyrus, premotor cortex, and the area Spt, all of which, as discussed above, are found to respond during speech perception and production. The dorsal stream is suggested to only have modulatory effects on speech perception when perceived stimuli are ambiguous (Hickok et al., 2011). Instead, its primary function is to support speech motor learning and speech motor control, and sensory feedback (both auditory and somatosensory) is crucial in guiding the tuning of speech gestures so that the sounds are accurately reproduced. Hence, similar to what is proposed in the associative account of imitation (discussed in Section 2.1.3), the perception-production link between corresponding auditory and motor representations of speech actions is established through sensorimotor experience during which people speak and immediately receive auditory feedback of their own speech. Additionally, such sensorimotor experience also contributes to the establishment of the link between corresponding somatosensory and motor representations of speech actions. However, since people do not receive somatosensory signals when perceiving others’ speech

but only when themselves are speaking, discussions of the link between somatosensory and motor representations of speech actions are out of the scope of this thesis that focuses on automatic imitation of perceived speech.

The situation is different for visual speech for speakers cannot directly see their own faces unless through mirrors or other optical surfaces. Alternatively, the link between corresponding visual and motor representations of speech actions could be built, as suggested by Ray & Heyes (2011), through a common stimulus which in this case are acoustic consequences of speech actions. In addition, the link could also be built through sensorimotor experience that occurs during social activities where imitation occurs between interactive partners or during group activities where different individuals perform synchronous actions. Particularly for imitation between individuals, laboratory studies have shown that people sometimes overtly but unintentionally imitate visual and/or auditory speech they perceive, and such imitation is indicated in changes in the perceivers' own speech, as measured from their articulator kinematics or voice spectra (Delvaux & Soquet, 2007; Gentilucci & Bernardis, 2007; Shockley, Sabadini, & Fowler, 2004). Thus, there seems to be abundant sources of sensorimotor experience in social life contributing to the development of the visuomotor link underlying speech.

### ***2.2.3 Sensory feedback in speech production***

No study has specifically investigated whether sensorimotor experience modulates the perception-production link underlying speech, but studies have observed changes in both perception and production after people repeatedly receive altered auditory feedback while speaking (see below). Thus, there are clues to the flexibility of the perception-production link and how sensorimotor experience may modulate this link. Here, I will first introduce sensorimotor adaptation studies, explain the underlying mechanism in accordance with speech production models, and then discuss the flexibility of the speech perception-production link observed in this line of research.

As mentioned earlier, the dorsal stream where sensorimotor integration occurs during speech processing is proposed to primarily support speech motor learning and speech motor control, and sensory feedback is crucially involved in guiding and monitoring articulatory movements (Hickok & Poeppel, 2007). The role of sensory feedback in speech production has been examined in studies using sensorimotor adaptation paradigms in which participants learn to adjust their production to compensate for real-time alterations in sensory feedback and the adjusted production persists after sensory feedback is returned to normal or masked in noise (e.g., Houde & Jordan, 1998, 2002; Jones & Munhall, 2000; Purcell & Munhall, 2006; Tremblay, Shiller, & Ostry, 2003). In seminal studies by Houde & Jordan (1998, 2002),

participants received altered auditory feedback in which formants they produced were shifted such that the vowel's perceived phonetic identity was changed and they gradually shifted their own production in a direction opposite to the auditory feedback change. Resulting from vocal tract resonances, formants are acoustic energy concentrations around specific frequencies, and vowels can be defined by first three formants (i.e.,  $F_1$ ,  $F_2$ , and  $F_3$  [Borden et al., 1994]). In Houde & Jordan, when participants' / $\epsilon$ / utterance in the word "pep" was altered (i.e., decreased  $F_1$  and increased  $F_2$ ) and ended up sounding more like /i/, they modified their utterance, producing a vowel sounding more like /a/ whose  $F_1$  is higher and  $F_2$  is lower than those of / $\epsilon$ /. Crucially, such modification in production was retained when auditory feedback was later masked in noise (or returned to normal in Purcell & Munhall, 2006).

Sensorimotor adaptation shown in Houde & Jordan (1998, 2002) indicates the importance of auditory feedback in speech motor control. In addition to vowel formants (Houde & Jordan, 1998, 2002; Purcell & Munhall, 2006; Schuerman, Nagarajan, McQueen, & Houde, 2017; Villacorta, Perkell, & Guenther, 2007), studies have also observed sensorimotor adaptation after shifting the vocal pitch during English vowel production (Jones & Munhall, 2000) or during Mandarin tone production (Jones & Munhall, 2002, 2005), or after modifying consonant features such as frication (e.g., modifying /s/ to sound more like /ʃ/ in Shiller, Sato, Gracco, & Baum, 2009) or voicing (e.g., modifying "tipper" to sound like "dipper" or vice versa in Mitsuya, MacDonald, & Munhall, 2014). Other research has also exhibited sensorimotor adaptation following somatosensory feedback perturbations, such as using a mechanical device to interfere with participants' jaw movement (Lametti, Nasir, & Ostry, 2012; Nasir & Ostry, 2006, 2009; S. Tremblay et al., 2003). Hence, in addition to auditory feedback, somatosensory feedback as well plays a crucial role in speech motor control.

The model of speech production, the Directions Into Velocities of Articulators (DIVA) model (Guenther, Ghosh, & Tourville, 2006; Tourville & Guenther, 2011), proposes that the goal of speech production is to generate a target sound, which is achieved through the collaboration between feedback and feedforward control systems. A motor control system only relying on sensory feedback would be problematic because sensory feedback is noisy and delayed due to environmental noise and neural processing respectively, whereas speech movements are rapid and require precise temporal control. Hence, an internal feedforward control system that predicts sensory outcomes of motor commands has been incorporated in the DIVA model and in other speech production models as well (e.g., Hickok, 2014; Houde & Nagarajan, 2011; Pickering & Garrod, 2013; Tian & Poeppel, 2010). According to the DIVA model, associations between motor commands and their sensory consequences are initially learned during infant babbling. Auditory targets are then learned through exposure to

native speech sounds and comparing them with one's own speech, and somatosensory targets are learned during the correction of self-production based on auditory feedback. Thus, sensory feedback is crucial in supporting speech acquisition. However, once sensory targets are learned, the feedforward system starts to play a more crucial role in speech motor control. During speech production, the feedforward mechanism sends both motor commands to motor regions for actual movement execution and sensory predictions of those motor commands to sensory areas. The predicted and actual sensory consequences are compared, and any discrepancy (i.e., errors) between the two will be used to correct motor commands and update the feedforward system. Therefore, though sensory feedback appears to be insufficient to support fluent speech, it is essential for updating the internal forward system when persistent errors occur due to permanent changes in articulators (e.g., body growth) or sudden sensory feedback alterations as in the case of sensorimotor adaptation.

According to the DIVA model, sensorimotor adaptation occurs because feedback alterations cause errors between predicted and actual sensory consequences. Hence, the corresponding motor commands are corrected in attempt to generate target sounds, and the feedforward system is accordingly updated as well after repeated productions. Particularly, the retained production change after removal of alterations has been suggested to indicate the existence of the feedforward system whose sensory predictions are modified by experience (Houde & Nagarajan, 2011; Pickering & Garrod, 2013). Focusing on production change following sensorimotor adaptation, the above-mentioned studies demonstrate that the speech production system remains flexible in adults and is subject to change when there is an alteration in sensory feedback (Houde & Jordan, 1998, 2002). However, what still remains unclear is whether there is perception change as well following sensorimotor adaptation, and if so, what drives this perception change. Several studies have evaluated participants' perceptual ability following sensorimotor adaptation and found a perception change that is related to the change in production rather than the alteration in auditory feedback (Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014; Schuerman et al., 2017; Shiller et al., 2009).

For example, while participants were reading a list of consonant-vowel or consonant-vowel-consonant words containing the initial fricative /s/ (e.g., "see"), Shiller et al. (2009) altered some participants' (i.e., the AF group) auditory feedback by shifting down the first spectral moments of the fricative towards /ʃ/ as in "she" and observed compensatory production. Additionally, participants also finished a phoneme identification test before and after sensorimotor adaptation to characterise their perceptual boundaries between /s/ and /ʃ/. The identification test employed synthetic speech stimuli differing from each other along a continuum from /s/ and /ʃ/ embedded within the carrier, "a shed" or "a said", and participants had to decide whether each heard stimulus started with /s/ or /ʃ/. The AF group exhibited sensorimotor adaptation in response to the auditory feedback shift, and results from the

identification test demonstrated that the group's perceptual boundary between /s/ and /ʃ/ moved towards the same direction as the auditory feedback shift (i.e., towards /ʃ/). In other words, while participants tried to produce an extreme /s/ by shifting up the first spectral moments of this fricative, their perceptual range of /s/ became wider and included some initially ambiguous sounds with relatively low first spectral moments. Importantly, no such perception change was observed in the passive listening group who received the same altered auditory stimuli as the AF group but did not produce any words. Thus, it was suggested that the perception change in the AF group was not due to auditory feedback alterations but was rather due to the change in production.

Similar findings of changes in both perception and production after sensorimotor adaptation have also been observed by Lametti, Rochet-Capellan, et al. (2014) who employed vowels and by Schuerman et al. (2017) who examined coarticulation. The authors of both studies have also suggested a motor contribution to the change in perception. Schuerman et al. proposed that sensorimotor adaptation is a type of sensorimotor experience and what happens following this experience is a sensorimotor remapping between a motor command used to compensate for altered auditory feedback and its acoustic outcome which itself is also altered. In other words, there is a certain degree of separation between corresponding motor commands and their acoustic outcomes that usually follow under natural conditions, such as when sensory feedback is not altered (Hickok, 2014), and a certain degree of flexibility in the perception-production link, allowing for mapping between noncorresponding motor commands and altered acoustic signals that do not naturally follow. Though these sensorimotor adaptation studies did not demonstrate how this sensorimotor experience modulates participants' imitative capacity, they have indicated that sensorimotor remapping between sensory and motor representations of different speech actions is possible. Hence, according to these findings, the perception-production link underlying speech actions seems to result from and is possibly subject to change through associative sensorimotor learning.

### **2.3 Summary and the thesis outline**

People demonstrate an imitative capacity to quickly map observed actions onto their motor repertoire, indicating the existence of a perception-production matching link between sensory and motor representations of the same actions. In controlled laboratory settings, the imitative capacity can be measured behaviourally using automatic imitation effects, in which responses are facilitated by the observation of the same action relative to the observation of a different action. Automatic imitation effects are suggested to occur due to perception-induced motor activation, with a larger effect indicating greater motor activation. Greater perception-induced

motor activation further indicates a stronger perception-production matching link presumably underlying the imitative capacity.

The origin of humans' imitative capacity and its proposed neural substrate, the human mirror neuron system (and mirror neurons themselves as well), has been the source of much debate. Various accounts have been proposed to explain how the imitative capacity emerges. One account that has been widely tested in laboratory settings is what I will refer to throughout this thesis as the *associative account*. This account proposes that the imitative capacity is acquired through sensorimotor experience during which concurrent observation and execution of the same actions occur and thus this capacity is experience dependent. Sources of sensorimotor experience are suggested to be abundant in social life. In other words, in addition to direct self-observation, people also acquire sensorimotor experience during social interactions and thus the same associative mechanism is proposed to be involved in imitation of both visually transparent manual gestures and visually opaque orofacial movements. Furthermore, sensorimotor experience has been shown to enhance mirror responses to the observation of newly learned novel movements, relative to sensory or motor experience only. Thus, the same associative mechanism is also proposed to be involved in imitation of both familiar and novel actions that were or were not part of the observers' motor repertoire.

Studies have observed automatic imitation effects for various manual gestures and also for speech actions perceived through different modalities, demonstrating that people imitate a wide range of actions they perceive. SRC tasks used to measure automatic imitation of speech vary across studies, and it thus remains to be further elucidated whether automatic imitation effects obtained using different speech SRC tasks measure the same motor process triggered by the perception of speech distractors. More importantly, sensorimotor training studies have demonstrated that mirror and counter-mirror training respectively enhances and reduces automatic imitation of manual gestures. Consistent with the associative account of imitation, these findings suggest that the matching properties of the perception-production link remain flexible in adults and are subject to change through sensorimotor training. However, sensorimotor training studies so far have only looked at familiar, visually transparent manual gestures that participants have observed and executed throughout their life. It thus remains to be further determined whether this associative mechanism is also involved in imitation of visually opaque orofacial movements such as speech and whether the associative mechanism is also engaged in imitation of novel actions that were not part of the observers' motor repertoire.

The overarching aims of this thesis are 1) to establish the role of sensorimotor experience in modulating the imitative capacity underlying communicative orofacial movements, namely speech actions, that are either familiar or novel to perceivers and 2) to elucidate at what level

experience-induced modification occurs in the perception-production link presumably underlying the imitative capacity. To achieve these aims, five experiments were conducted, each of which targeted one or two objectives outlined below:

- **Objective 1:** to determine whether and to which extent automatic imitation effects are confounded with SSC effects when distractors are associated with both prompts and responses in a speech SRC task (Experiment 1 in Chapter 3).
- **Objective 2:** to determine whether automatic imitation effects are confounded with – but cannot be reduced to – SSC effects when distractors are associated with both prompts and responses in a speech SRC task (Experiment 2 in Chapter 3 and Experiment 3 in Appendix A).
- **Objective 3:** to determine whether sensorimotor experience is critically involved in establishing the perception-production matching link underlying visually perceived speech actions (Chapter 4).
- **Objective 4:** to determine whether sensorimotor experience is critically involved in establishing the perception-production matching link underlying novel speech actions (Chapter 5).
- **Objective 5:** to explore at what level experience-induced modification occurs in the perception-production link (Chapter 5).

## **Chapter 3 The perceptual and the motor processes triggered by visual distractors in automatic imitation of speech**

### **3.1 Introduction**

As discussed in Section 2.2.2, the motor account of speech perception proposes that the articulatory motor system is critically involved during the perception of speech (Lieberman et al., 1967; Liberman & Mattingly, 1985), whereas the perceptual account does not make such an assumption (e.g., Diehl et al., 2004; Holt & Lotto, 2008; Massaro, 1998). As discussed in Section 2.2.1, using speech SRC tasks, Kerzel & Bekkering (2000) conducted a series of experiments to examine motor involvement in speech perception (see Table 3.1). In Experiment 1 in the series, participants responded to a syllable prompt “Ba” or “Da” by reading out that syllable and ignored the speaker who was mouthing either /ba/ or /da/ in a distractor video. It was predicted that, if there is motor activity involved in speech perception, an SRC effect (i.e., automatic imitation) should emerge in which performance is modulated by the relative correspondence between distractors and responses. Faster responses were observed for compatible trials than for incompatible trials. However, it was also possible that the performance was modulated by the relative correspondence between distractors and syllable prompts, thus resulting in an SSC effect (see detailed explanations below). To exclude this possibility, Kerzel & Bekkering in Experiment 2 replaced syllable prompts with symbol prompts “##” and “&&” to eliminate the associations between distractors and prompts while keeping the associations between distractors and responses. An SRC effect was observed in their Experiment 2, thus supporting the motor account of speech perception and indicating people’s tendency to imitate perceived speech.

However, as discussed in Section 2.2.1, it remains to be further elucidated whether using linguistic prompts (e.g., printed syllables) in speech SRC tasks leads to a different automatic imitation effect than using non-linguistic prompts (e.g., printed symbols); that is, it remains unknown whether the associations between distractors and linguistic prompts would give rise to an SSC effect that then interacts with the SRC effect resulting from the associations between distractors and responses. It is important to determine whether speech SRC tasks using linguistic prompts measure the motor process triggered by the perception of speech distractors, because studies have been using syllable prompts to investigate automatic imitation of speech (e.g., Adank et al., 2018; Jarick & Jones, 2009). In addition, as discussed in Section 3.1.2 below, it is also important to differentiate speech SRC effects from SSC effects and thus to determine that automatic imitation of speech results from relative perceptuomotor correspondence, rather than higher-level conceptual correspondence, between distractors and responses.

In the following sections, I will first explain how SSC and SRC effects are proposed to occur at different processing stages and thus respectively measure different processes triggered by the perception of speech distractors. I will explain how higher-level conceptual correspondence could possibly account for automatic imitation of speech and discuss potential problems with this conceptual assumption. I will then explain why and how SSC and SRC effects behaviourally demonstrate different time courses and how they could be used to investigate perceptual and motor processes triggered by perceived speech. Finally, I will introduce the current study that aimed to examine and compare time courses of compatibility effects measured using speech SRC tasks where distractors were associated with responses and/or prompts.

**Table 3.1** Task paradigms used by Kerzel & Bekkering (2000) and Kerzel (2002) to investigate SSC and SRC effects.

	Participant number	Printed Prompt	Video Distractor	Required Response	Compatibility Effect (ms)	
Kerzel & Bekkering (2000)						
Experiment 1	8	“Ba” & “Da”	/ba/ & /da/	Say /ba/ & /da/	SSC+SRC	35
Experiment 2	8	“##” & “&&”	/ba/ & /da/	Say /ba/ & /da/	SRC	42
Kerzel (2002)						
Letter condition	10	“Ba” & “Da”	/ba/ & /da/	Press left & right keys	SSC	49

*Notes.* In each task, participants were instructed to respond to a prompt by making an instructed response and to ignore a distractor.

### 3.1.1 Stimulus-response and stimulus-stimulus compatibility

As mentioned in Section 2.1.1, the Dimensional Overlap model proposes a dual-route account to explain SSC and SRC effects that presumably occur at different processing stages with the former occurring at a relatively earlier stimulus-identification stage and the latter at a response-production stage (Kornblum et al., 1990, 1999). Specifically, in a task where both prompts and distractors are employed, the presentation of a prompt activates an arbitrarily assigned response through a controlled process based on the instructions. Meanwhile, the presentation of a distractor automatically activates a corresponding response, if there is stimulus-response (S-R) dimensional overlap between distractors and responses, which is broadly defined as similarities in mental representations between the two. With S-R overlap, it means that there exist *excitatory links* through which the activation of a distractor automatically leads to the activation of the corresponding response. An SRC effect thus arises

from the interaction between the controlled and the automatic processes, with performance being facilitated when responses activated through the two processes are compatible and disrupted when the two activated responses are incompatible. In Experiment 2 in Kerzel & Bekkering (2000) where symbol prompts were used, automatic imitation occurred because seeing a speech action activates the corresponding motor representation of that action which is either compatible or incompatible with the one required in response, hence leading to facilitation or delay in performance respectively.

In contrast, stimulus-stimulus (S-S) overlap indicates excitatory links between corresponding distractors and prompts that are thought to represent different stimulus features. The presentations of distractors and prompts lead to activation of interacting compatible or incompatible stimulus features, producing either facilitation or interference respectively in SSC effects. Kerzel (2002) instructed participants to press a left or right key in response to a syllable prompt “Ba” or “Da” and to ignore the speaker who was mouthing either /ba/ or /da/ in a distractor video (see Table 3.1). Since manual responses are not commonly associated with any syllables (maybe not in people who use sign languages), an SSC effect was expected to result from S-S overlap between distractors and syllable prompts. Faster responses were found for compatible trials (e.g., a prompt “Ba” was presented when the speaker was mouthing /ba/ in a distractor video) than for incompatible trials (e.g., a prompt “Da” was presented when the speaker was mouthing /ba/ in a distractor video). The observed SSC effect was suggested to occur because performance was modulated by the relative phonological correspondence between distractors and syllable prompts. Accordingly, it seemed that perception-induced motor activation could not fully account for the compatibility effect observed in Experiment 1 in Kerzel & Bekkering (2000). Instead, following the Dimensional Overlap model, Kerzel suggested that both SSC and SRC effects might have contributed to the compatibility effect observed in Experiment 1 in Kerzel & Bekkering where both S-S and S-R overlap was present.

Interestingly, the SSC effect (49 ms) in Kerzel (2002) was comparable in size with the SRC effect (42 ms) in Experiment 2 in Kerzel & Bekkering (2000). Yet, a smaller compatibility effect (35 ms) was observed in Experiment 1 in Kerzel & Bekkering where distractors overlapped with both prompts and responses. That is, the effects of S-S and S-R overlap did not add up but turned out to be a smaller compatibility effect that I will refer to as an SSC+SRC effect in the following text. Part of the reason might be that responding to a syllable prompt by reading out that syllable in the SSC+SRC effect was much easier than responding to a symbol prompt by speaking out an arbitrarily assigned syllable in the SRC effect or responding to a syllable prompt by pressing an arbitrarily assigned key in the SSC effect. The overlapping syllable prompts and articulatory responses that were compatible in every trial in the SSC+SRC effect facilitated overall performance. Hence, participants were more resistant to the influence of distractors, resulting in a smaller combined SSC+SRC effect.

Response time data revealed overall faster responses for the SSC+SRC effect than for either the SSC or the SRC effect. However, because the above-mentioned SSC, SRC, and SSC+SRC effects (see Table 3.1) in Kerzel and in Kerzel & Bekkering were measured in separate experiments with small sample sizes ( $n \leq 10$ ), one has to take into account potential group differences when comparing these findings. To elucidate whether automatic imitation effects are confounded with SSC effects when both S-S and S-R overlap is present, Experiment 1 reported in this chapter adopted the SRC tasks used in Experiment 1 and 2 by Kerzel & Bekkering to examine SSC+SRC and SRC effects in a within-subject design.

### **3.1.2 *The multimodal nature of speech***

Following the Dimensional Overlap model, the previous section has focused on how theoretically SSC and SRC effects originate from S-S and S-R overlap respectively. However, it could also be argued that both SSC and SRC effects measured using speech SRC tasks might have resulted from conceptual compatibility mediated by higher-order representations and thus the two effects have the same origin. Using hand movements and moving dots as distractors, Jansson, Wilson, Williams, & Mon-Williams (2007, Experiment 2) observed compatibility effects for hand opening/closing responses in both conditions. The finding that symbolic, non-biological stimuli also elicited a compatibility effect seemed to undermine the hypothesis that automatic imitation occurs because action perception activates corresponding motor representations, which then interacts with responses cued by prompts. Instead, Jansson et al. proposed that seeing an opening hand or two dots moving apart activates the conceptual/semantic representation of “opening”, which then facilitates the production of a response associated with a compatible semantic representation and delays the production of a response with an incompatible semantic representation.

Following the assumption of conceptual compatibility, automatic imitation of speech occurs because seeing a syllable articulation activates the conceptual/phonological representation of that syllable, which then facilitates the production of a response with a compatible phonological representation and delays the production of a response with an incompatible phonological representation. For instance, in a speech SRC task where symbol prompts are used, the presentation of a distractor automatically activates the corresponding phonological representation. In addition, the presentation of a prompt activates the assigned phonological representation before activating the motor representation of a response. The automatic process triggered by a distractor and the controlled process triggered by a prompt interact with each other at the phonological level and have to reach an agreement at this level before continuing to activate the motor representation of a response. The same scenario could

be applied to the SSC effect which, as suggested by Kerzel (2002), occurs due to the relative phonological correspondence between distractors and syllable prompts.

Alternatively, as mentioned by Kerzel (2002), it is also possible that phonological information could be motor in nature, if one were to accept the Motor Theory of Speech Perception in its strong form – that is, any kind of speech perception is perceiving speech gestures. Consequently, it would suggest that SSC and SRC effects observed in speech SRC tasks have the same origin and should thus behave similarly. However, Kerzel also pointed out that this kind of assumption blurs the distinction between perception and action. More crucially, it simplifies the question by ignoring the multimodal nature of speech processing and trying to define an ‘ultimate’ form of speech within one modality. The assumption of conceptual compatibility bears the same problem, except that, rather than a pure motor account, it proposes a pure conceptual account. As mentioned in Section 2.2.2, the Dual-Route Model of speech proposes that perceived speech is processed in both ventral and dorsal streams, with the former involved in speech recognition and comprehension and the latter involved in perceptuomotor translation (Hickok & Poeppel, 2007). The Dual-Route Model provides the neural basis for multimodal speech processing and thus SSC and SRC effects observed from speech SRC tasks may have different origins and behaviourally measure different processes triggered by the perception of speech distractors.

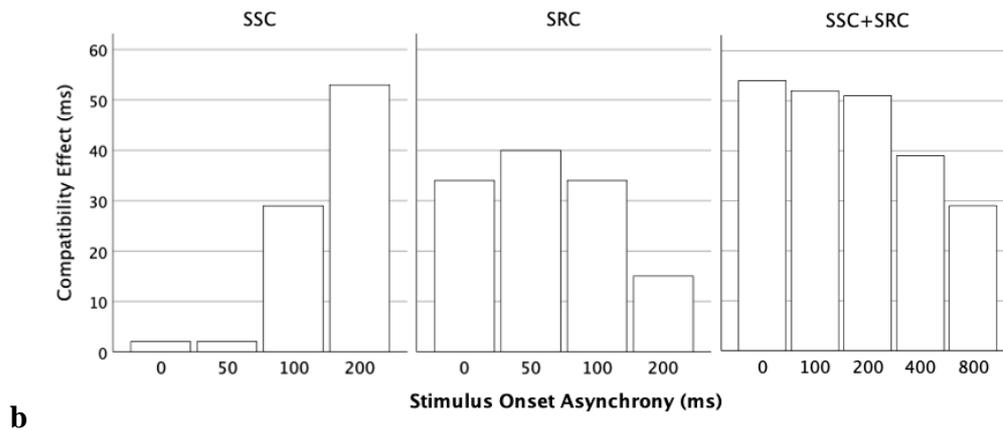
Experiment 2 reported in this chapter further determined whether automatic imitation effects are confounded with, but cannot be reduced to, SSC effects when both S-S and S-R overlap is present. Findings of such would suggest that, consistent with predictions derived from the Dual-Route Model of speech processing, the perception of speech distractors in SRC tasks triggers both perceptual and motor processes, which then interferes with participants’ performance when there is S-S or S-R overlap respectively. Modifying the SRC tasks used by Kerzel & Bekkering (2000) and by Kerzel (2002), Experiment 2 gave participants four different SRC tasks to examine SSC, SRC, and SSC+SRC effects in a within-subject design (see Section 3.1.4 for more details of the experiment design). Additionally, Experiment 2 also explored the possibility of using different speech SRC tasks to investigate multimodal speech processing, i.e., audiovisual versus visuomotor.

### ***3.1.3 Time courses of compatibility effects***

As discussed in Section 3.1.1, SSC and SRC effects are assumed to result from the interaction between a controlled and an automatic process when there is S-S or S-R overlap respectively. Following this assumption, the Dimensional Overlap model further proposes that the two effects occur at different processing stages and thus exhibit different time courses (see Figure 3.1). To summarise the main arguments first, SSC effects are theoretically proposed to occur

at an early stimulus-identification stage and behaviourally expected to be more prominent for late responses; in contrast, SRC effects are proposed to occur at a relatively late response-production stage and expected to be more prominent for early responses. That is to say, SSC and SRC effects can be distinguished based on behavioural results according to the Dimensional Overlap model, and detailed explanations are as follows.

**a** For Figure 3.1a, please see Figure 7 in Kornblum et al. (1999).



**Figure 3.1 a** Figure 7 from Kornblum et al. (1999). Demonstrations of how SSC (left) and SRC (right) effects change across SOAs. Light and dark grey rectangles were further added to respectively mark the stimulus-identification and the response-production stages proposed in the Dimensional Overlap model. **b** Illustrations of time courses of SSC (left) and SRC (right) effects reported in Experiment 1 by Kornblum et al. and SSC+SRC (right) effects in Experiment 2 also by Kornblum et al.

Specifically, in the controlled process where the presentation of a prompt leads to the production of an assigned response, the Dimensional Overlap model postulates a stage-like processing pattern in which a stimulus-identification stage precedes a response-production stage. When there is S-S overlap in a task, SSC effects are assumed occur at the stimulus-identification stage, where perceptual processing occurs and the motor system has not got involved yet; in contrast, when there is S-R overlap, SRC effects occur at the response-production stage, where the motor system is involved in response selection and execution. In other words, SSC and SRC effects respectively measure the perceptual and the motor processes triggered by the presentation of an overlapping distractor, with a larger effect indicating greater activation in the corresponding process.

Importantly, the activation of a distractor in the automatic process is proposed to first increase and then decay because of participants' inattention. As a result, the extent of the interaction between the controlled and the automatic processes changes over time for either S-S or S-R overlap. Furthermore, because SSC and SRC effects are assumed to occur at different stages, the two effects consequently exhibit different time courses. The time course of a compatibility effect can be examined by varying the time delays (i.e., stimulus onset

asynchronies [SOAs]) between presentations of distractors and prompts; that is, by varying the temporal relationship between the automatic and the controlled processes (Hommel, 1997; Kornblum et al., 1999). Kornblum et al. in Experiment 1 designed two similar tasks that measured different effects. In the task measuring SSC effects, participants had to press the left or right key in response to colour patches green or blue that might appear on the upper or lower half of the screen; in the meantime, they had to ignore colour patch positions and colour words “GREEN” or “BLUE”. Hence, there was semantic S-S overlap between colour patches and colour words. In the task measuring SRC effects, participants had to press the left or right key in response to colour patches green or blue that might appear on the left or right half of the screen; in the meantime, they had to ignore colour patch positions and non-colour words “DETAIL” or “NOVEL”. Hence, there was spatial S-R overlap between colour patch positions and response key positions. Using the same set of SOAs (i.e., 0, 50, 100, and 200 ms) in both tasks, Kornblum et al. observed the largest SSC effect at 200-ms SOA in the task with S-S overlap only and the largest SRC effect at 50-ms SOA in the task with S-R overlap only (see Figure 3.1b). Their findings demonstrated that SSC effects were larger for responses at longer SOAs (i.e., late responses relative to the distractor onset) and SRC effects were larger for responses at shorter SOAs (i.e., early responses relative to the distractor onset). Kornblum et al. concluded that “the effects of S-S overlap slowly rose and then fell, whereas the effects of S-R overlap displayed an extremely rapid rise, followed by a fall [, and] the effects of S-S and S-R overlap thus appear to follow very different time courses (p. 710).” In other words, SSC effects developed across SOAs at a slower pace and reached maximum later (i.e., at a longer SOA) than SRC effects.

In addition to Kornblum et al. (1999), such distinctive time courses for SSC and SRC effects have also been observed in other studies (e.g., De Jong et al., 1994; Hommel, 1997; Kornblum, 1994). In the following text, I will explain the development of SSC and SRC effects together across different SOAs, but readers should be aware that this applies to the situation when SSC and SRC effects are measured alone in separate tasks. Putting two effects together to explain helps better understand why SSC effects develop across SOAs at a slower pace than SRC effects. In the following paragraphs, I will then explain what happens when SSC and SRC effects are combined in one task. According to the Dimensional Overlap model, when the automatic and the controlled processes occur simultaneously (see 0-ms SOA in the top panel in Figure 3.1a), SSC effects that result from the interaction between the two processes at the early stimulus-identification stage are expected to be small because the distractor activation level (i.e., the activation in the automatic process) is low and just starts increasing; in contrast, SRC effects that result from the interaction at the late response-production stage are expected to be relatively large because the distractor activation level should have increased or even peaked by then. When the automatic process starts little earlier

than the controlled process (see medium SOA in the middle panel in Figure 3.1a), SSC effects at this SOA are expected to be larger than SSC effects at the 0-ms SOA because the distractor activation level has increased; in contrast, SRC effects at this SOA are expected to be smaller than SRC effects at the 0-ms SOA because the distractor activation level has decreased. When the automatic process starts considerably earlier than the controlled process (long SOA in the bottom panel in Figure 3.1a), both SSC and SRC effects are expected to be small because the interaction between the two processes at either stage is likely to occur after the distractor activation decays. Therefore, compared to SRC effects, SSC effects result from an interaction that occurs at an earlier stage and thus will coincide with the greatest activation level in the automatic process at a longer SOA. As a result, SSC effects develop more slowly across SOAs and reach maximum at a longer SOA than SRC effects.

Measuring SSC, SRC, and SSC+SRC effects across the same set of SOAs, experiments reported in this chapter aimed to illustrate and examine time courses of these compatibility effects in a within-subjects design. It was expected that, in Experiment 2, SSC effects develop more slowly across SOAs and reach maximum at a longer SOA than SRC effects. The time course of SSC+SRC effects will be explained in the following paragraphs.

To examine how S-S and S-R overlap affects each other when both are present in one task, experimental psychologists usually combine them factorially, meaning that two S-S compatibility levels and two S-R compatibility levels are combined into four conditions: S-S compatible + S-R compatible, S-S compatible + S-R incompatible, S-S incompatible + S-R compatible, and S-S incompatible + S-R incompatible (Hommel, 1997; Kornblum et al., 1999). For example, Kornblum et al. in Experiment 2 instructed participants to press the left or right key in response to colour patches green or blue that might appear on the left or right half of the screen; in the meantime, they had to ignore colour patch positions and colour words “GREEN” or “BLUE”. Thus, there was semantic S-S overlap between colour patches and colour words and spatial S-R overlap between colour patch positions and response key positions. Interestingly, when S-S and S-R overlap are combined factorially in one task, SSC and SRC effects do not simply add up but rather interact with each other. That is, SRC effects are larger for S-S compatible trials than for S-S incompatible trials. This is because S-S compatible trials advance the occurrence of the response-production stage relative to the S-S neutral condition, and thus SRC effects in this condition are likely to increase; whereas S-S incompatible trials delay the occurrence of the response-production stage, and thus SRC effects in this condition are likely to decrease.

However, the task used to measure SSC+SRC effects in Experiment 1 in Kerzel & Bekkering (2000) did not have S-S and S-R overlap combined factorially. Instead, there were only two compatibility conditions: S-S compatible + S-R compatible, and S-S incompatible + S-R incompatible. Following the logic discussed in the previous paragraph, S-S compatible

trials would advance the occurrence of the response-production stage relative to the S-S incompatible trials and thus S-R compatible trials would lead to more facilitation. However, S-S incompatible trials would delay the occurrence of the response-production stage and thus S-R incompatible trials could lead to less interference. In other words, it is difficult to compare the size of SSC+SRC effects with that of either SSC or SRC effects. Instead, the time course of SSC+SRC effects is expected to differ from the time course of either SSC or SRC effects. Figure 3.1b demonstrates the time course of SSC+SRC effects (i.e., [S-S incompatible + S-R incompatible] – [S-S compatible + S-R compatible]) observed in Experiment 2 in Kornblum et al. (1999), indicating a relatively flatter development curve across SOAs than that of SSC or SRC effects observed in Experiment 1 in the same paper. Previous studies (Hommel, 1997; Kornblum et al., 1999) have suggested that, when both S-S and S-R overlap is present, S-S overlap contributes to the compatibility effects at longer SOAs and S-R overlap contributes to the compatibility effects at shorter SOAs, thus giving rise to a relatively flatter developmental curve than that of SSC or SRC effects.

Experiment 1 in this chapter examined time courses of SRC and SSC+SRC effects in a within-subjects design and expected SSC+SRC effects to develop across SOAs more slowly and reach maximum at a longer SOA than SRC effects. Additionally, Experiment 2 in this chapter examined time courses of SSC, SRC, and SSC+SRC effects in a within-subject design and expected SSC+SRC effects to develop across SOAs more quickly than SSC effects and more slowly than SRC effects and probably, reach maximum at a shorter SOA than SSC effects and at a longer SOA than SRC effects.

#### **3.1.4 The current study**

Two experiments reported in this chapter aimed to examine SSC, SRC, and SSC+SRC effects and their time courses in a within-subjects design. Specifically, as mentioned in Section 2.3, two objectives of the current study are:

- **Objective 1:** to determine whether and to which extent automatic imitation effects are confounded with SSC effects when distractors are associated with both prompts and responses in a speech SRC task (Experiment 1).
- **Objective 2:** to determine whether automatic imitation effects are confounded with – but cannot be reduced to – SSC effects when distractors are associated with both prompts and responses in a speech SRC task (Experiment 2).

Combining the first two experiments by Kerzel & Bekkering (2000), Experiment 1 examined time courses of SRC and SSC+SRC effects. Experiment 2 examined time courses of SSC,

SRC, and SSC+SRC effects. Previous research has found that the effects of S-S and S-R overlap interact with each other and that this interaction is more likely to be detected when the S-S overlap by itself elicits a large SSC effect (Hommel, 1997). That is, the effect of S-S overlap has to be potent enough to modulate the effect of S-R overlap. The McGurk effect demonstrates robust visual modulations on the processing of auditory speech, suggesting strong links between auditory and visual speech (McGurk & MacDonald, 1976). Hence, in Experiment 2, visual prompts were replaced with auditory prompts, i.e., pure tones and spoken syllables, in an attempt to obtain larger SSC effects. Participants in Experiment 2 were instructed to look at the speaker's face while responding to auditory prompts. Since there was no visual prompt to force participants to look at the speaker's face, an eye-tracker was used to monitor their eye movements.

### **3.2 Experiment 1**

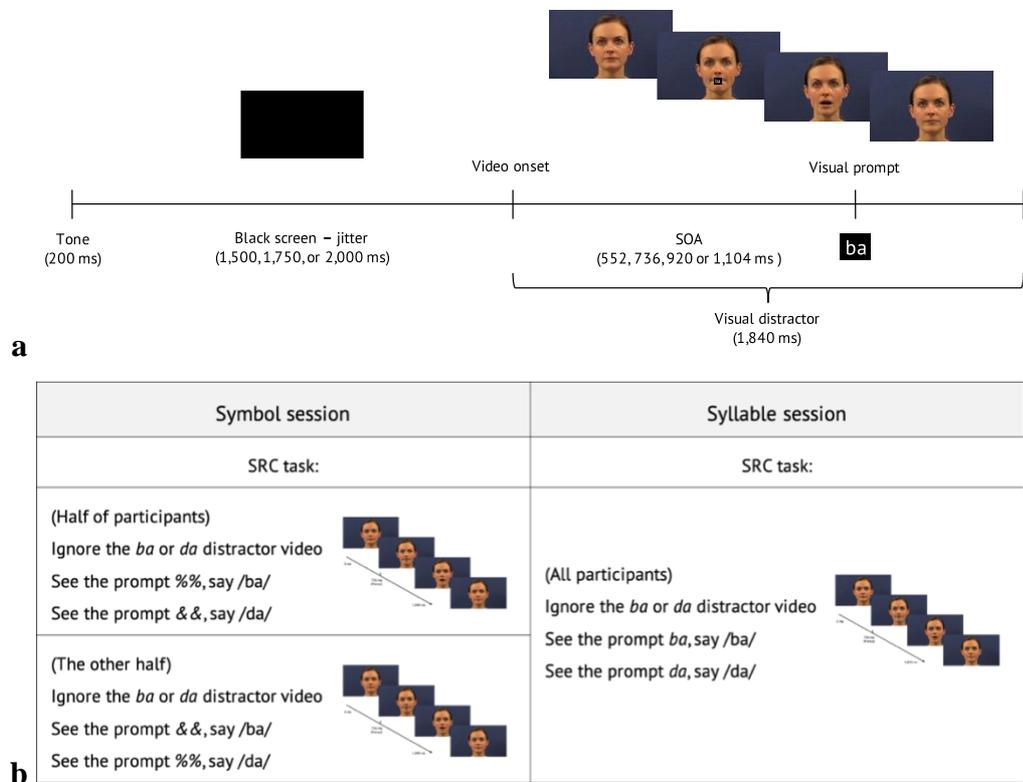
The aim of Experiment 1 was to examine time courses of SRC and SSC+SRC effects in a within-subjects design and to thus determine whether and to which extent automatic imitation effects are confounded with SSC effects when both S-S and S-R overlap is present (Objective 1). Specifically, this experiment compared the effects of using symbol versus syllable prompts in speech SRC tasks. Participants were instructed to say /ba/ or /da/ in response to either symbol (“%%” or “&&”) or syllable (“ba” or “da”) prompts in separate tasks while ignoring the speaker who was mouthing either /ba/ or /da/ in distractor videos. SRC effects were expected to result from S-R overlap between distractors and responses only in the symbol condition. SSC+SRC effects were expected to result from both S-S overlap between distractors and prompts and S-R overlap between distractors and responses in the syllable condition. Importantly, in the syllable condition, though there was also overlap between prompts and responses, the two were always compatible with each other (i.e., speaking out the prompted syllable) and hence no compatibility effect was expected to result from this overlap. Four SOAs were used to demonstrate time courses of SRC and SSC+SRC effects. It was expected that, if the perception of speech distractors triggers both perceptual and motor processes (i.e., the dual-route hypothesis of speech), SRC effects in the symbol condition would develop across SOAs more quickly and reach maximum at a shorter SOA than SSC+SRC effects in the syllable condition. This finding would then further suggest that automatic imitation effects are confounded with SSC effects when both S-S and S-R overlap is present. Detailed predictions and corresponding statistical outcomes are given in the following section.

### 3.2.1 Method

#### Participants

Thirty-one participants were recruited, but one was excluded for not finishing the study due to technical problems. Hence, the final analysis included 30 participants (24 female;  $M_{\text{age}} = 22.93$  years,  $SD_{\text{age}} = 3.68$ ,  $\text{range}_{\text{age}} = 18\text{--}33$ ). All were native British English speakers with self-reported normal or corrected-to-normal vision, normal hearing, and no speech or language disorders or other neurological disorders. Participants received £7.5 or course credit. The University Research Ethics Committee approved the procedures, and all participants gave written informed consent.

#### Material and stimuli



**Figure 3.2 a** Schematic timeline of a trial presented in the syllable sessions in Experiment 1. The visual prompt *ba* is positioned extending from the speaker's bottom lip to her top lip. **b** The Experiment 1 procedure with a symbol and a syllable SRC task session. The order of two sessions was counterbalanced across participants.

Stimuli (see Figure 3.2a) included silent distractor videos of a speaker saying /ba/ and /da/, printed symbol prompts %% and &&, and printed syllable prompts *ba* and *da*. The videos (25 frame per second) were filmed with a Canon Legria HF G30 video camera, edited in iMovie on an Apple iMac, and scaled down in resolution to 1,280 x 720 in AVI format to fit the experimental program's requirements. A female native British English speaker was shown in

the videos from her neckline upward in colour on a blue background, and her nose was approximately centred on the screen. Both videos lasted 1,840 ms and started and ended with the speaker's mouth closed in a resting configuration. At 552 ms, the speaker's mouth began to move in the *da* video, and her lips began moving towards each other in the *ba* video. The last frame before the release of airflow (i.e., the release of a consonant burst) stopped by lips in the *ba* video and by tongue and teeth in the *da* video was around 736 ms, and vowel articulation commenced around 920 ms. The speaker was still articulating the vowel at 1,104 ms in both videos, and her facial expression returned to its resting position at 1,400 ms. Participants' viewing was unrestrained at a distance of 60 cm from the screen, and the speaker's face was shown in 14.34° x 11.14° of visual angle. In both videos, the speaker's mouth was shown in 0.99° x 3.81° of visual angle at its resting configuration and in 2.03° x 3.23° during the vowel articulation. The most salient visual difference between two videos was before consonant bursts, with the speaker's mouth shown in 0.46° x 3.81° in the *ba* video and in 1.51° x 3.59° in the *da* video.

The printed prompts *ba*, *da*, %%, and && were 300 dpi JPEG images printed in white boldfaced Arial font on a black background and shown in 0.38° x 0.38° of visual angle. To make participants focus on the speaker's lip area, the prompts were positioned extending from the speaker's bottom lip to her top lip. The prompt was presented at one of four SOAs (552, 736, 920 or 1,104 ms relative to video onset; see Figure 3.2a) in each trial. In both videos, SOAs were equally spaced across the video duration and respectively coincided with the start of the speaker's lip movement, the release of burst, the onset of vowel articulation, and the time point when the vertical extent of the speaker's mouth was at the maximum. The experiment was performed using Presentation (Version 18.0, Neurobehavioral Systems).

### ***Design and procedure***

The experiment included two sessions (symbol versus syllable; see Figure 3.2) and took place in a soundproofed, light-controlled booth. Participants were facing the computer monitor at a distance of 60 cm, and the microphone was placed 10 – 15 cm in front of participants' mouth and adjusted individually so that the microphone was not obstructing their vision. Written instructions were presented on the PC monitor before each session commenced.

In the symbol session where SRC effects were expected, participants were instructed to speak out /ba/ or /da/ as soon as they saw a symbol prompt (%% or &&) and to ignore the speaker's articulation (*ba* or *da*) in a distractor video. Hence, there were two compatibility conditions: a compatible condition where the speaker's articulation in a distractor video matched the participants' response, and an incompatible condition where the speaker's articulation differed from the response. Half of participants were instructed to say /ba/ in

response to the %% prompt and say /da/ to the && prompt, and the other half were assigned with the reverse mapping.

Each trial started with a 200-ms tone with a frequency of 500 Hz at 70 dB of sound pressure level played through Sennheiser HD25-SP II headphones. The screen remained black for one of three jittered durations (1,500, 1,750 or 2,000 ms) that were included to reduce the trial onset's predictability. One of the distractor videos was then presented. The prompt was presented at one of the four SOAs (552, 736, 920, or 1,104 ms relative to the video onset) for 200 ms, and the screen went black at the end of the video. There were six blocks with 40 trials each (240 trials in total). Forty-eight trial types (2 Prompts x 2 Distractors x 4 SOAs x 3 Jitters) were repeated five times in a randomised order. Ten practice trials were randomly chosen and given before the first block, and the session lasted about 20 min.

The syllable session where SSC+SRC effects were expected was identical to the symbol session, except that participants were instructed to speak out /ba/ or /da/ as soon as they saw the corresponding syllable prompt *ba* or *da*. The order of two sessions was counterbalanced across participants. The whole experiment with two sessions lasted about 40 min.

### ***Data recording and analysis***

Responses were recorded via a voice-key using a RØDE NT1-A Condenser Microphone and a Focusrite Scarlett 2i4 USB Computer Audio Interface pre-amplifier plugged into the sound card input of a Dell PC at 44.1kHz with 16 bits. Audio recording started at the video onset and lasted for 3,000 ms. The voice-key was triggered when the system detected an audio input at .2 of Presentation's total range. Response times (RTs) were measured relative to the prompt onset. If no response was detected, a warning saying *No response given* would be presented for 500 ms. In addition, a warning saying *Response too early* would appear if participants responded within 200 ms of the prompt onset.

Responses were checked manually using Praat (Boersma & Weenink, 2018). Errors included incorrect responses and missed trials. In Kerzel & Bekkering (2000), trials with RTs < 100 ms or > 1,200 ms were defined as errors for they were likely to be anticipatory or neglected responses respectively. The same 100-ms criterion for anticipatory responses was adopted in the current experiment. Additionally, to establish a criterion that could be applied to SRC tasks requiring either manual or vocal responses, trials with RTs more than three standard deviations away from corresponding condition means per syllable response were considered as outliers and defined as errors. Therefore, errors in the current experiment included incorrect responses, missed trials, anticipatory responses, and outliers. A natural log-

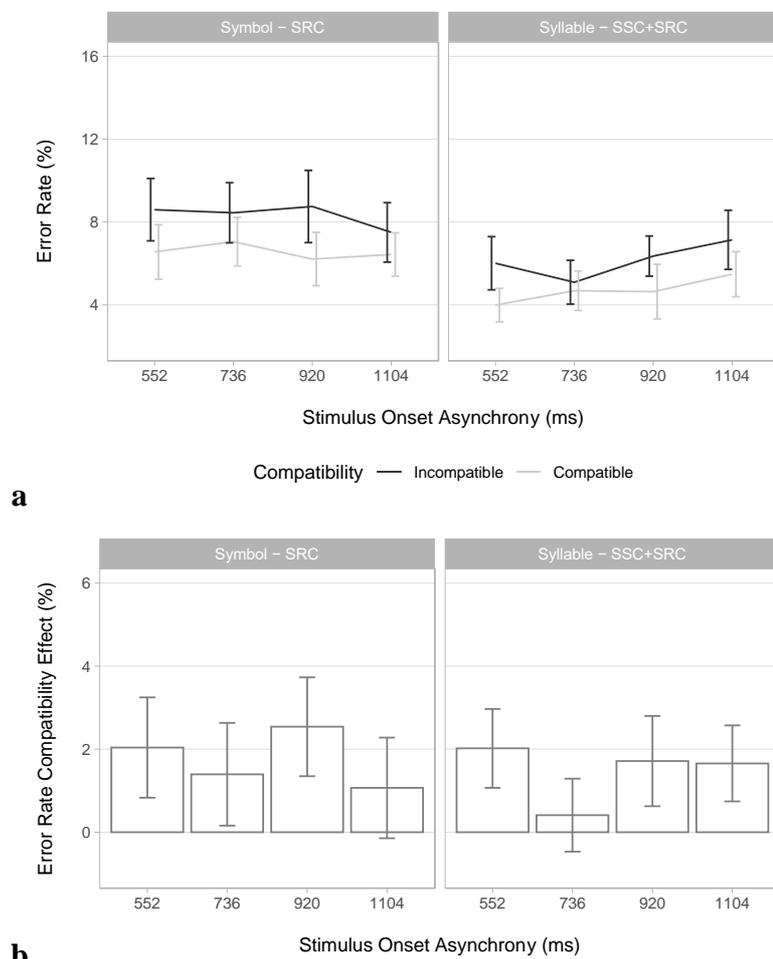
transformation was applied to RTs for statistical analyses but figures present back-transformed RTs.

Error rates (ERs, i.e., proportions of errors) and RTs were subjected to separate repeated-measures analysis of variance (ANOVA) with prompt type (symbol versus syllable), compatibility (compatible versus incompatible), and SOA (552, 736, 920 or 1,104 ms, also referred to as SOA1, 2, 3, and 4) as within-subjects variables. The significance level was set to  $p < .05$ . Greenhouse-Geisser correction for non-sphericity and Bonferroni-correction for multiple comparisons were applied whenever appropriate. Expected statistical outcomes for each prediction for Objective 1 are listed below:

- **Prediction 1:** following the dual-route hypothesis that the perception of speech distractors triggers both perceptual and motor processes, SRC effects were expected to reach maximum at a shorter SOA than SSC+SRC effects, suggesting that automatic imitation effects are confounded with SSC effects in the syllable condition – *the interaction between prompt type x compatibility x SOA should be significant, and follow-up t tests should reveal the largest SRC effect in the symbol condition at a shorter SOA than the largest SSC+SRC effect in the syllable condition.*
- **Prediction 2:** following the motor hypothesis that the perception of speech distractors triggers a motor process only, SRC and SSC+SRC effects were expected to demonstrate the same time course, suggesting that no SSC effects emerge in the syllable condition but the same automatic imitation effects arise in both symbol and syllable conditions – *the interaction between prompt type x compatibility x SOA should not be significant.*
- **Prediction 3:** following the perceptual hypothesis that the perception of speech distractors triggers a perceptual process only, SRC effects were expected to reach maximum at a shorter SOA than SSC+SRC effects, suggesting that automatic imitation effects can be reduced to pure SSC effects in the syllable condition – *the interaction between prompt type x compatibility x SOA should be significant, and follow-up t tests should reveal the largest SRC effect in the symbol condition at a shorter SOA than the largest SSC+SRC effect in the syllable condition.*
- **Prediction 4:** following the conceptual hypothesis that the perception of speech distractors triggers a higher-level conceptual process only, SRC and SSC+SRC effects were expected to demonstrate the same time course, suggesting that both automatic imitation and SSC effects result from the relative conceptual correspondence between distractors and responses/prompts – *the interaction between prompt type x compatibility x SOA should not be significant.*

### 3.2.2 Results

#### Error rate analysis



**Figure 3.3 a** ERs  $\pm$  standard errors (SEs) in each experimental condition. Two panels represent the symbol (left) and the syllable (right) sessions where SRC and SSC+SRC effects were expected respectively. **b** ER compatibility effects (i.e., incompatible - compatible)  $\pm$  SEs at each SOA for each prompt type.

**Table 3.2** Three-way ANOVA summary for ERs, as a function of prompt type, compatibility, and SOA.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	$Epsilon$	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Prompt	1	29	1.00	488.86	997.69	14.21	.001	.329
Compatibility	1	29	1.00	309.12	859.51	10.43	.003	.265
SOA	3	87	0.89	9.59	1116.54	0.25	.862	.009
Prompt x Compatibility	1	29	1.00	2.89	753.12	0.11	.741	.004
Prompt x SOA	3	87	0.86	86.17	1642.55	1.52	.215	.050
Compatibility x SOA	3	87	0.90	30.15	1425.08	0.61	.608	.021
Prompt x Compatibility x SOA	3	87	0.83	12.13	1126.05	0.31	.816	.011

*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of

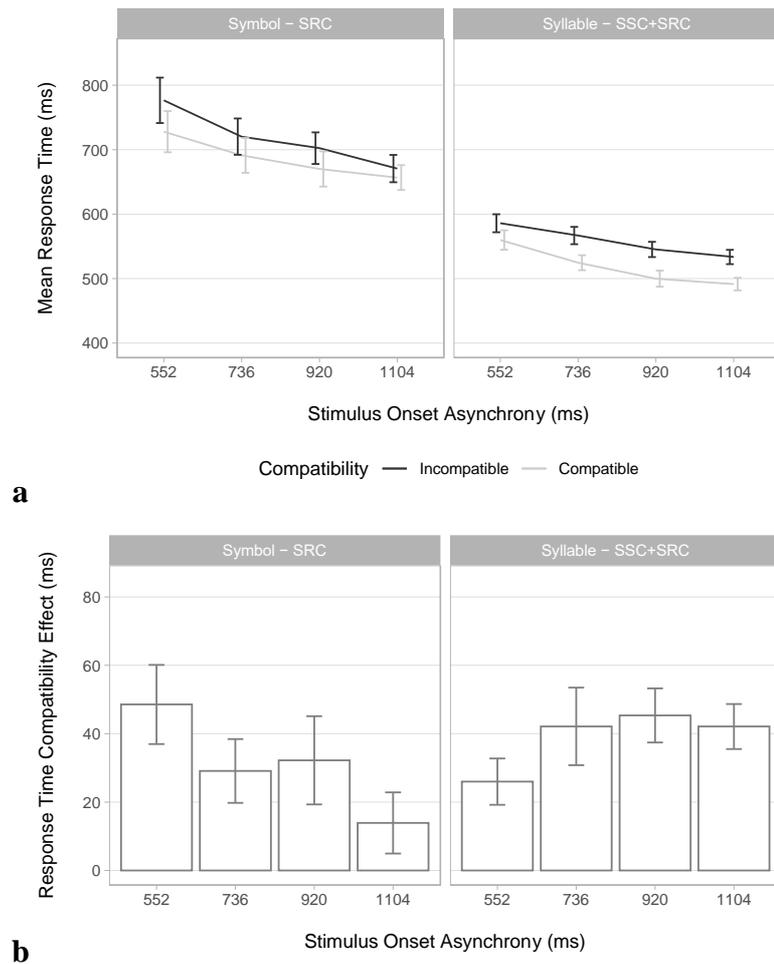
freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

On average, participants made 6.9% errors (incorrect responses: 3.1%; missed trials: 0.7%; anticipatory responses: 2.0%; outliers: 1.1%). ER analyses (see Figure 3.3, Table 3.2, and Table 3.3) revealed a main effect of prompt type, with a higher ER for symbol prompts than for syllable prompts. The main ER compatibility effect revealed a higher ER for incompatible trials than for compatible trials ( $\Delta = 1.61$ ). Neither the main effect of SOA nor any interaction was significant.

**Table 3.3** ERs, SEs, and 95% confidence intervals (CIs) for each level within prompt type and compatibility.

	Level	Mean	SE	95% CI
Prompt	Symbol	7.44	1.16	[5.08, 9.81]
	Syllable	5.42	0.93	[3.53, 7.32]
Compatibility	Incompatible	7.24	1.16	[4.87, 9.60]
	Compatible	5.63	0.92	[3.76, 7.51]

## Response time analysis



**Figure 3.4 a** Mean RTs ± SEs in each experimental condition. Two panels represent the symbol (left) and the syllable (right) sessions where SRC and SSC+SRC effects were expected respectively. **b** RT compatibility effects ± SEs at each SOA for each prompt type.

**Table 3.4** Three-way ANOVA summary for log-transformed RTs, as a function of prompt type, compatibility, and SOA.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	$Epsilon$	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Prompt	1	29	1.00	7.74	1.63	137.99	< .001	.826
Compatibility	1	29	1.00	0.41	0.16	76.32	< .001	.725
SOA	1.79	51.88	0.60	0.84	0.53	45.86	< .001	.613
Prompt x Compatibility	1	29	1.00	0.03	0.12	6.21	.019	.176
Prompt x SOA	2.40	69.52	0.80	0.00	0.24	0.53	.626	.018
Compatibility x SOA	3	87	0.84	0.01	0.22	0.87	.460	.029
Prompt x Compatibility x SOA	3	87	0.95	0.02	0.23	3.10	.031	.097

*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator.  $Epsilon$  indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.

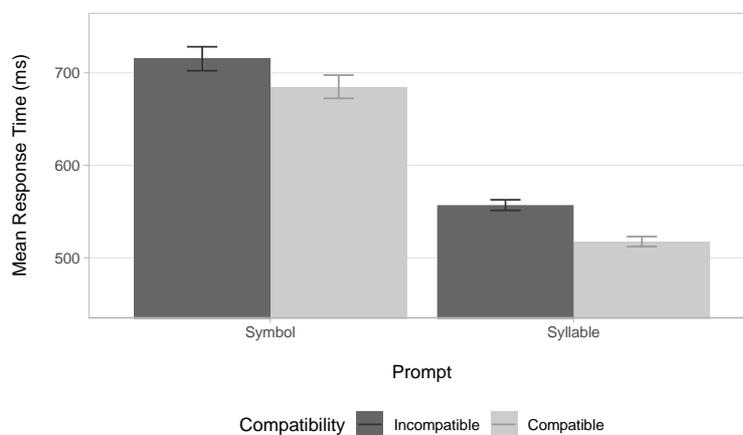
$SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

After errors were removed, log-transformed RT analyses (see Figure 3.4, Table 3.4, and Table 3.5) revealed a main effect of prompt type, with slower RTs for symbol prompts than for syllable prompts. The main RT compatibility effect revealed slower RTs for incompatible trials than for compatible trials ( $\Delta = 36$  ms). Follow-up  $t$  tests for the main effect of SOA revealed faster RTs for longer SOAs (all  $p < .05$ ), except that the difference between SOA3 and SOA4 was not significant.

**Table 3.5** Mean log-transformed RTs, back-transformed RT, SEs, and 95% CIs for each level within prompt type, compatibility, and SOA.

	Level	Mean	RT (ms)	SE	95% CI
Prompt	Symbol	6.53	688	0.03	[6.47, 6.60]
	Syllable	6.28	534	0.02	[6.24, 6.32]
Compatibility	Incompatible	6.44	624	0.03	[6.38, 6.49]
	Compatible	6.38	588	0.03	[6.33, 6.43]
SOA (ms)	552	6.47	645	0.03	[6.41, 6.53]
	736	6.42	612	0.03	[6.36, 6.47]
	920	6.38	591	0.03	[6.33, 6.43]
	1104	6.36	578	0.02	[6.32, 6.40]

The two-way interaction between prompt type and compatibility was significant (see Figure 3.5 and Table 3.6), and follow-up  $t$  tests revealed a smaller RT compatibility effect for symbol prompts ( $\Delta = 30$  ms,  $t(29) = 6.29$ ,  $p < .001$ ) than for syllable prompts ( $\Delta = 40$  ms,  $t(29) = 6.73$ ,  $p < .001$ ).



**Figure 3.5** Mean RTs  $\pm$  SEs for each prompt type in each compatibility condition.

**Table 3.6** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between prompt type and compatibility.

Prompt	Compatibility	Mean	RT (ms)	SE	95% CI
Symbol	Incompatible	6.56	703	0.03	[6.49, 6.62]
	Compatible	6.51	673	0.03	[6.44, 6.58]
Syllable	Incompatible	6.32	554	0.02	[6.28, 6.36]
	Compatible	6.24	514	0.02	[6.20, 6.28]

Of primary interest, the three-way interaction between prompt type, compatibility, and SOA was significant (see Figure 3.4 and Table 3.7), and follow-up *t* tests revealed significant RT compatibility effects at SOA1-3 for symbol prompts and at SOA1-4 for syllable prompts (all  $p < .006$ ). Further comparisons of RT compatibility effects across SOAs for each prompt type revealed a smaller RT compatibility effect at SOA1 than at SOA3 ( $p = .004$ ) for syllable prompts (i.e., SSC+SRC effects) and a larger RT compatibility effect at SOA1 than at SOA4 only before Bonferroni correction for symbol prompts (i.e., SRC effects).

**Table 3.7** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between prompt type, compatibility, and SOA.

Prompt	Compatibility	SOA (ms)	Mean	RT(ms)	SE	95% CI
Symbol	Incompatible	552	6.63	757	0.04	[6.55, 6.71]
		736	6.56	706	0.04	[6.49, 6.63]
		920	6.54	691	0.03	[6.47, 6.61]
		1104	6.50	662	0.03	[6.43, 6.56]
	Compatible	552	6.57	711	0.04	[6.49, 6.65]
		736	6.52	678	0.04	[6.45, 6.59]
		920	6.49	656	0.04	[6.41, 6.56]
		1104	6.48	649	0.03	[6.42, 6.53]
Syllable	Incompatible	552	6.37	581	0.02	[6.32, 6.41]
		736	6.33	563	0.02	[6.29, 6.38]
		920	6.30	542	0.02	[6.25, 6.34]
		1104	6.27	531	0.02	[6.23, 6.32]
	Compatible	552	6.32	554	0.03	[6.27, 6.37]
		736	6.26	521	0.02	[6.21, 6.30]
		920	6.21	496	0.02	[6.16, 6.26]
		1104	6.19	489	0.02	[6.15, 6.23]

### 3.2.3 Discussion

The aim of Experiment 1 was to examine time courses of SRC and SSC+SRC effects in a within-subjects design and to thus determine whether and to which extent automatic imitation effects are confounded with SSC effects when both S-S and S-R overlap is present (Objective 1). SRC effects were expected to result from S-R overlap in the symbol condition and SSC+SRC effects were expected from both S-S and S-R overlap in the syllable condition. The

overall higher ER and slower RT for symbol prompts than for syllable prompts suggested that responding to a symbol prompt by speaking out an arbitrarily assigned syllable was more difficult than responding to a syllable prompt by reading out that syllable. RTs were faster for longer SOAs, which has been consistently observed in previous studies (Adank et al., 2018; Kerzel & Bekkering, 2000). The presentation of a distractor video could be considered as a warning signal of the appearance of a prompt. Longer SOAs meant that participants had more time to get ready and were thus better prepared and responded faster.

ER and RT compatibility effects were observed for both symbol and syllable prompts, and RT compatibility effects were smaller for symbol prompts than for syllable prompts. In other words, opposite to what Kerzel & Bekkering (2000) observed, the current experiment found that the SRC effect was smaller than the SSC+SRC effect. This finding suggested that both S-S and S-R overlap might have contributed to the compatibility effect in the syllable condition, resulting in an SSC+SRC effect that was larger than the SRC effect in the symbol condition. Critically, however, the SSC+SRC effect was only slightly, though significantly, larger than the SRC effect in the current experiment, while the SSC effect in Kerzel (2002) was comparable in size with the SRC effect in Kerzel & Bekkering. As discussed earlier, since responding to syllable prompts was easier than responding to symbol prompts, participants might be more resistant to the influence of distractors in the syllable condition. It was possible that, compared to using symbol prompts, using syllable prompts on the one hand increases the compatibility effect due to the additional S-S overlap between distractors and prompts whose relative compatibility varies across trials; on the other hand decreases the compatibility effect due to overlap between prompts and responses that are always compatible in every trial and thus facilitate participants' overall performance. Notably, as discussed in Section 3.1.3, since S-S and S-R overlap interact with each other when both are present, it is in fact difficult to decisively claim whether SSC+SRC effects will be larger than SSC or SRC effects or the combination of the two. Instead, a different time course is expected for SSC+SRC effects than for either SSC or SRC effects.

Of primary interest, current results suggested that SRC and SSC+SRC effects follow different time courses. For symbol prompts, SRC effects decreased across SOAs, and the largest effect at SOA1 was larger than the smallest effect at SOA4, though only before Bonferroni correction. In contrast, for syllable prompts, SSC+SRC effects seemed to increase across SOA1-3 and then decrease at SOA4, and the largest effect at SOA3 was larger than the smallest effect at SOA1. Consistent with Prediction 1 given in Section 3.2.1, SRC effects in the symbol condition were found to develop faster across SOAs and reach maximum at a shorter SOA than SSC+SRC effects in the syllable condition, as indicated by the significant interaction between prompt type, compatibility, and SOA. This finding suggested that automatic imitation effects are confounded with SSC effects in the syllable condition.

However, this finding was also consistent with Prediction 3 given in Section 3.2.1, suggesting that automatic imitation effects can be reduced to pure SSC effects in the syllable condition. To further test these two predictions, it is essential to demonstrate distinctive time courses of SSC, SRC, and SSC+SRC effects. Experiment 2 reported in the following section examined SSC, SRC, and SSC+SRC effects and their time courses in a within-subjects design.

### 3.3 Experiment 2

		Prompt type			
		Tone		Syllable	
Response modality	Manual	No overlap		S-S overlap	
		 500 or 1,000 Hz tone	 Spoken /ba/ or /da/	 Press left- or down-arrow key	 Press left- or down-arrow key
	Vocal	S-R overlap		Both S-S and S-R overlap	
		 500 or 1,000 Hz tone	 Spoken /ba/ or /da/	 Say /ba/ or /da/	 Say /ba/ or /da/

**Figure 3.6** The study design of Experiment 2 with two response modalities and two prompt types that factorially comprised four tasks. In each task, participants had to either manually press a key or vocally produce a syllable in response to either a pure tone or a spoken syllable prompt; meanwhile, participants had to ignore the speaker who was mouthing either /ba/ or /da/ in distractor videos. Shaded cells indicate expected S-S and/or S-R overlap in each task.

The aim of Experiment 2 was to examine time courses of SSC, SRC, and SSC+SRC effects in a within-subjects design and to thus determine whether automatic imitation effects are confounded with, but cannot be reduced to, SSC effects when both S-S and S-R overlap is present (Objective 2). Participants in the current experiment were instructed to either manually press the left- or down-arrow key or vocally produce /ba/ or /da/ in response to either pure tone (a high- or low-frequency tone) or spoken syllable (auditory /ba/ or /da/) prompts and to ignore the speaker's /ba/ and /da/ articulations in distractor videos (see Figure 3.6). Two response modalities and two prompt types factorially comprised four tasks, across which the same pair of distractor videos were presented. S-S overlap was assumed between distractors and syllable prompts, but not between distractors and tone prompts; S-R overlap was assumed between distractors and vocal responses, but not between distractors and manual responses. Accordingly, SSC effects were expected to result from S-S overlap in the syllable-manual task, SRC effects were expected from S-R overlap in the tone-vocal task, and SSC+SRC effects were expected from both S-S and S-R overlap in the syllable-vocal task. Tone prompts and manual responses were assumed not to be associated with any syllable, so no compatibility

effect was expected from the tone-manual control task. Thus, four tasks measured no compatibility, SSC, SRC, and SSC+SRC effects. Six SOAs were included to exhibit time courses for SSC, SRC, and SSC+SRC effects. It was expected that, following the dual-route hypothesis of speech, SSC effects would develop across SOAs more slowly and reach maximum at a longer SOA than SSC+SRC effects and SRC effects develop more quickly and reach maximum at a shorter SOA than SSC+SRC effects. This finding would then suggest that automatic imitation effects are confounded with, but cannot be reduced to, SSC effects when both S-S and S-R overlap is present. Detailed predictions and corresponding statistical outcomes are given in the following section.

An eye-tracker was used to monitor participants' eye-movements to examine whether they had been looking at the speaker's face and if so, which facial area (mouth or eye). Previous research has found that people tend to look at the speaker's mouth and eyes to search for linguistic and social information respectively (Birmingham & Kingstone, 2009; Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998). Three areas of interest (AOI) were selected to sort eye-tracking data in the current experiment: mouth, eye, and other. The lower half of the speaker's face was selected to define the mouth area, the upper half define the eye area, and the remaining part of the screen define the other area. The other area was included to examine if participants had been looking at the speaker's face at all.

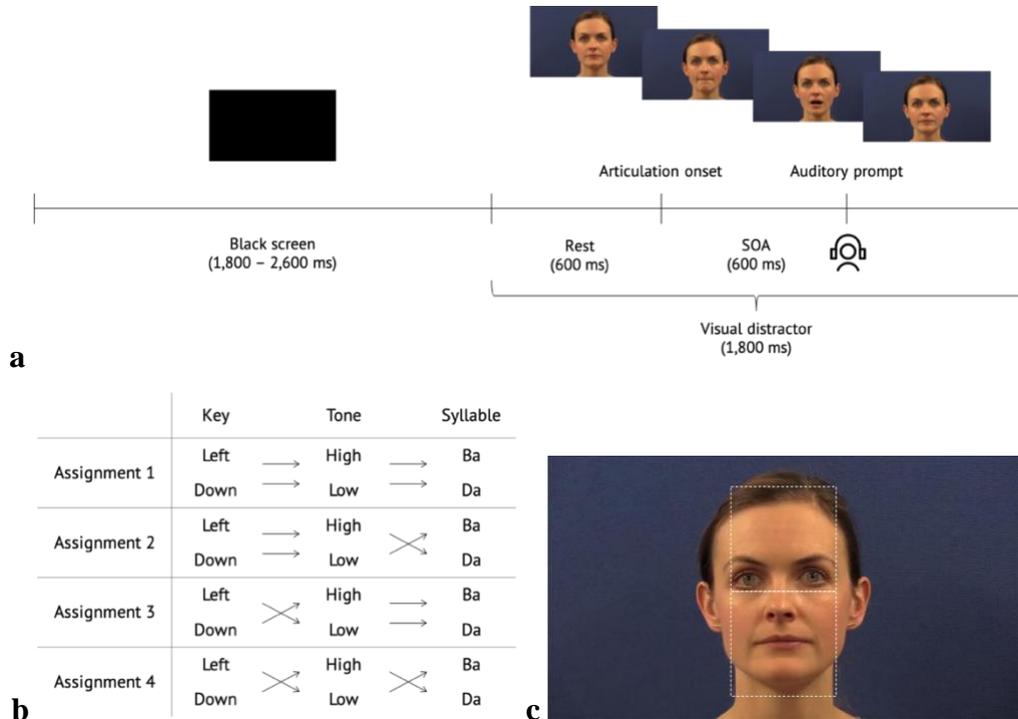
### **3.3.1 Method**

#### ***Participants***

Twenty-five participants were recruited, but one was excluded for having no eye-tracking data collected, which could be because the person was not looking at the screen at all. Hence, the final behavioural analysis included 24 participants (21 female;  $M_{\text{age}} = 21.46$  years,  $SD_{\text{age}} = 3.92$ ,  $\text{range}_{\text{age}} = 18\text{--}31$ ). All were native British English speakers with no report of speech or language disorders or other neurological disorders. Visual acuity was assessed, and all had a binocular vision rating of less than 0.1 on the logarithm of the minimum angle of resolution scale, equating to greater than 0.8 on the decimal scale (Colenbrander, 2002). Pure tone audiometry test was conducted using the MAICO MA 41 audiometer, and each ear was tested separately. All had normal hearing, i.e., average pure tone threshold of 20 dB HL or better at octave frequencies between 250 and 8,000 Hz in both ears (British Society of Audiology, 2011). Five participants were further excluded from eye-tracking analysis for their eye-tracking data failed to meet the selection requirements (see Method section for detailed explanations). Hence, the final eye-tracking analysis included 19 participants (17 female;  $M_{\text{age}} = 21.63$  years,  $SD_{\text{age}} = 4.19$ ,  $\text{range}_{\text{age}} = 18\text{--}31$ ). Participants received £15 or course credit. The

University Research Ethics Committee approved the procedures, and all participants gave written informed consent.

**Material and stimuli**



**Figure 3.7** **a** Schematic timeline of a trial in Experiment 2. **b** Four key-tone-syllable assignments counterbalanced across participants. **c** AOI used to sort eye-tracking data. The top and the bottom white squares define the eye and the mouth areas respectively, and the area outside these two squares were defined as the other area. The size of each square was 5.22° x 5.22° of visual angle.

Stimuli (see Figure 3.7a) included silent distractor videos of a speaker saying /ba/ and /da/, a 500- and a 1,000-Hz pure tone prompts, and spoken syllable prompts /ba/ and /da/. The same videos presented in Experiment 1 were employed in the current experiment but were further edited using MATLAB (R2015b). Each video lasted 1,800 ms, and the speaker was shown from her neckline upward in colour. Prior to articulation onset, the speaker remained in a resting configuration for 600 ms. In the *da* video, the airflow was released between the tongue and the upper teeth at 1,120 ms, the vertical extent of the mouth configuration reached its maximum at 1,280 ms, and the mouth began closing at 1,400 ms. In the *ba* video, the mouth’s vertical extent reached its minimum at 1,040 ms, the airflow was released between the lips at 1,120 ms, the mouth’s vertical extent then reached its maximum at 1,240 ms, and the mouth began closing at 1,400 ms. Viewing was unrestrained at a distance of 80 cm from a PC monitor, and the speaker’s face was shown in a visual angle of 10.78° x 8.36°. Additionally, the speaker’s mouth was shown in 0.74° x 2.86° at its resting configuration and 1.53° x 2.42° during the vowel articulation. The most salient visual difference between two videos was

before the consonant burst, with the mouth shown in  $0.34^\circ \times 2.86^\circ$  in the *ba* video and  $1.13^\circ \times 2.69^\circ$  in the *da* video.

Auditory prompts were either pure tones or spoken syllables. Spoken syllables /ba/ and /da/ were recorded with the same speaker in a sound-attenuated booth. Audio digitising was performed at 44.1 kHz with 16 bits. Syllable recordings were synthetically manipulated to be about the same length (400 ms) and root-mean-square normalised using Praat. Pure tones were created in MATLAB with frequencies of 500 and 1,000 Hz. Auditory prompts were delivered to participants through Beyerdynamic DT 297 PV MK II / 80 ohms headset at a comfortable level. The prompt was presented at one of six SOAs (0, 150, 300, 450, 600, or 750 ms relative to articulation onset) in each trial. SOAs were equally spaced across the video duration, with 0-ms SOA coinciding with the articulation onset in each video, 600-ms SOA coinciding with the consonant burst, and 750-ms SOA coinciding with the vowel articulation right before the speaker's mouth started to close. The experiment was performed using Presentation.

### ***Design and procedure***

Six variables were manipulated: prompt type, response modality, key-tone-syllable assignment, compatibility, SOA, and jitter. For prompt type, participants responded to either pure tone (a high- or low-frequency tone) or spoken syllable (/ba/ or /da/) prompts in different tasks. For response modality, participants responded by either manually pressing response keys (the left- or down-arrow key) or vocally producing syllables (/ba/ or /da/) in different tasks. Hence, these two variables comprised four tasks with different prompt-response pairs (see Figure 3.6): tone-manual, tone-vocal, syllable-manual, and syllable-vocal. Both *ba* and *da* distractor videos were presented in all tasks that were given in separate blocks, and the task order was counterbalanced across participants.

Key-tone-syllable assignments (see Figure 3.7b) were counterbalanced across participants to avoid potential confounds associated with any particular pairs, and data were collapsed over this variable in the analysis. In the tone-manual condition, half of participants pressed the left key in response to a high tone and pressed the down key to a low tone (Assignment 1 and 2), and the opposite was assigned to the other half (Assignment 3 and 4). In the tone-vocal condition, half of participants said /ba/ in response to a high tone and said /da/ to a low tone (Assignment 1 and 3), and the opposite was assigned to the other half (Assignment 2 and 4). In the syllable-manual condition, half of participants pressed the left key in response to an auditory /ba/ and pressed the down key to an auditory /da/ (Assignment 1 and 4), and the opposite was assigned to the other half (Assignment 2 and 3). In the syllable-

vocal condition, all participants said /ba/ in response to an auditory /ba/ and said /da/ to an auditory /da/.

Compatibility included two levels (compatible vs. incompatible) defined according to the assigned distractor-prompt and/or distractor-response pairs. In the tone-vocal condition, trials were defined as compatible when distractors and vocal responses corresponded (e.g., seeing /ba/ while saying /ba/) and incompatible when the two did not (e.g., seeing /ba/ while saying /da/). In the syllable-manual condition, trials were compatible when distractors and syllable prompts corresponded (e.g., seeing /ba/ while hearing /ba/) and incompatible when the two did not (e.g., seeing /ba/ while hearing /da/). In the syllable-vocal condition, trials were compatible when distractors corresponded with both syllable prompts and vocal responses (e.g., seeing /ba/ while saying /ba/ in response to an auditory /ba/) and incompatible when distractors corresponded with neither (e.g., seeing /ba/ while saying /da/ in response to an auditory /da/). Though no compatibility effect was expected from the tone-manual condition, two levels were defined. For instance, with Assignment 1 (see Figure 3.7b), trials were compatible when the *ba* video was presented and participants pressed the left key in response to a high tone and incompatible when the *ba* video was presented and participants pressed the down key in response to a low tone. This was because participants said /ba/ in response to a high tone in the tone-vocal condition and pressed the left key in response to an auditory /ba/ in the syllable-manual condition, and so both the high tone and the left key were associated with the syllable *ba*. Compatibility effects (i.e., incompatible – compatible) were calculated per participant per experimental condition and used as dependent variables for this experiment.

Six SOAs (0, 150, 300, 450, 600, or 750 ms) were used to determine the time intervals between the speaker's articulation onset in a distractor video and the time point when an auditory prompt was presented. The speaker's mouth began to move at 0-ms SOA, and the speaker was mouthing /a/ at 750-ms SOA. SOAs were randomized across trials. Additionally, inter-trial intervals were jittered randomly from 1,800 to 2,600 ms with 200-ms increments to prevent trial onset's predictability, and data were collapsed over this jitter variable in the analysis.

The experiment took place in a soundproofed, light-controlled booth. The prompt-response pair was given in written instructions presented on the screen before each task. In all tasks, while responding to auditory prompts, participants were instructed to look at the speaker's face but to ignore what she was saying in distractor videos. For manual responses, participants were instructed to use their right index finger only and to always place the finger between the left- and down-arrow keys before and after making a response. In each trial (see Figure 3.7a), the black screen first appeared for a jittered time interval and was then followed by a video distractor. The speaker's articulation started at 600 ms after the video onset, and

an auditory prompt was presented at one of six SOAs after the articulation onset. The screen turned black at the end of the video to signal the next trial onset. The whole session included four blocks that corresponded to four tasks with 240 trials each (960 trials in total). Within each block, 120 trial types (2 Prompts x 2 Distractors x 6 SOAs x 5 Jitters) were repeated twice and presented in a randomised order. Additionally, each block was divided into four runs with 30 trials each, and participants could take a break in between. Ten practice trials were given before each block, and the whole session lasted about 2 hours.

### *Data recording and analysis*

Vocal responses were recorded using the Beyerdynamic headset microphone that was connected to a Focusrite Scarlett 2i4 USB Audio Interface plugged into the sound card input of a Dell PC at 44.1kHz with 16 bits. Recordings started at the prompt onset and lasted 2,000 ms. RTs for vocal responses were time intervals between the prompt onset and the time point when the system detected .2 of the total input range recorded in Presentation. RTs for manual responses were time intervals between the prompt onset and the time point when participants pressed the first of two response keys. A warning saying *NO RESPONSE* was displayed for 1,000 ms for missed trials, *RESPONSE TOO EARLY* for RTs < 50 ms, and *RESPONSE TOO LATE* for RTs > 2,000 ms.

Vocal responses were checked manually using Praat. Errors included incorrect responses, missed trials, anticipatory responses with RTs < 100 ms, and outliers with RTs more than three standard deviations away from corresponding condition means per response. To be consistent with other studies reported in this thesis, a natural log-transformation was first applied to RTs. In addition, because manual and vocal responses were measured differently, the mean log-transformed RT per condition per participant was standardised to the overall mean of log-transformed RTs across each response modality, yielding a z-score for each participant each condition. RT compatibility effects were then calculated by subtracting standardised RTs in compatible trials from standardised RTs in incompatible trials. ER compatibility effects were also calculated by subtracting ERs in compatible trials from ERs in incompatible trials. ER and RT compatibility effects were then subjected to separate repeated-measures ANOVA with prompt type (tone versus syllable), response modality (manual versus vocal), and SOA (0, 150, 300, 450, 600, or 750 ms, also referred to as SOA1, 2, 3, 4, 5, and 6) as within-subjects variables. The significance level was set to  $p < .05$ . Greenhouse-Geisser correction for non-sphericity and Bonferroni-correction for multiple comparisons were applied whenever appropriate. Expected statistical results for each prediction for Objective 2 are listed below:

- **Prediction 5:** following the dual-route hypothesis, SSC effects were expected to develop across SOAs more slowly and reach maximum at a longer SOA than SSC+SRC effects and SRC effects develop more quickly and reach maximum at a shorter SOA than SSC+SRC effects, suggesting that automatic imitation effects are confounded with, but cannot be reduced to, SSC effects in the syllable-vocal condition – the interaction between prompt type x response modality x SOA should be significant, and follow-up t tests should reveal the largest SSC effect at a longer SOA and the largest SRC effect at a shorter SOA than the largest SSC+SRC effect.
- **Prediction 6:** following the motor hypothesis, SRC and SSC+SRC effects were expected to demonstrate the same time course and SSC effects were not expected to occur, suggesting that automatic imitation effects are not confounded with SSC effects in the syllable-vocal condition – the interaction between prompt type x response modality x SOA should not be significant; the main effect of prompt type and the interaction between prompt type x SOA should not be significant; yet, the main effect of response modality and the interaction between response modality x SOA should be significant.
- **Prediction 7:** following the perceptual hypothesis, SSC and SSC+SRC effects were expected to demonstrate the same time course and SRC effects were not expected to occur, suggesting that automatic imitation effects are reduced to pure SSC effects in the syllable-vocal condition – the interaction between prompt type x response modality x SOA should not be significant; the main effect of response modality and the interaction between response modality x SOA should not be significant; the main effect of prompt type and the interaction between prompt type x SOA should be significant.
- **Prediction 8:** following the conceptual hypothesis, SSC, SRC, and SSC+SRC effect were expected to demonstrate the same time course, suggesting that both automatic imitation and SSC effects result from the relative conceptual correspondence between distractors and prompts/responses – the interaction between prompt type x response modality x SOA should be significant, but follow-up t tests should reveal the largest SSC, SRC, and SSC+SRC effects at the same SOA.

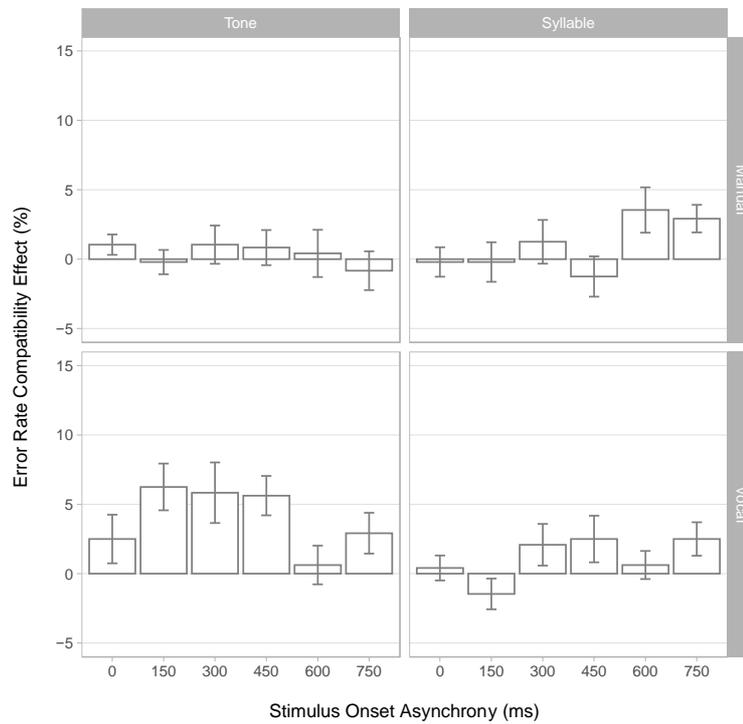
During the experiment, a Tobii X120 eye-tracker recorded participants' eye movements. Placed under the computer screen, the eye-tracker was 70 cm away from participants' eyes, and this distance was measured before each block. Eye movements were recorded by tracking the pupil and corneal reflection of the right eye at a sample rate of 120 Hz. Calibration was carried out using a customized 6-point configuration defining mouth and

eye areas (see Figure 3.7c) before each block. Eye-tracking data were first analysed to check for data loss due to blinks or head movements. Participants' head was not fixated by using a chin rest because they needed to make vocal responses in some task conditions. Therefore, quite a lot data loss was expected and relatively lenient criteria were used to process eye-tracking data. Participants were instructed to look at the speaker's face when she was present on the screen for 1,800 ms, so a maximum of 216 samples was expected from each trial. Since the speaker's articulation lasted 1,200 ms in each video, trials with more than 33.33% (i.e., 72 samples) data loss were excluded (percentage of trials included per participant per task:  $M = 64.8$ ,  $SD = 27.9$ ). Then, five participants with less than 30 trials remaining in any task condition were excluded from eye-tracking analysis. However, all participants were included in behavioural analysis for eye-tracking results were not of primary interest in the current study but were used to examine whether participants had followed the instructions. Thus, behavioural and eye-tracking data were analysed separately.

Eye-tracking sample sizes (i.e., numbers of samples) were subjected to repeated-measures ANOVA with prompt type (tone versus syllable), response modality (manual versus vocal), and AOI (mouth, eye, or other) as within-subjects variables. Prompt type and response modality were included to examine whether participants took different strategies depending on the assigned tasks. The significance level was set to  $p < .05$ . Greenhouse-Geisser correction for non-sphericity and Bonferroni-correction for multiple comparisons were applied whenever appropriate.

### 3.3.2 Results

#### Error rate analysis



**Figure 3.8** ER compatibility effects  $\pm$  SEs in each experimental condition. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where SSC+SRC effects were expected.

**Table 3.8** Three-way ANOVA summary for ER compatibility effects, as a function of prompt type, response modality, and SOA.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	$Epsilon$	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Prompt	1	23	1.00	177.78	1953.47	2.09	.161	.083
Response	1	23	1.00	487.67	1906.08	5.89	.024	.204
SOA	5	115	0.77	179.69	4647.40	0.89	.491	.037
Prompt x Response	1	23	1.00	434.03	1722.22	5.80	.024	.201
Prompt x SOA	3.36	77.31	0.67	608.68	4922.57	2.84	.038	.110
Response x SOA	5	115	0.83	439.41	5379.34	1.88	.103	.076
Prompt x Response x SOA	5	115	0.79	189.93	4341.32	1.01	.417	.042

*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

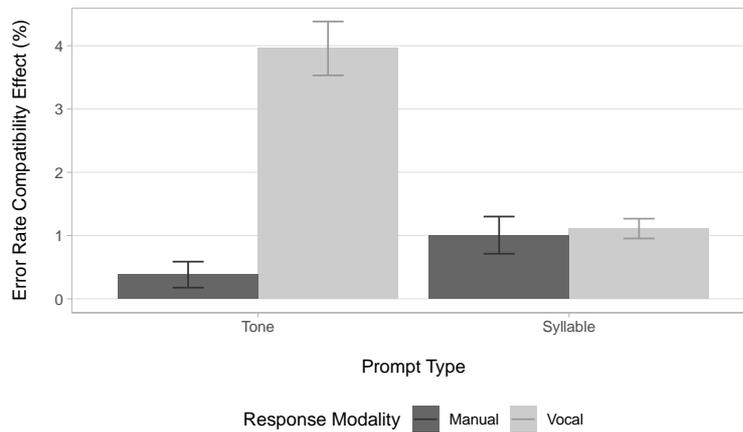
On average, participants made 5.1% errors (incorrect responses: 3.7%; missed trials: 0.5%; anticipatory responses: 0.9%; outliers: 0.0%). Analyses for ER compatibility effects (see

Figure 3.8, Table 3.8, and Table 3.9) revealed a main effect of response modality, with a smaller ER compatibility effect for manual responses than for vocal responses.

**Table 3.9** ER compatibility effects, SEs, and 95% CIs for each level within response modality.

	Level	Mean	SE	95% CI
Response	Manual	0.69	0.36	[-0.05, 1.44]
	Vocal	2.54	0.60	[1.30, 3.77]

The two-way interaction between prompt type and response modality (see Figure 3.9 and Table 3.10) was significant, and follow-up *t* tests revealed a smaller ER compatibility effect for manual responses than for vocal responses when responding to tone prompts ( $t(23) = 2.95, p = .007$ ) but no difference between manual and vocal responses when responding to syllable prompts ( $t(23) = 0.12, p = .90$ ).



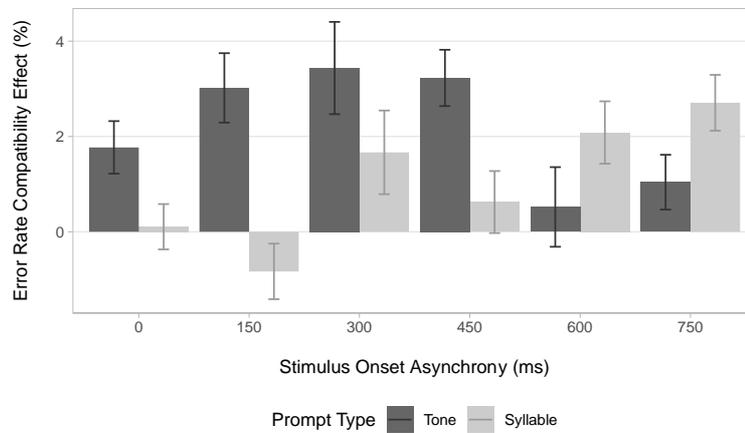
**Figure 3.9** ER compatibility effects  $\pm$  SEs for each prompt type for each response modality.

**Table 3.10** ER compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and response modality.

Prompt	Response	Mean	SE	95% CI
Tone	Manual	0.38	0.51	[-0.68, 1.44]
	Vocal	3.96	1.06	[1.76, 6.16]
Syllable	Manual	1.01	0.73	[-0.51, 2.53]
	Vocal	1.11	0.39	[0.30, 1.92]

The two-way interaction between prompt type and SOA (see Figure 3.10 and Table 3.11) was significant, and follow-up *t* tests revealed a larger ER compatibility effect for tone prompts than for syllable prompts at SOA2 ( $t(23) = 2.69, p = .013$ ). Further comparisons of ER compatibility effects across SOAs for each prompt type revealed no difference across

SOAs for tone prompts but a smaller ER compatibility effect at SOA2 than at SOA6 for syllable prompts ( $p < .047$ ).

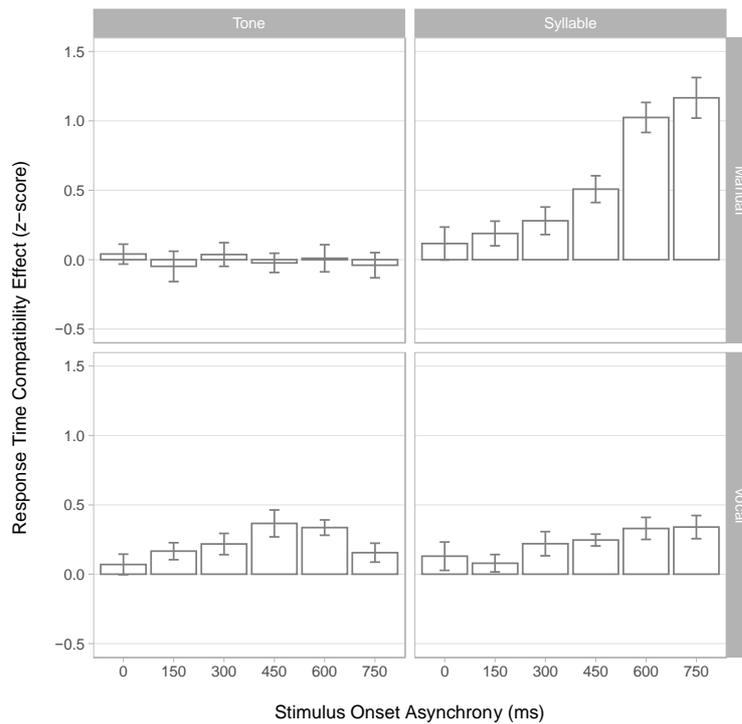


**Figure 3.10** ER compatibility effects  $\pm$  SEs at each SOA for each prompt type.

**Table 3.11** ER compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and SOA.

Prompt	SOA	Mean	SE	95% CI
Tone	1	1.77	0.79	[0.14, 3.40]
	2	3.02	1.04	[0.87, 5.18]
	3	3.44	1.39	[0.57, 6.30]
	4	3.23	0.84	[1.48, 4.98]
	5	0.52	1.19	[-1.95, 2.99]
	6	1.04	0.82	[-0.66, 2.74]
Syllable	1	0.10	0.68	[-1.31, 1.51]
	2	-0.83	0.83	[-2.56, 0.89]
	3	1.67	1.26	[-0.93, 4.27]
	4	0.63	0.93	[-1.30, 2.55]
	5	2.08	0.94	[0.15, 4.02]
	6	2.71	0.84	[0.98, 4.44]

## Response time analysis



**Figure 3.11** Standardised RT compatibility effects  $\pm$  SEs in each experimental condition. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where SSC+SRC effects were expected.

**Table 3.12** Three-way ANOVA summary for standardised RT compatibility effects, as a function of prompt type, response modality, and SOA.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	$Epsilon$	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Prompt	1	23	1.00	11.15	7.93	32.36	< .001	.585
Response	1	23	1.00	0.36	5.74	1.45	.241	.059
SOA	3.46	79.55	0.69	10.42	19.82	12.09	< .001	.345
Prompt x Response	1	23	1.00	10.73	4.85	50.86	< .001	.689
Prompt x SOA	5	115	0.73	8.16	22.56	8.32	< .001	.266
Response x SOA	5	115	0.75	3.10	17.23	4.14	.002	.153
Prompt x Response x SOA	3.68	84.63	0.74	5.32	17.89	6.84	< .001	.229

*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

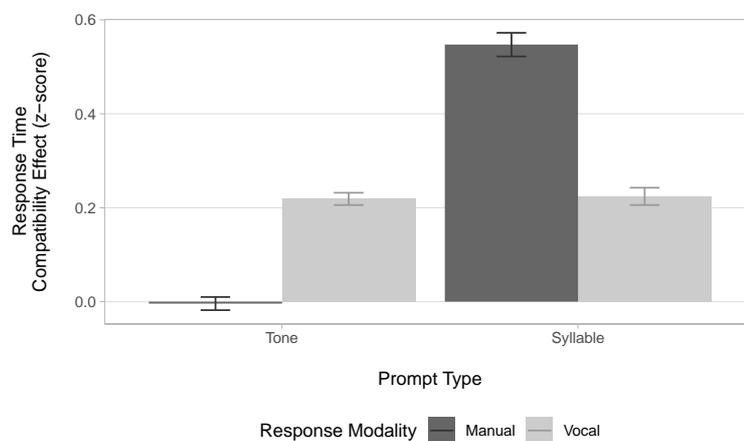
After errors were removed, analyses for standardised RT compatibility effects (see Figure 3.11, Table 3.12, and Table 3.13) revealed a main effect of prompt type, with a smaller RT

compatibility effect for tone prompts than for syllable prompts. Follow-up *t* tests for the main effect of SOA revealed smaller RT compatibility effects at SOA1-3 than at SOA5-6 (all  $p < .05$ ).

**Table 3.13** Standardised RT compatibility effects, SEs, and 95% CIs for each level within response modality and SOA.

	Level	Mean	SE	95% CI
Prompt	Tone	0.11	0.02	[0.06, 0.15]
	Syllable	0.39	0.05	[0.29, 0.48]
SOA	1	0.09	0.04	[0.01, 0.17]
	2	0.10	0.05	[-0.01, 0.20]
	3	0.19	0.04	[0.11, 0.27]
	4	0.27	0.04	[0.19, 0.36]
	5	0.43	0.05	[0.32, 0.53]
	6	0.41	0.05	[0.29, 0.52]

The two-way interaction between prompt type and response modality (see Figure 3.12 and Table 3.14) was significant, follow-up *t* tests revealed a smaller RT compatibility effect for tone prompts than for syllable prompts when performing manual responses ( $t(23) = 7.44$ ,  $p < .001$ ) but no difference between tone and syllable prompts when performing vocal responses ( $t(23) = 0.11$ ,  $p = .91$ ).

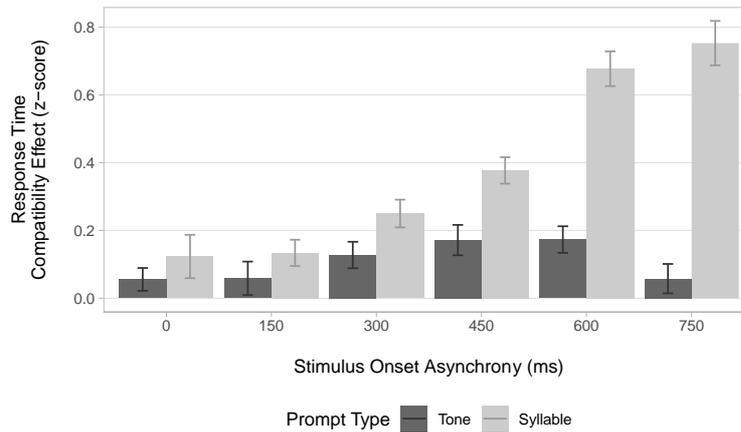


**Figure 3.12** Standardised RT compatibility effects  $\pm$  SEs for each prompt type for each response modality.

**Table 3.14** Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and response modality.

Prompt	Response	Mean	SE	95% CI
Tone	Manual	0.00	0.04	[-0.08, 0.07]
	Vocal	0.22	0.03	[0.15, 0.29]
Syllable	Manual	0.55	0.06	[0.42, 0.68]
	Vocal	0.22	0.05	[0.13, 0.32]

The two-way interaction between prompt type and SOA (see Figure 3.13 and Table 3.15) was significant, and follow-up *t* tests revealed smaller RT compatibility effects for tone prompts than for syllable prompts at SOA4-6 (all  $p < .03$ ). Further comparisons of RT compatibility effects across SOAs for each prompt type revealed no difference across SOAs for tone prompts but smaller RT compatibility effects at SOA1-4 than at SOA5-6 for syllable prompts (all  $p < .004$ ).

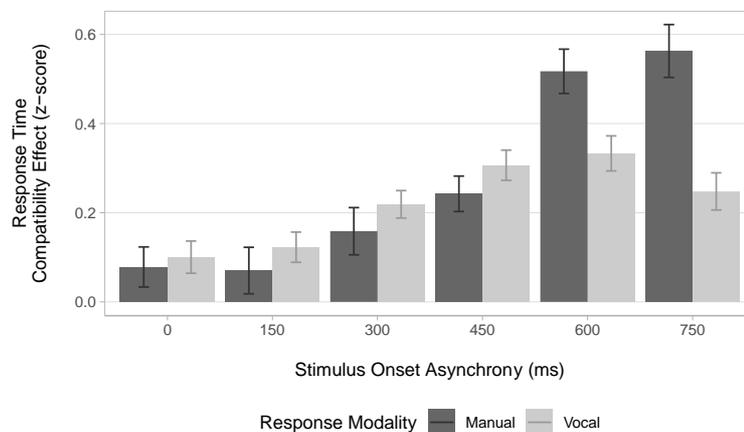


**Figure 3.13** Standardised RT compatibility effects ± SEs at each SOA for each prompt type.

**Table 3.15** Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and SOA.

Prompt	SOA	Mean	SE	95% CI
Tone	1	0.06	0.05	[-0.04, 0.16]
	2	0.06	0.07	[-0.09, 0.21]
	3	0.13	0.06	[0.01, 0.24]
	4	0.17	0.06	[0.04, 0.30]
	5	0.17	0.06	[0.06, 0.29]
	6	0.06	0.06	[-0.07, 0.19]
Syllable	1	0.12	0.09	[-0.07, 0.31]
	2	0.13	0.06	[0.02, 0.25]
	3	0.25	0.06	[0.13, 0.37]
	4	0.38	0.06	[0.26, 0.49]
	5	0.68	0.07	[0.53, 0.83]
	6	0.75	0.09	[0.56, 0.95]

The two-way interaction between response modality and SOA (see Figure 3.14 and Table 3.16) was significant, and follow-up *t* tests revealed larger RT compatibility effects for manual responses than for vocal responses at SOA5-6 (both  $p < .03$ ). Further comparisons of RT compatibility effects across SOAs for each response modality revealed smaller RT compatibility effects at SOA1-4 than at SOA5-6 for manual responses and smaller RT compatibility effects at SOA1 than at SOA4-5 for vocal responses (all  $p < .02$ ).



**Figure 3.14** Standardised RT compatibility effects ± SEs at each SOA for each response modality.

**Table 3.16** Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between response modality and SOA.

Response	SOA	Mean	SE	95% CI
Manual	1	0.08	0.06	[-0.06, 0.21]
	2	0.07	0.07	[-0.08, 0.22]
	3	0.16	0.08	[0.002, 0.32]
	4	0.24	0.06	[0.13, 0.36]
	5	0.52	0.07	[0.37, 0.66]
	6	0.56	0.09	[0.39, 0.74]
Vocal	1	0.10	0.05	[-0.01, 0.21]
	2	0.12	0.05	[0.02, 0.22]
	3	0.22	0.04	[0.13, 0.31]
	4	0.31	0.05	[0.21, 0.41]
	5	0.33	0.06	[0.22, 0.45]
	6	0.25	0.06	[0.13, 0.37]

Of primary interest, the three-way interaction between prompt type, response modality, and SOA (see Figure 3.11 and Table 3.17) was significant. Again, NO compatibility effects were expected from the tone-manual task, SSC effects were expected from the syllable-manual task, SRC effects were expected from the tone-vocal task, and SSC+SRC effects were expected from the syllable-vocal task. Comparisons of RT compatibility effects across four tasks at each SOA were further conducted and revealed that (all  $p < .05$ ):

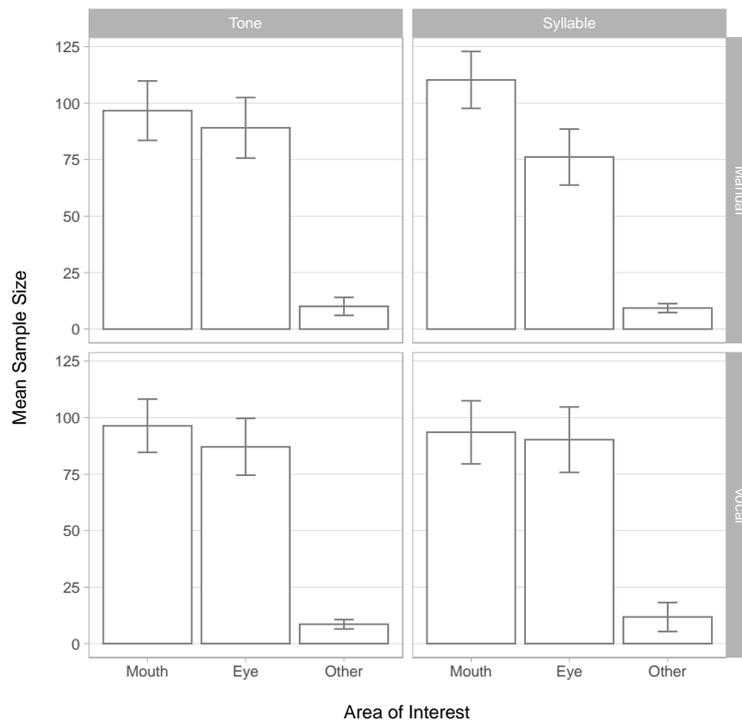
1. SSC effects were larger than NO effects at SOA4-6, and the largest difference was at SOA6.
2. SRC effects were larger than NO effects at SOA4-5, and the larger difference was at SOA4.
3. SSC+SRC effects were larger than NO effects at SOA4-5, and the larger difference was at SOA5.
4. SSC effects were larger than SRC effects at SOA5-6, and the larger difference was at SOA6.
5. SSC effects were larger than SSC+SRC effects at SOA5-6, and the larger difference was at SOA6.
6. SRC effects were not different from SSC+SRC effects at any SOA.

Consistent with Prediction 5, these results suggested that SSC effects reached maximum at a longer SOA and SRC effects reached maximum at a shorter SOA than SSC+SRC effects.

**Table 3.17** Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type, response modality, and SOA.

Prompt	Response	SOA	Mean	SE	95% CI
Tone	Manual	1	0.04	0.07	[-0.11, 0.19]
		2	-0.05	0.11	[-0.27, 0.18]
		3	0.04	0.09	[-0.14, 0.21]
		4	-0.02	0.07	[-0.17, 0.12]
		5	0.01	0.10	[-0.19, 0.21]
		6	-0.04	0.09	[-0.23, 0.15]
	Vocal	1	0.07	0.07	[-0.08, 0.22]
		2	0.17	0.06	[0.04, 0.29]
		3	0.22	0.08	[0.06, 0.38]
		4	0.37	0.10	[0.17, 0.57]
		5	0.34	0.06	[0.22, 0.45]
		6	0.16	0.07	[0.02, 0.30]
Syllable	Manual	1	0.12	0.12	[-0.13, 0.36]
		2	0.19	0.09	[0.01, 0.37]
		3	0.28	0.10	[0.08, 0.49]
		4	0.51	0.10	[0.31, 0.71]
		5	1.03	0.11	[0.80, 1.25]
		6	1.17	0.15	[0.86, 1.47]
	Vocal	1	0.13	0.10	[-0.08, 0.34]
		2	0.08	0.06	[-0.05, 0.21]
		3	0.22	0.09	[0.04, 0.40]
		4	0.25	0.04	[0.16, 0.33]
		5	0.33	0.08	[0.17, 0.49]
		6	0.34	0.08	[0.17, 0.51]

## Eye-tracking analysis



**Figure 3.15** Mean sample sizes  $\pm$  SEs for each AOI in each task. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where c SSC+SRC effects were expected.

**Table 3.18** Three-way ANOVA summary for eye-tracking same sizes, as a function of prompt type, response modality, and AOI.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	$Epsilon$	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Prompt	1	18	1.00	17.67	147.43	2.16	.159	.107
Response	1	18	1.00	22.67	306.22	1.33	.263	.069
AOI	1.14	20.48	0.57	351652.04	375820.64	16.84	< .001	.483
Prompt x Response	1	18	1.00	20.63	279.45	1.33	.264	.069
Prompt x AOI	1.09	19.59	0.54	1007.69	40979.85	0.44	.530	.024
Response x AOI	1.08	19.51	0.54	2054.28	18357.41	2.01	.171	.101
Prompt x Response x AOI	1.48	26.71	0.74	2577.99	51878.12	0.89	.392	.047

*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

Eye-tracking analyses (see Figure 3.15, Table 3.18, and Table 3.19) revealed a main effect of AOI, with more samples collected from the mouth and the eye areas than from the other area (both  $p < .001$ ). No other main effects or interactions were significant.

**Table 3.19** Mean sample sizes, SEs, and 95% CIs for each level within AOI.

	Level	Mean	SE	95% CI
AOI	Mouth	99.19	11.41	[75.21, 123.17]
	Eye	85.63	11.63	[61.19, 110.06]
	Other	9.93	3.25	[3.11, 16.76]

### 3.3.3 Discussion

The aim of Experiment 2 was to examine time courses of SSC, SRC, and SSC+SRC effects in a within-subjects design and to thus determine whether automatic imitation effects are confounded with, but cannot be reduced to, SSC effects when both S-S and S-R overlap is present (Objective 2). Participants responded to either pure tone or spoken syllable prompts by making either manual or vocal responses and ignored the speaker's syllable articulation in distractor videos. Two prompt types and two response modalities factorially comprised four tasks tested in separate blocks, but the same pair of distractor videos were used in all tasks. In this experiment, S-S overlap was assumed between distractors and syllable prompts but not between distractors and tone prompts; S-R overlap was assumed between distractors and vocal responses but not between distractors and manual responses. Therefore, an SSC effect was expected to occur in the syllable-manual task, an SRC effect was expected in the tone-vocal task, an SSC+SRC effect was expected in the syllable-vocal task, and no compatibility effect was expected in the tone-manual control task. Results are discussed in detail below.

#### *ER compatibility effects and the effect of S-R overlap*

The effect of S-R overlap assumed between distractors and vocal responses was indicated by the finding of a larger ER compatibility effect for vocal responses than for manual responses. However, the difference of ER compatibility effects between two response modalities only existed for tone prompts but not for syllable prompts (see Figure 3.9). In other words, the effect of S-R overlap was modulated by whether there was S-S overlap between distractors and prompts or not. On the one hand, the effect of S-S overlap presumably enhanced the overall ER compatibility effect for syllable prompts (i.e., SSC and SSC+SRC effects), relative to the overall effect for tone prompts. On the other hand, however, participants might be more resistant to the influence of distractors in the syllable-vocal task where prompts were compatible with responses in every trial, thus reducing the overall ER compatibility effect in this task (i.e., SSC+SRC effects). Consequently, the difference of ER compatibility effects between two response modalities disappeared for syllable prompts, ending up with comparable SSC and SSC+SRC effects. Previous studies have observed automatic imitation effects for auditory speech (Adank et al., 2018; Galantucci et al., 2009; Jarick & Jones, 2009), suggesting the existence of excitatory links between auditory and motor representations of the

same speech actions. Thus, repeating heard syllables in the syllable-vocal task should be easier than responding to arbitrarily assigned prompts in other tasks and accordingly, participants in the current experiment were likely to be more resistant to the influence of distractors in the syllable-vocal task.

Moreover, for the significant interaction between prompt type and SOA (see Figure 3.10), the ER compatibility effect was larger for tone prompts than for syllable prompts at SOA2. This finding was likely to be driven by the relatively large ER compatibility effect at SOA2 in the tone-vocal task where SRC effects were expected to occur at shorter SOAs. To look at this interaction in a different way, the ER compatibility effect was larger at SOA6 than at SOA2 for syllable prompts. This finding was likely to be driven by the relatively large ER compatibility effect at SOA6 in the syllable-manual task where SSC effects were expected to occur at longer SOAs. Therefore, though the interaction between prompt type, response modality, and SOA was not significant for ER compatibility effects, the significant interaction between prompt type and SOA seemed to indirectly suggest that the effects of S-S and S-R overlap in the current experiment followed different time courses. Specifically, consistent with the assumptions proposed by the Dimensional Overlap model, SSC effects tended to occur at longer SOAs than SRC effects.

### ***RT compatibility effects and the effect of S-S overlap***

The effect of S-S overlap assumed between distractors and syllable prompts was indicated by the finding of a larger RT compatibility effect for syllable prompts than for tone prompts. However, the difference of RT compatibility effects between two prompt types only existed for manual responses but not for vocal responses (see Figure 3.12). In other words, the effect of S-S overlap was modulated by whether there was S-R overlap between distractors and responses or not. Similar to what was proposed earlier, the effect of S-R overlap between distractors and vocal responses presumably enhanced the overall RT compatibility effect for vocal responses (i.e., SRC and SSC+SRC effects), relative to the overall effect for manual responses. In the meantime, however, participants were probably more resistant to the influence of distractors in the syllable-vocal task where prompts were compatible with responses in every trial, thus reducing the overall RT compatibility effect in this task (i.e., SSC+SRC effects). As a result, the difference of RT compatibility effects between two prompt types disappeared for vocal responses, resulting in comparable SRC and SSC+SRC effects.

Of primary interest, the interaction between prompt type, response modality, and SOA was significant for RT compatibility effects (see Figure 3.11). Consistent with Prediction 5 given in Section 3.3.1, follow-up analyses revealed that SSC effects reached maximum at a longer SOA and SRC effects reached maximum at a shorter SOA than SSC+SRC effects.

These findings suggested that automatic imitation effects are confounded with, but cannot be reduced to, SSC effects in the syllable-vocal condition where both S-S and S-R overlap was present.

Furthermore, the finding of larger RT compatibility effects at SOA5-6 than at SOA1-3 was likely to be driven by the increase of SSC effects at longer SOAs. For the significant interaction between prompt type and SOA (see Figure 3.13), RT compatibility effects were larger at SOA5-6 than at SOA1-4 for syllable prompts, which was also mainly driven by the increase of SSC effects at longer SOAs. To look at this interaction in a different way, the RT compatibility effects at SOA4-6 were larger for syllable prompts than for tone prompts. It was possible that, in addition to being driven by the increase of SSC effects at longer SOAs in the syllable-manual task, this finding could partially be attributed to the decrease of SRC effects at longer SOAs in the tone-vocal task.

Similarly, for the significant interaction between response modality and SOA (see Figure 3.14), RT compatibility effects were larger at SOA 5-6 than at SOA1-4 for manual responses, which was also mainly driven by the increase of SSC effects at longer SOAs. Interestingly, RT compatibility effects were larger at SOA4-5 than at SOA1 for vocal responses probably due to the increase of SRC and SSC+SRC effects across SOA1-5. More importantly, however, the RT compatibility effect was not statistically larger at SOA6 than at SOA1, suggesting a decrease of the compatibility effect at SOA6 relative to the effects at SOA4-5. This finding was potentially driven by the decrease of SRC effects at SOA6 in the tone-vocal condition, indicating a different time course for SRC effects than for SSC effects. To look at this interaction in a different way, the RT compatibility effects at SOA5-6 were larger for manual responses than for vocal responses. Similar to what was proposed earlier, it was possible that, in addition to being driven by the increase of SSC effects at longer SOAs in the syllable-manual task, this finding could partially be attributed to the decrease of SRC effects at longer SOAs in the tone-vocal task. Consistent with the Dimensional Overlap model, these findings also seemed to suggest that the effects of S-S and S-R overlap in the current experiment followed different time courses for RT compatibility effects.

Lastly, there seemed to be an asymmetry in which SRC effects appeared to be more prominent for ER data, whereas SSC effects were more prominent for RT data. It could be due to the nature of movements required in each response modality. For manual responses in the current experiment, participants had to move their right index finger either left or right and then press the arrow key. In other word, they could always have the right index finger ready to move, and it might be easier to suppress the unwanted right or left finger movement. In contrast, for vocal responses, participants had to prepare various articulators for a syllable articulation and the main articulators involved in producing syllables /ba/ and /da/ are different (lips versus tongue). Thus, it could be harder to suppress the unwanted syllable articulation

whose motor representation is more complex than that of a finger movement. Consequently, participants were more likely to perform a wrong vocal response rather than delaying it (i.e., more prominent SRC effects in ER data), and it was the other around for manual responses (i.e., more prominent SSC effects in RT data). Future studies could employ relatively more complex manual gestures that rely on coordination by various effectors (e.g., multiple fingers or even wrist and arm movements) and then investigate whether such an asymmetry would still exist for SSC and SRC effects.

### ***Gaze behaviours***

The eye-tracking analysis revealed that more samples were collected from the mouth and the eye areas than from the other area, suggesting that participants were following the instructions to look at the speaker's face. Notably, though not statistically significant, there seemed to be a tendency for people to look at the mouth area more in the syllable-manual task where SSC effects were expected. As mentioned earlier, people tend to look at the speaker's mouth and eyes to search for linguistic and social information respectively (Birmingham & Kingstone, 2009; Vatikiotis-Bateson et al., 1998). Thus, participants in the current experiment might unintentionally take different looking strategies after they were given task instructions, which might consequently affect compatibility effects in each task. To exclude this possibility, Experiment 3 reported Appendix A aimed to replicate findings from Experiment 2 while controlling participants' gaze behaviours.

### ***Experiment 3 (see Appendix A)***

The aim of Experiment 3 was to replicate behavioural findings from Experiment 2, but with participants instructed to look at the speaker's mouth, instead of face, while completing each task. Firstly, the eye-tracking analysis in Experiment 3 revealed that more eye-tracking samples were collected from the mouth area than from the eye or the other areas, suggesting that participants were following the instructions to look at the speaker's mouth while responding to auditory prompts. Interestingly, the eye-tracking analysis also revealed that more eye-tracking samples were collected from the mouth area when manual responses were required than when vocal responses were required. This finding was consistent with the observation in Experiment 2 that people tended to look at the mouth area more in the syllable-manual task. It could be that, when required to say something, participants unavoidably tended to look away from the speaker's mouth because the speaker's articulation might interfere with their own articulatory performance.

Behavioural results from Experiment 2 and 3 were similar, except that a few interactions involving SOA did not reach significance in Experiment 3. Different from

Experiment 2, the interaction between prompt type, response modality, and SOA in Experiment 3 was not significant for RT compatibility effects. Additionally, different from Experiment 2, Experiment 3 did not find a significant interaction between prompt type and SOA for ER compatibility effects or between response modality and SOA for RT compatibility effects. It seemed that the time courses for compatibility effects in Experiment 3 were less clear than those in Experiment 2. In Appendix B, further analyses for both behavioural and eye-tracking results were conducted by combining data from Experiment 2 and 3. These analyses were conducted to examine whether participants' gaze behaviours were statistically different between two experiments and if so, whether this difference in gaze behaviours led to different compatibility effects between two experiments.

#### ***Further analysis of Experiment 2 and 3 (see Appendix B)***

Further analysis of behavioural data from both Experiment 2 and 3 revealed similar results to those observed in Experiment 2 and 3 when analysed separately. Most importantly, neither the main effect of experiment nor any interaction involving experiment was significant. The interaction between prompt type, response modality, and SOA was marginally significant for ER compatibility effects and significant for RT compatibility effects. Consistent with Prediction 5 given in Section 3.3.1, follow-up analyses revealed that SSC effects reached maximum at a longer SOA and SRC effects reached maximum at a shorter SOA than SSC+SRC effects, thus suggesting that automatic imitation effects are confounded with, but cannot be reduced to, SSC effects in the syllable-vocal condition. Eye-tracking analysis suggested that participants in Experiment 2 and 3 were following the instructions to look at the speaker's face or to focus on her mouth respectively, but this difference in gaze behaviours did not affect compatibility effects elicited by the perception of speech distractors.

### **3.4 General discussion**

Two experiments reported in this chapter examined SSC, SRC, and SSC+SRC effects and their time courses in a within-subjects design. Specifically, two objectives were:

- **Objective 1:** to determine whether and to which extent automatic imitation effects are confounded with SSC effects when both S-S and S-R overlap is present (Experiment 1).
- **Objective 2:** to determine whether automatic imitation effects are confounded with – but cannot be reduced to – SSC effects when both S-S and S-R overlap is present (Experiment 2).

Experiment 1 examined time courses of SRC and SSC+SRC effects by comparing the effects of using symbol and syllable prompts in different speech SRC tasks. Participants were instructed to say /ba/ or /da/ in response to printed symbol (“%%” or “&&”) or syllable (“ba” or “da”) prompts in separate conditions while ignoring the speaker’s /ba/ and /da/ articulations in distractor videos. S-R overlap was assumed between distractors and responses, and S-S overlap was assumed between distractors and syllable prompts. Accordingly, SRC effects were expected to occur in the symbol condition and SSC+SRC effects were expected in the syllable condition. Compatibility effects were observed in both prompt conditions and were found to be larger in the syllable condition than in the symbol condition. Most importantly, consistent with Prediction 1 given in Section 3.2.1, SSC+SRC effects in the syllable condition developed more slowly across SOAs and reached maximum at a longer SOA than SRC effects in the symbol condition. This finding suggested that automatic imitation effects are confounded with SSC effects when both S-S and S-R overlap is present. However, this finding was also consistent with Prediction 3 given in Section 3.2.1, suggesting that automatic imitation effects can be reduced to pure SSC effects in the syllable condition.

Experiment 2 examined SSC, SRC, and SSC+SRC effects and their time courses. Participants were instructed to respond to either pure tone (a low- or high-frequency tone) or spoken syllable (auditory /ba/ or /da/) prompts by either manually pressing a key (press the left- or down-arrow key) and vocally producing a syllable (say /ba/ or /da/) in different tasks. In the meantime, participants had to ignore the speaker’s /ba/ and /da/ articulations in distractor videos presented in all tasks. Thus, two prompt types and two response modalities factorially comprised four tasks. S-S overlap was assumed between distractors and syllable prompts, and S-R overlap was assumed between distractors and vocal responses. Thus, SSC effects were expected to occur in the syllable-manual task, SRC effects were expected in the tone-vocal task, SSC+SRC effects were expected in the syllable-vocal task, and no compatibility effects were expected in the tone-manual control task. The effect of S-R overlap was observed for ER compatibility effects, which was further modulated by whether there was S-S overlap between distractors and prompts. The effect of S-S overlap was observed for RT compatibility effects, which was further modulated by whether there was S-R overlap between distractors and responses. Most importantly, consistent with Prediction 5 given in Section 3.3.1, SSC effects reached maximum at a longer SOA and SRC effects reached maximum at a shorter SOA than SSC+SRC effects, thus suggesting that automatic imitation effects are confounded with, but cannot be reduced to, SSC effects in the syllable-vocal condition where both S-S and S-R overlap is present.

Altogether, findings reported in this chapter suggested that the perception of speech distractors in a speech SRC task triggers both perceptual and motor processes and may interfere with participants’ performance when there is S-S overlap between distractors and

prompts and/or when there is S-R overlap between distractors and responses. In addition, these findings further suggested that it is possible to use SSC and SRC effects within the Dimensional Overlap framework to investigate multimodal processing of speech. Detailed discussions of current findings are as follows.

#### ***3.4.1 Time courses of compatibility effects***

In line with the Dimensional Overlap model, current experiments demonstrated different time courses for SSC and SRC effects, indicating that the two effects respectively measure the perceptual and the motor processes triggered by the perception of the same visual speech distractors. SRC effects observed in the syllable condition in Experiment 1 and in the tone-vocal task in Experiment 2 suggested that visual speech distractors interacted with the production of vocal responses, giving rise to automatic imitation effects. Since previous automatic imitation studies have suggested both audiomotor and visuomotor links underlying speech (e.g., Adank et al., 2018; Jarick & Jones, 2009), one may argue that SSC effects in the syllable-manual task in Experiment 2 actually resulted from the interaction between motor units activated by auditory syllable prompts and visual speech distractors. This argument is based on the Motor Theory of Speech Perception in its strong form that any kind of speech perception is perceiving speech gestures. Consequently, the difference between SSC effects in the syllable-manual task and SSC+SRC effects in the syllable-vocal task would be whether the selected speech motor units were then transferred into the assigned manual responses or were directly executed. The process of transferring into manual responses should not be affected by compatibility levels. As a result, SSC and SSC+SRC effects should exhibit the same time course, which was inconsistent with what was observed in Experiment 2. Therefore, a pure motor account of speech perception could not explain current findings. Instead, I proposed that SSC effects in the syllable-manual task were at least partially a product of audiovisual interaction between corresponding auditory and visual speech representations, measuring the perceptual process where visual speech interacts with the processing of auditory speech. I could not, however, completely deny that there might be interaction between audiomotor and visuomotor processes in SSC effects but would expect this influence to be small relative to audiovisual interaction.

Furthermore, I proposed that audiovisual interaction between visual speech distractors and auditory syllable prompts also contributed to SSC+SRC effects in the syllable-vocal task, in addition to visuomotor interaction between visual speech distractors and vocal responses. In other words, both S-S and S-R overlap contributed to SSC+SRC effects, resulting in a time course different from the time course for either SSC or SRC effects. Following the Dimensional Overlap model, it was possible that S-S and S-R overlap interacted with each

other in SSC+SRC effects, with the former contributing to compatibility effects at longer SOAs and the latter contributing compatibility effects at shorter SOAs. Future studies using syllable prompts, either printed or spoken, to investigate automatic imitation effects should explain their findings with the consideration of whether their experimental manipulations also affect the effect of S-S overlap between distractors and syllable prompts. The current study only attempted to use concepts in the computational version of the Dimensional Overlap model (Kornblum et al., 1999) to illustrate the time courses of various compatibility effects and so to indicate multimodal processing underlying speech. If future studies attempt to replicate the current study or to better elucidate these effects, a computational approach would help explain the underlying processes with greater detail.

### **3.4.2 *Automatic imitation in cross-modal settings***

Placing visual prompts in the centre of the speaker's mouth, previous studies have found comparable automatic imitation effects for auditory, visual, and audiovisual speech (Adank et al., 2018; Jarick & Jones, 2009). In line with prior research, SRC effects observed in the tone-vocal task in Experiment 2 added to the converging evidence that automatic imitation is automatic in the way that it occurs regardless of whether attention is directed to the same modality as the distractors. Additionally, similar results between Experiment 2 and 3 suggested that the main visible articulators involved in producing perceived speech actions need not be in the centre of the visual field to elicit automatic imitation. These findings were consistent with those from Leighton & Heyes (2010) who reported automatic imitation effects for both mouth and hand opening/closing movements when distractors were presented in the periphery of the visual field. Part of the reason for this finding was that because the current experiments only employed two syllables and thus participants did not need to analyse visual stimuli in detail in order to tell whether the presented distractor was /ba/ or /da/. Automatic imitation studies employing more syllables may observe a larger automatic imitation effect when participants have to focus on the speaker's mouth than when they just look at the face in general.

In the Dimensional Overlap model, automaticity and attention are suggested to be closely related, and an automatic process could under some circumstances be enhanced or attenuated respectively by focusing attention on or diverting attention from the object that triggers the process, but under no circumstances could the process be ignored. A similar view has been adopted by Heyes (2011) when examining the putative automaticity of automatic imitation. The current study demonstrated that automatic imitation of visual speech occurred in a crossmodal setting with attention directed to the auditory modality, but whether and how such manipulation of attention would modulate automatic imitation effects need further

investigation. Future research could investigate how automatic imitation of speech changes from a unimodal condition (with prompts and distractors presented in the same modality) to a crossmodal setting (with prompts and distractors presented in different modalities) in a within-subjects design and examine whether changes in automatic imitation are the same for auditory, visual, and audiovisual speech.

### ***3.4.3 Audiovisual speech in stimulus-stimulus compatibility effects***

As discussed in Section 2.2.2, seeing speakers' facial movements has been shown to improve comprehension of auditory speech masked with noise (Sumbly & Pollack, 1954), and mismatched visual cues may give rise to the McGurk illusion. Consistent with prior research, SSC effects observed in the syllable-manual task in Experiment 2 also demonstrated effective visual modulations on the processing of auditory speech. Questions then arise as to whether these three measures indicate the same multisensory processing underlying speech. Tiippana (2014) has defined the McGurk effect as a categorical change in auditory perception due to mismatched visual information. Hence, it is the occurrence of perceptual illusion in the McGurk effect that indicates the degree of audiovisual integration during speech processing. After reviewing both behavioural and neural research, Alsius, Paré, & Munhall (2018) have concluded that the McGurk illusion is likely to be mediated by a different process than the one underlying naturally congruent audiovisual speech and is instead linked to individuals' multisensory integration ability in other non-speech domains (C. Tremblay, Champoux, Bacon, & Theoret, 2007).

In contrast, according to the Dimensional Overlap model, SSC effects observed in Experiment 2 occurred because seeing the speaker's articulation in distractor videos automatically activated the corresponding auditory units that could be either the same or different from the ones participants heard and had to respond to. Thus, a larger SSC effect indicates greater activation in the corresponding auditory units by visual speech, which presumably further indicates stronger excitatory links between corresponding auditory and visual speech representations. Whether the McGurk effect and SSC effects observed in current experiments measure the same process underlying audiovisual speech requires further investigation, and future research could examine whether audiovisual SSC effects correlate with the McGurk effect, lipreading ability, and sentence comprehension of audiovisual speech with auditory stimuli masked in noise. Investigation of whether SSC effects could be generalized to other speech sounds is also important.

Furthermore, similar results between Experiment 2 and 3 also suggested that the main visible articulators involved in producing perceived speech actions need not be in the centre of the visual field to elicit SSC effects in audiovisual speech. Native Japanese speakers have

been found to be less susceptible to the McGurk effect than native English speakers (Sekiyama & Tohkura, 1991, 1993). Measuring participants' gaze behaviours during audiovisual speech, Hisanaga, Sekiyama, Igasaki, & Murayama (2016) found that English participants tended to look at the speaker's mouth area whereas their Japanese counterparts showed no bias towards any facial area during audiovisual speech. Interestingly, in the same study, having Japanese participants focus on the speaker's mouth did not make them more susceptible to the McGurk effect, suggesting that Japanese people generally make less use of visual information during audiovisual speech than English people. Hisanaga et al. have proposed that it is probably because visual speech in Japanese is less informative than visual speech in English and thus Japanese people rely less on visual speech than English people. In other words, it also suggests that people do look at the speaker's mouth as much as they could get enough information from visual stimuli. Therefore, making people focus the speaker's mouth more than they naturally do would not help them get more linguistic information or elicit larger SSC effects in the current experiment or a greater McGurk effect. Future studies could further investigate whether cultural differences observed for the McGurk effect also exist for SSC effects in audiovisual speech.

Moreover, previous studies using stimuli that involve semantic processing have observed an asymmetric pattern in which visual distractors tend to yield larger SSC effects with participants attending to the auditory modality than vice versa (Donohue, Appelbaum, Park, Roberts, & Woldorff, 2013; Yuval-Greenberg & Deouell, 2009). Donohue et al. have suggested that the quality of information might play a role in inducing SSC effects, but specifically in their study, visual stimuli seemed to be weighed more since printed words were presented parafoveally whereas spoken words were presented bilaterally and at a relatively high volume. Future studies could examine whether such an asymmetric pattern applies to pre-lexical audiovisual speech and whether, in the case of speech, it rather depends on characteristics that differentiate speech sounds. For example, speech sounds with salient visual features may elicit larger SSC effects when participants attend to auditory stimuli and have to ignore visual stimuli than vice versa; conversely, for speech sounds with salient auditory features, it probably works the other way around.

#### **3.4.4 *Visual modulations on auditory speech***

In speech science, the question of how visual inputs modulate auditory speech processing remains debatable. As discussed in Section 2.2.2, the perceptual account of speech perception proposes that visual inputs are directly integrated with auditory signals to achieve the least ambiguous result that matches an abstract prototype stored in memory and thus, the combination of imperfect auditory and visual speech stimuli will yield better performance

than either unimodal condition. Alternatively, the motor account proposes that visual inputs modulate auditory speech processing by involving the motor system through a forward-model mechanism. Specifically, visual inputs activate corresponding motor commands whose efference copies are then sent to the auditory regions to predict sensory consequences of the motor commands when executed overtly. Perception of auditory speech is enhanced when the predicted and perceived sensory patterns match, and misperception may arise, as in the case of the McGurk effect, when there is mismatch.

Critically, the perceptual and motor accounts are not mutually exclusive. The Dual-Route Model of visual speech has been proposed that both multisensory and sensorimotor integration processes exist through which visual inputs influence auditory speech processing (Okada & Hickok, 2009). Past neuroimaging and brain stimulation studies have found that the left superior temporal sulcus (i.e., a language-related auditory region) might be the place where multisensory integration occurs during audiovisual speech (Beauchamp, Nath, & Pasalar, 2010; Calvert et al., 1999; Matchin, Groulx, & Hickok, 2014). Additionally, the left inferior frontal gyrus (i.e., part of the speech production network) has been suggested to be a critical region underlying sensorimotor integration during speech processing (Murakami et al., 2018; Okada & Hickok, 2009; Skipper, van Wassenhove, Nusbaum, & Small, 2007). SSC+SRC effects in the syllable-vocal condition in Experiment 2 behaviourally suggested that seeing a speaker's articulation automatically activates corresponding auditory and motor units and therefore supported the assumption that visual information may influence auditory speech through both multisensory and sensorimotor integration processes. To further investigate the roles of superior temporal sulcus and inferior frontal gyrus in audiovisual speech, future research could use brain stimulation techniques such as TMS to modulate the neural activity in each area and see how each stimulation would affect SSC and/or SRC effects during speech processing.

### **3.5 Conclusion**

In conclusion, results from the current set of experiments demonstrate that SSC, SRC, and SSC+SRC effects measured using speech SRC tasks can be explained within the Dimensional Overlap framework and thus suggest that the observed SSC and SRC effects respectively measure the perceptual and the motor processes triggered by the same speech distractors. Therefore, current findings support the hypothesis that automatic imitation of speech occurs due to perception-induced motor activation. More importantly, current findings support the dual-route hypothesis that the perception of speech distractors triggers both perceptual and motor processes and interferes with participants' performance when there is S-S or S-R overlap respectively. Crucially, even when linguistic prompts such as printed or spoken

syllables are used, as in the case of SSC+SRC effects, current findings also suggest the involvement of perception-induced motor activation which, however, is likely to interact with a perceptual process triggered by the same speech distractors. Hence, future studies using speech-related prompts need to consider the possibilities of whether and how their experimental manipulations may modulate the effects of S-S and/or S-R overlap. As will be shown in the following chapters, current findings have provided important information for the design of the training studies reported in Chapter 4 and Chapter 5.

More importantly, there are implications for research in the speech domain. Current findings also suggest that speech SRC tasks can be used to behaviourally investigate multimodal (i.e., multisensory and sensorimotor) speech processing within a unified theoretical framework. Firstly, in addition to sensorimotor processing in automatic imitation effects, speech SRC tasks can be extended to the investigation of multisensory speech perception as in the case of SSC effects and thus provide a different approach to study audiovisual speech at the pre-lexical level apart from the McGurk effect. Additionally, as discussed in Section 2.2.2, Dual-Route Models of speech processing propose a ventral stream that involves speech recognition and comprehension and a dorsal stream that involves perceptuomotor translation during speech processing. Speech SRC tasks measuring SSC, SRC, and SSC+SRC effects may provide an avenue to behaviourally investigate the interaction between the two processing streams and therefore complement findings from neural research of speech processing.

Furthermore, as mentioned in Chapter 1, SRC paradigms have been widely used in both experimental and cognitive psychology for decades. Kornblum et al. have described research on SSC and SRC effects as “fragmented” (1999, p. 688) because previous research has viewed different compatibility effects as unrelated empirical phenomena and examined these effects in isolation. The Dimensional Overlap model has been proposed in attempt to account for various compatibility effects in a unified theoretical framework and thus to understand the basic common cognitive processing structure shared by all these effects. The current set of experiments demonstrates how various compatibility effects can be used to investigate multimodal speech processing. In the meantime, it also means that the current work demonstrates how different compatibility effects can be examined together using one single object, namely, speech, which is multimodal by nature. Compatible with the Dimensional Overlap framework, current findings indicate the dependence of speech on a common, domain-general processing structure that is also shared by other cognitive functions. Future studies could implement the computational Dimensional Overlap model to further explore the commonalities and differences between speech processing and other cognitive functions relying on the same processing structure.

Overall, the current research adds to the work of previous automatic imitation studies in providing direct comparisons of various compatibility effects measured using speech SRC tasks in a within-subjects design. In doing so, the current work demonstrates the possibility of employing various compatibility effects, which have been widely used but often examined separately in cognitive psychology, to investigate multimodal speech processing and therefore also suggests a way in which findings from the speech domain can be linked to findings of more general cognitive processing.

## **Chapter 4 Sensorimotor training effects on automatic imitation of native speech**

*The study reported in this chapter has been published in the journal *Psychonomic Bulletin & Review* with the title “Sensorimotor training modulates automatic imitation of visual speech” (see Wu, Evans, & Adank, 2019).*

### **4.1 Introduction**

As discussed in Section 2.1.3, humans possess an imitative capacity to quickly map observed actions onto their motor repertoire, and different accounts have been proposed to explain the origin of this imitative capacity. The genetic account of imitation proposes that the matching properties of the observation-execution link underlying the imitative capacity are innate and genetically predisposed (Ferrari et al., 2003; Meltzoff, 2002; Meltzoff & Decety, 2003; Rizzolatti et al., 1996); whereas the associative account proposes that the matching properties are essentially a product of sensorimotor experience acquired through direct self-observation and through social interactions (Catmur et al., 2009; Heyes, 2001, 2005, 2010; Ray & Heyes, 2011). A closer examination of arguments from each account suggests that one dispute between the genetic and the associative accounts concerns the role of experience in the development of the imitative capacity and the type of experience critically involved (Gallese et al., 2011). Specifically, the genetic account suggests that sensory and/or motor experience plays a facilitative or tuning role in the development of the imitative capacity. It means that the imitative capacity will eventually emerge even in the absence of sensory or motor experience, but such experience may help hasten the development, improve behavioural performance, or enhance perceptual acuity (Gottlieb, 1976). Alternatively, the associative account proposes that sensorimotor experience plays an inductive or forging role in the development of the imitative capacity, meaning that sensorimotor experience determines whether the imitative capacity will emerge (Gottlieb, 1976). Furthermore, according to the associative account, sensorimotor experience that consists of concurrent observation and execution of the same actions forges the matching properties of the observation-execution link. Conversely, sensorimotor experience that consists of concurrent observation and execution of different actions leads to an excitatory link between noncorresponding sensory and motor representations of different actions.

As discussed in Section 2.1.4, sensorimotor training studies have investigated the flexibility of the observation-execution link underlying manual gestures and suggested that the link is subject to change through sensorimotor training. However, one problem that

remains unsolved concerns the flexibility of the perception-production link underlying orofacial movements that, unlike visually transparent manual gestures, are visually opaque. In the following sections, I will first discuss expertise effects that provide naturalistic examples showing how responses in the human mirror system change as a result of experience and then further explain laboratory-based sensorimotor training studies that have specifically examined the role of sensorimotor experience in modulating the imitative capacity. Following the discussions of expertise effects and sensorimotor training studies, I will introduce the current study that aimed to elucidate whether the flexibility observed for the observation-execution link underlying manual gestures extends to the perception-production link underlying visual speech.

#### **4.1.1 Expertise effects**

As discussed in Section 2.1.2, the human mirror neuron system is suggested to be the neural substrate mediating imitation. Studies investigating the role of experience in the development of the mirror system have demonstrated that mirror responses to action observation change as a result of experience. Haslinger et al. (2005) were the first to investigate mirror responses to the observation of piano playing and non-piano playing finger movements in professional pianists and musically naive controls. Compared to the controls, professional pianists showed greater mirror responses when observing piano playing finger movements than when observing non-piano playing finger movements. Similarly, Calvo-Merino et al. (2005) found greater activation in the mirror system in expert dancers (i.e., ballet versus capoeira dancers) when they observed dance movements they had been trained to perform (e.g., ballet dancers watching ballet movements) relative to movements they had not (e.g., ballet dancers watching capoeira movements). Both studies demonstrated *expertise effects* in which professionals show greater mirror responses to the observation of movements within their area of expertise, relative to the observation of movements outside their area of expertise or relative to naive control participants. Additionally, in another study, Calvo-Merino et al. (2006) found greater activation in the mirror system in dancers when they observed dance movements they had performed relative to opposite-gender movements they had seen equally often but had never performed. This finding indicated that sensory experience alone cannot sufficiently explain expertise effects. Based on their findings, Calvo-Merino et al. (2005, 2006) suggested that how the mirror system responds to action observation is influenced by the observers' specific motor expertise, emphasizing the role of motor experience in facilitating the development of the mirror system. However, since mirrors are widely used during dance training, sensorimotor experience could have been adequately obtained as well by dancers tested in Calvo-Merino et al. (2005, 2006). Pianists tested in Haslinger et al. (2005) could also have

acquired sensorimotor experience by directly observing their own finger movements while playing pianos. Thus, though expertise effects associate response change in the mirror system with experience, they cannot unequivocally establish whether it is motor or sensorimotor experience that modulates mirror responses to action observation.

Using TMS to stimulate the lip motor area, Swaminathan et al. (2013) found higher cortical excitability measured from lip muscles during the viewing of sentences spoken in a known language (i.e., English) than during the viewing of sentences spoken in an unknown language (i.e., Hebrew). Following the associative hypothesis that sensorimotor experience modulates the perception-production matching link, Swaminathan et al. proposed that the difference in the articulatory excitability between two conditions was due to different strength of perception-production links underlying known and unknown languages. Expertise effects are observed based on complex movements (e.g., dance and speech) acquired and used in real life situations. Hence, compared to automatic imitation effects that are observed with simple manual gestures or syllable articulations, expertise effects provide relatively naturalistic examples showing how the perception-production link changes as a result of experience. However, because these complex movements are acquired and used in real life situations, it is difficult to control for types of experience involved in modulating the observation-execution matching link. In other words, expertise effects cannot unambiguously reject or support either the genetic or the associative account of the origin of humans' imitative capacity. Therefore, what is needed to complement findings of expertise effects and to further investigate the role of experience in the development of the imitative capacity is controlling for training experience in laboratory settings, and sensorimotor training studies discussed below provide such an example.

#### **4.1.2 *Sensorimotor training***

As discussed in Section 2.1.4, a series of sensorimotor training studies has been conducted to examine the role of sensorimotor experience in modulating the imitative capacity (see Catmur, 2013 for a review). Using automatic imitation effects as a behavioural measure, these training studies have tested the flexibility of the observation-execution link assumed to underlie the imitative capacity. Sensorimotor training studies are based on the assumptions that, if humans' imitative capacity is a product of associative sensorimotor learning, a short period of sensorimotor training that associates observation and execution of different actions should lead to change in automatic imitation; however, if humans are born with the capacity to imitate, then the mechanism should be resistant to the change that might weaken its capacity and automatic imitation should not change after sensorimotor training (Heyes et al., 2005). Thus, instead of controlling for types of experience involved in modulating the observation-

execution link, sensorimotor training studies have manipulated the predictive relationship between observed and executed actions and examined whether this manipulation modulates automatic imitation.

As introduced in Section 2.1.4, Heyes et al. (2005) conducted the first sensorimotor training study. They found that automatic imitation of hand opening/closing movements was abolished (i.e., response times for compatible and incompatible trials were not significantly different) following a short period of counter-mirror training that associated observation and execution of different movements. This finding was in line with the associative hypothesis that concurrent experience of observing and executing different actions results in inhibitory links between sensory and motor representations of the same actions. Consequently, observing an action no longer activates motor representations of that same action and thus there is no facilitation in responses in compatible trials relative to incompatible trials, which eventually reduces automatic imitation. Also in Heyes et al., automatic imitation of the same hand movements was preserved following mirror training that associated observation and execution of the same movements. However, because automatic imitation was not measured before training, it was hard to conclude whether automatic imitation changed, or say, increased following mirror training.

Press, Gillmeister, & Heyes (2007) employed both human and robotic hand opening/closing movements to elicit automatic imitation and observed a human bias with a larger automatic imitation effect elicited by human movements than by robotic movements (also observed in Press, Bird, Flach, & Heyes, 2005). After giving participants either mirror or counter-mirror training with robotic movements, Press et al. found that automatic imitation elicited by robotic movements increased following mirror training and decreased following counter-mirror training and the human bias was eliminated after mirror training. The authors proposed that participants initially had less experience in which action execution was paired with the observation of the same non-biological action than with the observation of the same human action. Hence, automatic imitation for robotic movements was initially smaller than that of human movements, but mirror training enhanced the automatic imitation effect elicited by robotic movements and thus eliminated the human bias. The finding that automatic imitation for robotic movements increased following mirror training was in line with the associative hypothesis that concurrent experience of observing and executing the same actions results in excitatory links between sensory and motor representations of the same actions. Accordingly, in Press et al., mirror training strengthened the link between observation of a robotic movement and execution of the same movement by the observer, thus enhancing the corresponding automatic imitation effect.

By showing that mirror and counter-mirror training leads to divergent changes in automatic imitation of manual gestures, sensorimotor training studies seem to favour the

associative account over the genetic account. However, as discussed in Section 2.1.3, further research is needed as all training studies so far have been based on visually transparent manual gestures, whereas one of the problems that remain unclear concerns the flexibility of the observation-execution link underlying orofacial movements that are visually opaque. Different developmental trajectories have been proposed for the mirror systems underlying manual and orofacial movements, with the former being learned and the latter being innate (Casile et al., 2011). In contrast, the associative account emphasizes the contributions of social factors to the development of humans' capacity to imitate both orofacial and manual movements (Heyes, 2005; Ray & Heyes, 2011), and thus the imitative capacity can be acquired through sensorimotor experience regardless whether movements are visually transparent or opaque.

#### **4.1.3 The current study**

The current study focused on automatic imitation of visually perceived speech actions. These actions are communicative orofacial gestures seen in face-to-face conversations that cannot be seen by the talkers themselves and are therefore visually opaque. As discussed in Section 2.2.1, previous studies have observed automatic imitation effects for visual, auditory, and audiovisual speech, demonstrating the imitative tendency to imitate perceived speech actions. Additionally, fMRI and TMS studies have demonstrated that watching and/or listening other people speak activates articulatory motor regions (e.g., Fadiga et al., 2002; Nuttall et al., 2016; Okada & Hickok, 2009; Pulvermüller et al., 2006; Wilson et al., 2004), suggesting that speech perception and production systems are closely linked. Furthermore, findings from sensorimotor adaptation studies discussed in Section 2.2.3 have indicated a certain degree of flexibility in the speech perception-production link, allowing for sensorimotor mappings between noncorresponding motor commands and altered acoustic signals. Thus, according to these findings, the perception-production link underlying speech actions seems to result from and is possibly subject to change through associative sensorimotor learning.

As mentioned in Section 2.3, the objective of the current study was:

- **Objective 3:** to determine whether sensorimotor experience is critically involved in establishing the perception-production matching link underlying visually perceived speech actions.

The speech SRC task used in Experiment 1 in Kerzel & Bekkering (2000) and also used in Experiment 1 reported in Chapter 3 was adopted to establish participants' initial automatic imitation effects before assigning them to either a counter-mirror (i.e., say /ba/ when the

speaker says /da/ and vice versa) or a mirror (i.e., say /ba/ when the speaker says /ba/ and likewise for /da/) training group. In the speech SRC task, participants were instructed to respond to printed syllable prompts “ba” or “da” by producing the corresponding syllables and to ignore the speaker who was simultaneously mouthing either /ba/ or /da/ in distractor videos. Results from studies reported in Chapter 3 suggested that using syllable prompts will result in a combined SSC+SRC effect where performance is modulated by the relative correspondence between distractors and prompts and by the relative correspondence between distractors and responses. In other words, the automatic imitation effect will be confounded with an SSC effect resulting from the overlap between distractors and prompts. However, despite the confounding SSC effect, syllable prompts were used in the current study. This decision was made based on the concern that using symbol prompts may introduce *familiarity effects* in which participants become familiar with the arbitrarily assigned symbol-syllable pairs (e.g., say /ba/ when seeing the prompt “##”) and this increased familiarity may bring systematic change in automatic imitation from pre-training to post-training sessions in both groups. In contrast, SSC effects resulting from the overlap between distractors and syllable prompts were expected to change to a lesser extent after training than the familiarity effects. That is, participants were expected to have plenty of experience of reading text and so their ability to respond to a syllable prompt by producing that syllable was not expected to change much from pre-training to post-training sessions. Automatic imitation effects were measured again using the same task but approximately 24 hours after training. Automatic imitation effects were measured before and after training to ensure whether the final results were confounded with any pre-training differences between two groups. It was predicted that, following the associative hypothesis, automatic imitation effects were expected to increase following mirror training and decrease following counter-mirror training. Alternatively, following the genetic hypothesis, automatic imitation effects were expected to remain the same between two groups after training. Additionally, according to findings from studies reported in Chapter 3, automatic imitation of speech varies when prompts are presented at different time points relative to the distractor onset (i.e., SOAs) and so different SOAs were included to examine whether training effects would interact with automatic imitation effects at different SOAs.

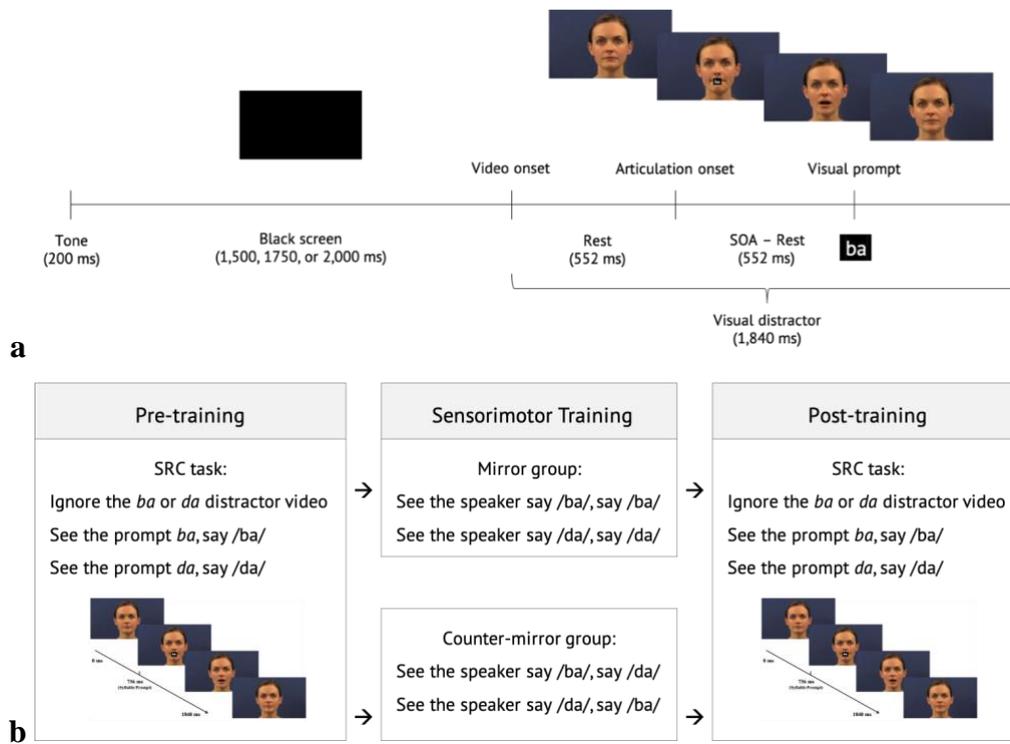
## 4.2 Method

### *Participants*

An a-priori power analysis was conducted using G\*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) with an effect size  $\eta_p^2 = .108$ , obtained from a pilot study. Sixty-two participants were needed to detect significant sensorimotor training effects on automatic imitation with a

power of .80 and an  $\alpha$  of .05. Sixty-eight participants were recruited, but one was excluded for not being a native British English speaker, one for having dyslexia, one for not attending the post-training session and three for performing at chance level during training. The final analysis included 31 participants in the mirror group (23 female,  $M_{age} = 21.71$ ,  $SD_{age} = 4.97$ ,  $range_{age} = 17-34$ ) and 31 in the counter-mirror group (19 female,  $M_{age} = 21.61$ ,  $SD_{age} = 3.30$ ,  $range_{age} = 18-30$ ). All were native British English speakers with self-reported normal or corrected-to-normal vision, normal hearing and no speech or language disorders, or other neurological disorders. Participants received £20 or course credit. The University Research Ethics Committee approved the procedures and all participants gave written informed consent.

### Material and stimuli



**Figure 4.1** **a** Schematic timeline of a trial presented in either the pre-training or post-training session. **b** The experimental procedure with pre-training, training, and post-training sessions. Participants were randomly assigned to either mirror or counter-mirror group.

Stimuli (see Figure 4.1) included the same silent distractor videos of a speaker saying /ba/ and /da/ and printed syllable prompts *ba* and *da* used in the syllable condition in Experiment 1 reported in Chapter 3. The experiment was performed using Presentation.

### ***Design and procedure***

The experiment included two testing sessions (pre- versus post-training) and one training session (see Figure 4.1b) and took place in a soundproofed, light-controlled booth. The pre- and post-training sessions were identical to the syllable session in Experiment 1 in Chapter 3; except that, for the post-training session, participants were tested a day later and only five practice trials were given. Training took place immediately following the first testing session, and participants were randomly assigned to one of the two training groups. Participants in the counter-mirror group had to say /ba/ as soon as they saw the speaker say /da/ and vice versa, whereas participants in the mirror group had to repeat the syllable produced by the speaker in the distractor video. No prompt was presented during training, and participants were only responding to the speaker. The jittered intertrial intervals (1,500, 1,750, or 2,000 ms) were randomly given to prevent the predictability of the trial onset. There were twelve blocks with 80 trials each (960 trials in total) in the training session, and six trial conditions (2 Videos x 3 Jitters) were repeated in a randomised order. Participants could choose whether to take a break between blocks, and after the first eight blocks, a short silent animation of Tom and Jerry downloaded from the Internet was played silently before they continued to the second half of the session. Five practice trials were given before the first block and the entire session lasted about 90 min. In total, the whole experiment lasted about two hours on the first day and 30 min on the second day.

### ***Data recording and analysis***

Technical equipment used to record responses and data processing methods were identical to those in Experiment 1 in Chapter 3. RTs in testing sessions were measured relative to the prompt onset, and RTs in training sessions were measured relative to the video onset. For both testing and training sessions, incorrect responses, missed trials, anticipatory responses with RTs < 100 ms, and outliers with RTs more than three standard deviations away from corresponding condition means per syllable response were classified as errors and excluded from the RT analysis. A natural log-transformation was applied to RTs for statistical analyses, but figures present back-transformed RTs.

ERs and RTs collected from testing sessions were subjected to separate repeated-measures ANOVA with test (pre- versus post-training), compatibility (compatible versus incompatible), and SOA (552, 736, 920 or 1,104 ms, also referred to as SOA1, 2, 3, and 4) as within-subjects variables and with training (mirror versus counter-mirror) as a between-subjects variable. Expected statistical results for each prediction for Objective 3 are listed below:

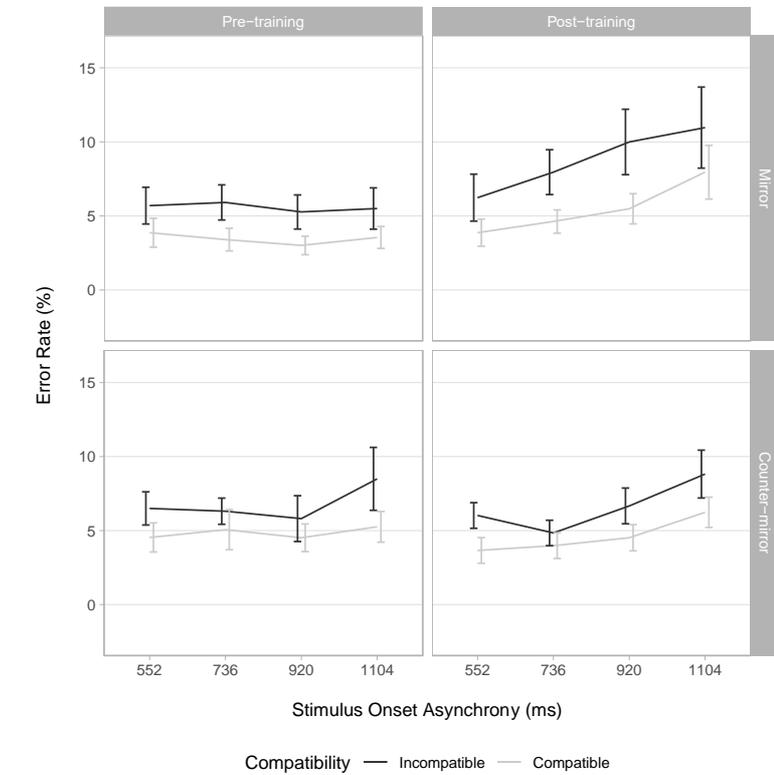
- **Prediction 9:** following the associative hypothesis, automatic imitation effects increase following mirror training and decrease following counter-mirror training – the interaction between training, test, and compatibility should be significant, and the interaction between training, test, compatibility, and SOA may or may not be significant.
- **Prediction 10:** following the genetic hypothesis, automatic imitation effects remain the same between two groups after training – neither the interaction between training, test, and compatibility nor the interaction between training, test, compatibility, and SOA should be significant.

ERs and RTs collected from training sessions were first grouped into four blocks and were then subjected to separate repeated-measures ANOVA with block (four blocks) as a within-subjects variable and with training (mirror versus counter-mirror) as a between-subjects variable. The significance level was set to  $p < .05$ . Greenhouse-Geisser correction for non-sphericity and Bonferroni-correction for multiple comparisons were applied whenever appropriate.

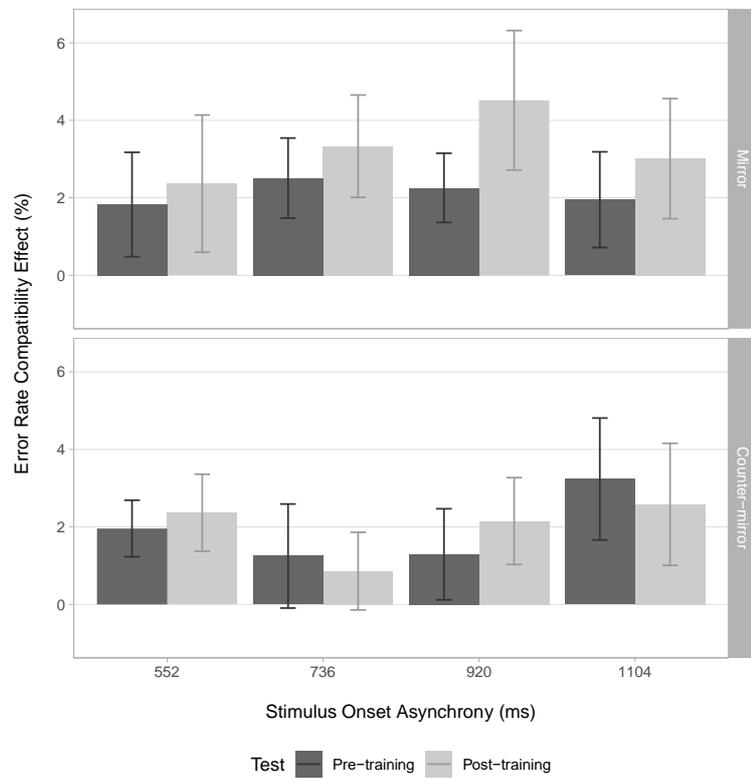
## 4.3 Results

### 4.3.1 Testing sessions

#### Error rate analysis



**a**



**b**

**Figure 4.2 a** ERs  $\pm$  SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training (bottom-right) sessions in the counter-mirror group. **b** ER compatibility effects  $\pm$  SEs for the pre- and post-training sessions at each SOA in each group.

**Table 4.1** Four-way ANOVA summary for ERs, as a function of training, test, compatibility, and SOA.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	<i>Epsilon</i>	$SS_{Num}$	$SS_{Den}$	<i>F</i>	<i>p</i>	$\eta^2_p$
Training	1	60	1.00	4.25	23830.66	0.01	.918	.000
Test	1	60	1.00	354.19	4223.91	5.03	.029	.077
Test x Training	1	60	1.00	496.46	4223.91	7.05	.010	.105
Compatibility	1	60	1.00	1360.17	4670.74	17.47	< .001	.226
Compatibility x Training	1	60	1.00	35.75	4670.74	0.46	.501	.008
SOA	1.73	103.69	0.58	633.24	6005.22	6.33	.004	.095
SOA x Training	1.73	103.69	0.58	33.59	6005.22	0.34	.684	.006
Test x Compatibility	1	60	1.00	23.21	1450.68	0.96	.331	.016
Test x Compatibility x Training	1	60	1.00	19.29	1450.68	0.80	.375	.013
Test x SOA	2.11	126.75	0.70	386.18	3496.25	6.63	.002	.099
Test x SOA x Training	2.11	126.75	0.70	89.77	3496.25	1.54	.217	.025
Compatibility x SOA	2.14	128.13	0.71	21.08	4275.40	0.30	.759	.005
Compatibility x SOA x Training	2.14	128.13	0.71	64.06	4275.40	0.90	.415	.015
Test x Compatibility x SOA	3	180	0.89	19.18	2450.37	0.47	.704	.008
Test x Compatibility x SOA x Training	3	180	0.89	5.49	2450.37	0.13	.939	.002

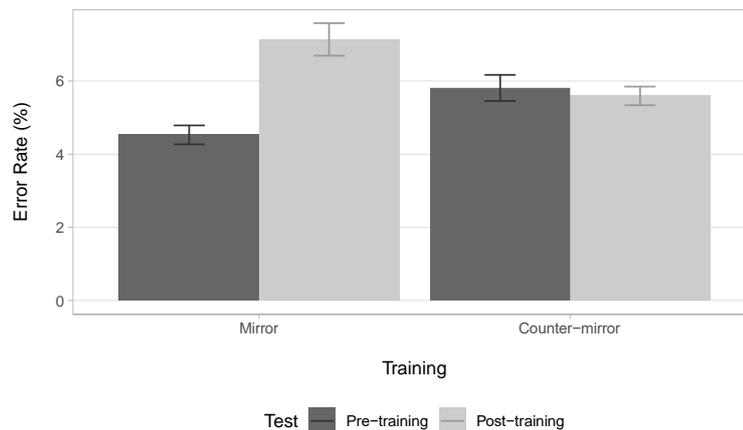
*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. *Epsilon* indicates Greenhouse-Geisser multiplier for degrees of freedom, *p*-values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

On average, participants made 5.7% errors (incorrect responses: 2.5%; missed trials: 0.7%; anticipatory responses: 1.6%; outliers: 0.9%). ER analyses (see Figure 4.2, Table 4.1, and Table 4.2) revealed a main effect of test, with a lower ER for the pre-training session than for the post-training session. The main ER compatibility effect revealed a higher ER for incompatible trials than for compatible trials ( $\Delta = 2.34$ ). Follow-up *t* tests for the main effect of SOA revealed a higher ER at SOA4 than at SOA1-3 (all  $p > .05$ ).

**Table 4.2** ERs, SEs, and 95% CIs for each level within test, compatibility, and SOA.

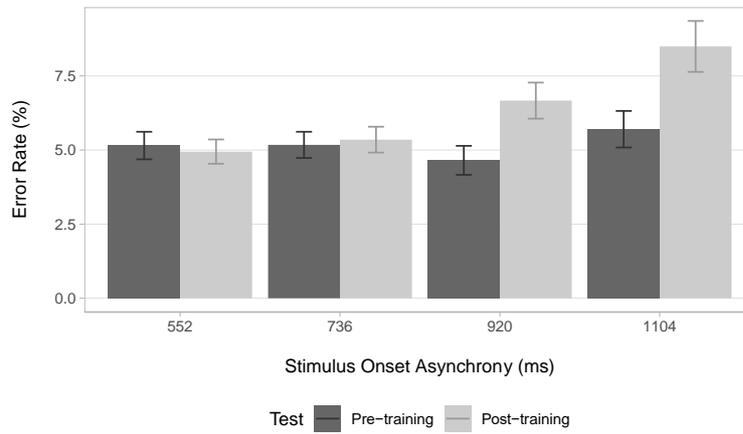
	Level	Mean	SE	95% CI
Test	Pre-training	5.17	0.63	[3.90, 6.44]
	Post-training	6.36	0.74	[4.89, 7.84]
Compatibility	Incompatible	6.94	0.83	[5.27, 8.61]
	Compatible	4.60	0.51	[3.57, 5.62]
SOA (ms)	552	5.05	0.55	[3.95, 6.15]
	736	5.26	0.58	[4.10, 6.43]
	920	5.66	0.70	[4.27, 7.05]
	1104	7.10	0.94	[5.22, 8.98]

The two-way interaction between training and test (see Figure 4.3 and Table 4.3) was significant, and follow-up *t* tests revealed a lower ER for the pre-training session than for the post-training session in the mirror group ( $\Delta = 2.61$ ,  $t(30) = 3.47$ ,  $p = .001$ ), but no difference between two testing sessions in the counter-mirror group ( $\Delta = 0.22$ ,  $t(30) = 0.29$ ,  $p = .772$ ). Further comparisons of ERs between two groups for each testing session revealed no difference between two groups in either the pre- or post-training session.

**Figure 4.3** ERs ± SEs for each testing session in each group.**Table 4.3** ER, SEs, and 95% CIs for each level within the interaction between training and test.

Training	Test	Mean	SE	95% CI
Mirror	Pre-training	4.53	0.90	[2.73, 6.32]
	Post-training	7.14	1.04	[5.06, 9.22]
Counter-mirror	Pre-training	5.81	0.90	[4.02, 7.60]
	Post-training	5.59	1.04	[3.51, 7.67]

The two-way interaction between test and SOA (see Figure 4.4 and Table 4.4) was significant, and follow-up *t* tests revealed lower ERs for the pre-training session than for the post-training session at SOA3-4 (both  $p = .007$ ).

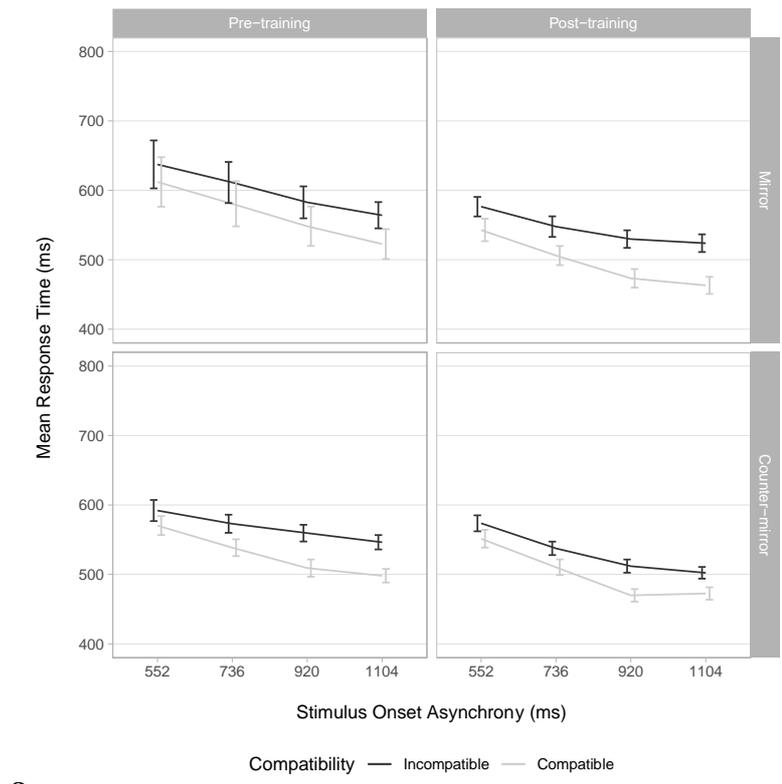


**Figure 4.4** ERs  $\pm$  SEs at each SOA in each testing session.

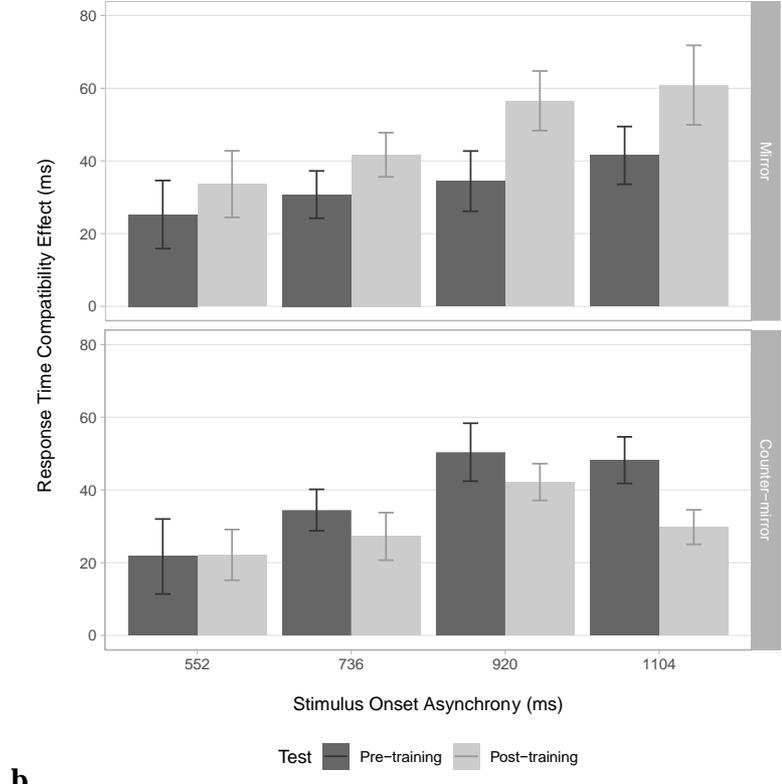
**Table 4.4** ERs, SEs, and 95% CIs for each level within the interaction between test and SOA.

Test	SOA (ms)	Mean	SE	95% CI
Pre-training	552	5.15	0.67	[3.82, 6.48]
	736	5.17	0.63	[3.92, 6.43]
	920	4.65	0.70	[3.26, 6.05]
	1104	5.70	0.87	[3.96, 7.44]
Post-training	552	4.95	0.59	[3.77, 6.13]
	736	5.35	0.61	[4.12, 6.58]
	920	6.67	0.86	[4.95, 8.39]
	1104	8.50	1.23	[6.05, 10.94]

**Response time analysis**



**a**



**b**

**Figure 4.5 a** Mean RTs  $\pm$  SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training

(bottom-right) sessions in the counter-mirror group. **b** RT compatibility effects  $\pm$  SEs for the pre- and post-training sessions at each SOA in each group.

**Table 4.5** Four-way ANOVA summary for log-transformed RTs, as a function of training, test, compatibility, and SOA.

Main Effect/Interaction	<i>df</i> <sub>Num</sub>	<i>df</i> <sub>Den</sub>	<i>Epsilon</i>	<i>SS</i> <sub>Num</sub>	<i>SS</i> <sub>Den</sub>	<i>F</i>	<i>p</i>	$\eta^2_p$
Training	1	60	1.00	0.134	15.243	0.528	.470	.009
Test	1	60	1.00	1.52	4.86	18.77	< .001	.238
Test x Training	1	60	1.00	0.09	4.86	1.12	.294	.018
Compatibility	1	60	1.00	1.33	0.52	154.73	< .001	.721
Compatibility x Training	1	60	1.00	0.01	0.52	1.34	.252	.022
SOA	1.76	105.75	0.59	2.30	0.80	173.89	< .001	.743
SOA x Training	1.76	105.75	0.59	0.01	0.80	0.39	.653	.006
Test x Compatibility	1	60	1.00	0.01	0.21	1.89	.175	.030
Test x Compatibility x Training	1	60	1.00	0.03	0.21	7.88	.007	.116
Test x SOA	2.54	152.54	0.85	0.04	0.39	6.25	.001	.094
Test x SOA x Training	2.54	152.54	0.85	0.01	0.39	1.96	.133	.032
Compatibility x SOA	2.43	145.75	0.81	0.09	0.38	14.57	< .001	.195
Compatibility x SOA x Training	2.43	145.75	0.81	0.00	0.38	0.67	.539	.011
Test x Compatibility x SOA	2.51	150.30	0.84	0.00	0.35	0.20	.866	.003
Test x Compatibility x SOA x Training	2.51	150.30	0.84	0.01	0.35	1.46	.232	.024

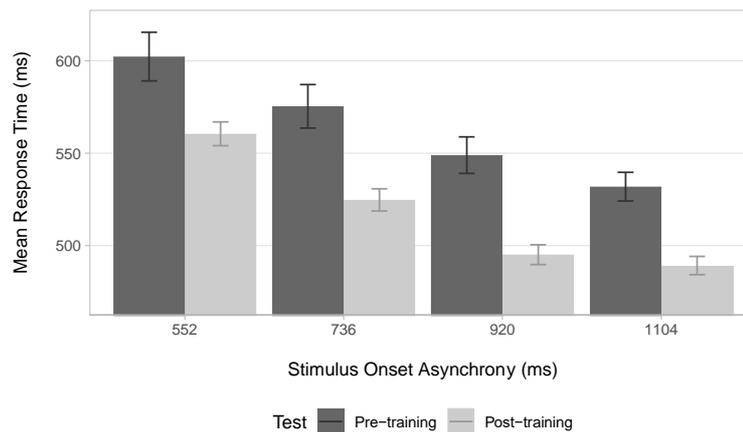
*Note.* *df*<sub>Num</sub> indicates degrees of freedom numerator. *df*<sub>Den</sub> indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, *p*-values and degrees of freedom in the table incorporate this correction. *SS*<sub>Num</sub> indicates Type III sum of squares numerator. *SS*<sub>Den</sub> indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with *p* < .05 are highlighted.

After errors were removed, log-transformed RT analyses (see Figure 4.5, Table 4.5, and Table 4.6) revealed a main effect of test, with slower RTs for the pre-training session than for the post-training session. The main RT compatibility effect revealed slower RTs for incompatible trials than for compatible trials ( $\Delta = 39$  ms). Follow-up *t* tests for the main effect of SOA revealed faster RTs for long SOAs (all *p* < .001).

**Table 4.6** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within test, compatibility, and SOA.

	Level	Mean	RT(ms)	SE	95% CI
Test	Pre-training	6.32	555	0.02	[6.28, 6.36]
	Post-training	6.24	513	0.02	[6.21, 6.27]
Compatibility	Incompatible	6.32	553	0.02	[6.29, 6.35]
	Compatible	6.24	514	0.02	[6.21, 6.28]
SOA (ms)	552	6.35	572	0.02	[6.31, 6.39]
	736	6.30	542	0.02	[6.26, 6.33]
	920	6.25	515	0.02	[6.21, 6.28]
	1104	6.23	506	0.01	[6.20, 6.25]

The two-way interaction between test and SOA (see Figure 4.6 and Table 4.7) was significant, and follow-up *t* tests revealed slower RTs for the pre-training session than for the post-training session at all SOAs (all  $p < .006$ ), but the decrease in RTs after training was smaller at SOA1 than at SOA2-3 (both  $p < .036$ ).

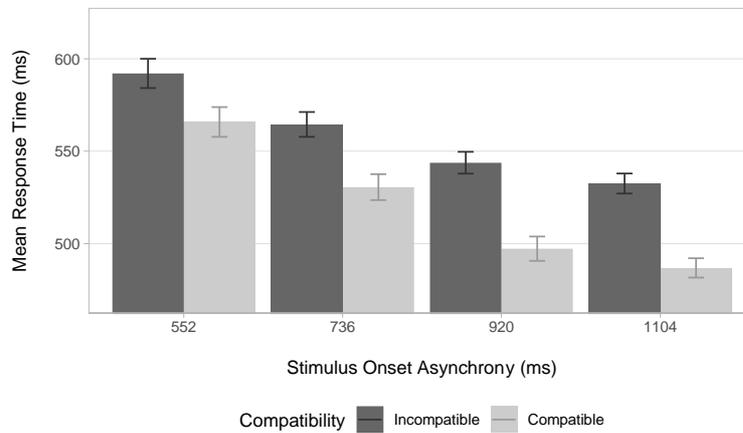


**Figure 4.6** Mean RTs ± SEs at each SOA for each testing session.

**Table 4.7** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between test and SOA.

Test	SOA (ms)	Mean	RT (ms)	SE	95% CI
Pre-training	552	6.38	590	0.03	[6.33, 6.43]
	736	6.34	565	0.02	[6.29, 6.38]
	920	6.29	541	0.02	[6.25, 6.34]
	1104	6.27	526	0.02	[6.23, 6.30]
Post-training	552	6.32	556	0.02	[6.29, 6.35]
	736	6.26	521	0.02	[6.22, 6.29]
	920	6.20	492	0.02	[6.17, 6.23]
	1104	6.19	486	0.01	[6.16, 6.22]

The two-way interaction between compatibility and SOA (see Figure 4.7 and Table 4.8) was significant, and follow-up *t* tests revealed significant RT compatibility effects at all SOAs (all  $p < .001$ ), but RT compatibility effects were smaller at SOA1-2 than at SOA3-4 (all  $p < .024$ ).



**Figure 4.7** Mean RTs  $\pm$  SEs at each SOA for each compatibility condition.

**Table 4.8** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between compatibility and SOA.

Compatibility	SOA (ms)	Mean	RT (ms)	SE	95% CI
Incompatible	552	6.37	586	0.02	[6.34, 6.41]
	736	6.33	560	0.02	[6.30, 6.36]
	920	6.29	540	0.02	[6.26, 6.32]
	1104	6.27	529	0.01	[6.24, 6.30]
Compatible	552	6.33	559	0.02	[6.29, 6.37]
	736	6.26	525	0.02	[6.23, 6.30]
	920	6.20	492	0.02	[6.16, 6.24]
	1104	6.18	483	0.02	[6.15, 6.21]

Of primary interest, the three-way interaction between training, test, and compatibility (see Figure 4.8 and Table 4.9) was significant, and follow-up *t* tests revealed significant RT compatibility effects for both pre- and post-training sessions in both groups (all  $p < .001$ ). Further comparisons of RT compatibility effects between pre- and post-training sessions for each group revealed that the RT compatibility effect increased (by 32%) from 37 ms to 49 ms following mirror training ( $\Delta = 12$  ms,  $t(30) = 2.55$ ,  $p = .016$ ) and decreased (by 21%), but not significantly, from 39 ms to 31 ms following counter-mirror training ( $\Delta = 8$  ms,  $t(30) = 1.25$ ,  $p = .221$ ). A follow-up one-way ANOVA was further conducted to test whether mirror training was more effective than counter-mirror training and revealed comparable changes in RT compatibility effects between two groups,  $F(1, 60) = 2.64$ ,  $p = .175$ .



**Figure 4.8** Mean RTs  $\pm$  SEs for each compatibility condition in each testing session in each group.

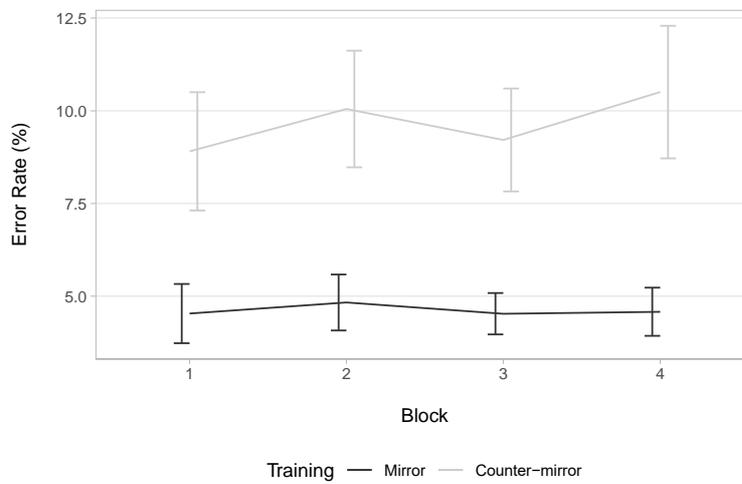
**Table 4.9** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level in the three-way interaction between training, test, and compatibility.

Training	Test	Compatibility	Mean	RT(ms)	SE	95% CI
Mirror	Pre-training	Incompatible	6.37	585	0.03	[6.31, 6.43]
		Compatible	6.31	548	0.03	[6.24, 6.37]
	Post-training	Incompatible	6.29	539	0.02	[6.25, 6.33]
		Compatible	6.19	490	0.02	[6.15, 6.24]
Counter-mirror	Pre-training	Incompatible	6.33	563	0.03	[6.28, 6.39]
		Compatible	6.26	524	0.03	[6.20, 6.33]
	Post-training	Incompatible	6.27	528	0.02	[6.23, 6.31]
		Compatible	6.21	497	0.02	[6.16, 6.25]

Though the four-way interaction between training, test, compatibility, and SOA (see Figure 4.5 and Table C.1 in Appendix C) was not significant, further comparisons of RT compatibility effects between pre- and post-training sessions for each SOA in each group were conducted. It was revealed that the RT compatibility effects at SOA3-4 increased significantly following mirror training (all  $p < .03$ ) and the RT compatibility effect at SOA4 decreased significantly following counter-mirror training ( $p = .029$ ).

### 4.3.2 Training sessions

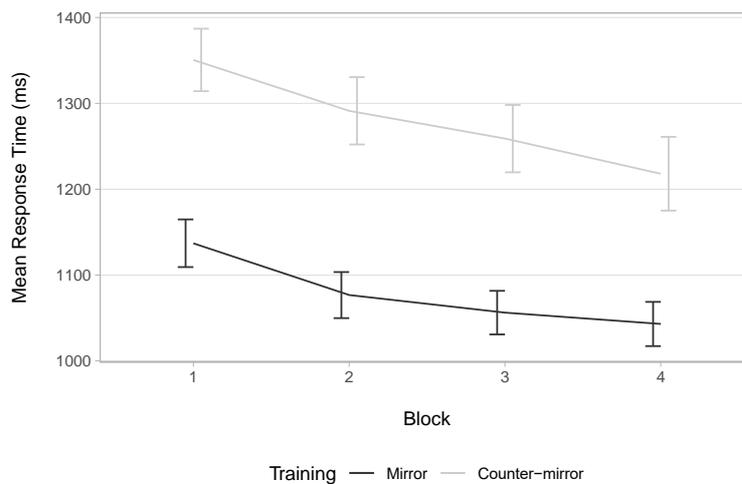
#### Error rate analysis



**Figure 4.9** ERs  $\pm$  SEs for each block in each training group.

On average, participants made 7.1% errors (incorrect responses: 4.0%; missed trials: 0.2%; anticipatory responses: 1.6%; outliers: 1.3%). ER analyses (see Figure 4.9) revealed a main effect of training,  $F(1, 60) = 9.57, p = .003, \eta^2_p = .138$ , with a lower ER for the mirror group ( $M = 4.62, SE = 1.16, 95\% CI = [2.31, 6.93]$ ) than for the counter-mirror group ( $M = 9.67, SE = 1.56, 95\% CI = [7.36, 11.98]$ ). Neither the main effect of block nor the interaction between training and block was significant.

#### Response time analysis



**Figure 4.10** Mean RTs  $\pm$  SEs for each block in each training group.

After errors were removed, log-transformed RT analyses (see Figure 4.10) revealed a main effect of training,  $F(1, 60) = 18.86, p < .001, \eta_p^2 = .239$ , with faster RTs for the mirror group ( $M = 6.97$ , back-transformed RT = 1064 ms,  $SE = 0.03$ , 95% CI = [6.92, 7.03]) than for the counter-mirror group ( $M = 7.14$ , back-transformed = 1261 ms, RT =  $SE = 0.03$ , 95% CI = [7.09, 7.19]). The main effect of block (see Table 4.10),  $F(1.96, 117.37) = 24.21, p < .001, \eta_p^2 = .287$ , revealed faster RTs for later blocks, but RTs in block 3 were not statistically different from RTs in block 2 and block 4. The interaction between training and block was not significant.

**Table 4.10** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each block.

	Level	Mean	RT(ms)	SE	95% CI
Block	1	7.11	1227	0.02	[7.08, 7.15]
	2	7.06	1164	0.02	[7.02, 7.10]
	3	7.04	1140	0.02	[7.00, 7.08]
	4	7.01	1111	0.02	[6.97, 7.06]

#### 4.4 Discussion

The objective of the current study was:

- **Objective 3:** to determine whether sensorimotor experience is critically involved in establishing the perception-production matching link underlying visually perceived speech actions.

Measured in a speech SRC task, automatic imitation effects were used to index perception-induced motor activation with a larger effect indicating greater motor activation and thus a stronger perception-production matching link presumably underlying the imitative capacity. A between-subjects training study was conducted in which participants were assigned to either a mirror or a counter-mirror training group and automatic imitation of visually perceived speech actions was measured before and after training. As discussed in Section 2.1.3, the associative account of imitation proposes that the matching properties of the perception-production link underlying the imitative capacity are acquired through sensorimotor experience; whereas the genetic account proposes that such matching properties are innate, though sensory or motor experience may facilitate the development of the matching. Thus, following the associative hypothesis, automatic imitation effects were expected to increase following mirror training and decrease following counter-mirror training. Alternatively, following the genetic hypothesis, automatic imitation effects were expected to remain the same between two groups after training.

In line with previous findings (e.g., Heyes et al., 2005; Press et al., 2007), current results demonstrated that automatic imitation increased following mirror training and decreased slightly following counter-mirror training. Therefore, consistent with Prediction 9 given in Section 4.2, current findings supported the associative hypothesis that the perception-production link underlying orofacial movements such as speech actions could be altered through sensorimotor experience.

#### ***4.4.1 Mirror and counter-mirror training***

The associative account proposes that sensorimotor experience of observing and executing the same actions establishes and reinforces excitatory matching links between sensory and motor representations of the same actions (Press et al., 2007). Accordingly, the mirror training in the current study strengthened the excitatory matching links that consequently enhanced observation-induced motor activation which in turn led to more facilitation in compatible trials relative to incompatible trials. Hence, automatic imitation was enhanced following mirror training. Furthermore, the associative account also proposes that sensorimotor experience of observing and executing different actions not only leads to excitatory non-matching links between sensory and motor representations of different actions, but also establishes inhibitory matching links between sensory and motor representations of the same actions (Heyes et al., 2005). Hence, the counter-mirror training in the current study probably contributed to inhibitory matching links that consequently reduced observation-induced motor activation, which in turn led to less facilitation in compatible trials relative to incompatible trials. Therefore, automatic imitation was reduced, though only slightly (the change from pre-training to post-training sessions was not significant when collapsed across SOAs [ $p = .221$ ] but significant at SOA4 [ $p = .029$ ]), following counter-mirror training.

#### ***4.4.2 Counter-mirror training and visual speech***

Though consistent with predictions from the associative account, the sensorimotor training effects on automatic imitation of speech actions differed from those on automatic imitation of manual gestures observed in previous research (e.g., Gillmeister, Catmur, Liepelt, Brass, & Heyes, 2008; Heyes et al., 2005). In Heyes et al., automatic imitation of hand opening/closing movements was eliminated following counter-mirror training; whereas in the current study, automatic imitation of speech actions was only slightly reduced after counter-mirror training (960 trials) that was considerably longer than the training in Heyes et al. (432 trials). The perception-production link underlying speech actions seemed to be more resilient to counter-mirror training than the link underlying manual gestures. Orofacial movements are perceptually opaque and therefore, following the associative account, people mainly gain

visual input of their own movements during social interactions. In contrast, sufficient visual feedback of one's own manual movements could be obtained through direct self-observation. Mouth and hand mirror neurons in monkeys are found to be connected to different anatomical pathways, suggesting that these two types of neurons may have distinct functions, and especially mouth mirror neurons are connected to brain regions involved in emotion/reward processing that plays a role in social activities (Ferrari et al., 2017). Imaging studies reported mirror activation in homologous regions in humans during perception and production of emotional facial expressions (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Singer et al., 2004) and of speech and music (Ackermann & Riecker, 2004; Callan et al., 2006; Riecker et al., 2000). Hence, while both manual and orofacial observation-execution links are products of sensorimotor learning, the degrees of social engagement may possibly influence the learning effects. Future studies could manipulate social factors of the training environment and examine whether and which social manipulations modulate sensorimotor training effects on automatic imitation of different types of actions.

Apart from being perceptually opaque, the speech actions used in the current study are relatively complex learned actions that consist of sequences of motor acts involving at least three articulators (jaw, tongue, lips) and that create sequences of sensory percepts. In contrast, manual actions employed in previous studies can be regarded as movements that are less complex (e.g., hand opening or closing) and require little sensory or motor learning. According to the associative account, the initial establishment of perception-production links underlying novel actions constitutes imitation learning. Novel actions are assumed to comprise new sequences of familiar actions each of which has its own perception-production link. Observing a novel action involves sequence learning that links each component in the new order. Consecutive activation of each sensory representation activates the associated motor representation in the same order, which then leads to the linkage between each motor representation. This process is suggested to be important for the observer to copy and overtly perform the observed novel movement, and the linkage of these motor representations constitutes motor learning that eventually results in a new motor unit. Instead of establishing new perception-production links for novel actions, counter-mirror training in the current study was used to re-wire the links between sensory and motor units of learned actions. The question then was whether this re-wiring process occurred at the global level where the links between sensory and motor units were altered or at the individual local levels where the links between sensory and motor sequence components were altered. The /ba/ and /da/ articulations differ in their initial consonant components: the lips moving towards each other in /ba/ and the tongue tip moving towards the ridge behind the upper teeth in /da/. Therefore, in the case of syllable articulations, the question could be reformed as whether counter-mirror training in the current study re-wired the perception-production links underlying the whole syllable units or the links

underlying the consonant components only. Future studies could examine whether and how sensorimotor training with one syllable pair (e.g., /ba/ and /da/) would modulate automatic imitation of the trained syllable pair and automatic imitation measured with the other untrained syllable pair that involves the same consonants (e.g., /bi/ and /di/). If sensorimotor training modulates the perception-production links underlying the whole syllable units, change in automatic imitation of the trained pair would be expected but not in automatic imitation of the untrained pair. Alternatively, if sensorimotor training modulates the links underlying the consonant components only, change in automatic imitation of both trained and untrained syllable pairs would be expected.

In addition, the orofacial movements employed in the current study were syllable articulations, the basic language units used in everyday speech communication. In other words, they are movements used to convey communicative intentions and meanings. Using TMS to elicit MEPs measured from both left and right hand muscles, Möttönen, Farmer, & Watkins (2010) observed no difference in excitability between left and right motor cortex when participants were watching bimanual gestures but not aware of their communicative nature. However, after participants learned that these were sign language gestures and were taught the meanings of half of them, excitability was increased in the left, but not right, motor cortex for both learned and unlearned gestures. Möttönen et al. proposed that observers' awareness that the perceived manual gestures were communicative movements probably strengthened sensorimotor resonance in the left motor cortex and thus suggested that perception-induced motor activation is likely to be affected by cognitive factors. In the case of the current study, it could be syllable articulations' communicative nature that enhanced participants' imitative tendency and thus interfered with counter-mirror training. Future research could first elucidate whether automatic imitation of non-communicative orofacial movements, such as lip and tongue protrusion, would be smaller than that of syllable articulations and then investigate whether the perception-production matching link underlying non-communicative orofacial movements is more susceptible to counter-mirror training. Likewise for manual gestures, studies could examine whether automatic imitation of manual gestures would increase if participants become aware of their communicative nature and whether the observation-execution link underlying communicative manual gestures is more resilient to counter-mirror training.

Furthermore, different selections of SRC tasks between Heyes et al. (2005) and the current study could be part of the reason why counter-mirror training was more effective in the former. Heyes et al. employed a simple response task that elicited faster responses than the choice response task adopted in the current study, and faster responses were found to result in smaller automatic imitation effects (Brass et al., 2001). In their meta-analysis paper of automatic imitation, Cracco et al. (2018) confirmed that choice response tasks tend to elicit

larger automatic imitation effects than simple response tasks. The authors suggested that it is because choice response tasks involve processes of motor selection and motor planning both of which interact with the perception of actions shown in distractor videos, whereas simple response tasks only involve motor planning. Accordingly, it was possible that counter-mirror training was less effective in the current study because the adopted speech SRC task involved more processes that were being altered and was thus more resilient to counter-mirror training. Moreover, syllable prompts were selected in the current speech SRC task with awareness that these prompts will introduce a confounding SSC effect. This decision was made with an aim to avoid learning effects in which participants may become familiar with the assigned symbol-syllable pairs (e.g., say /ba/ when seeing a prompt “##”). Thus, in the current study, the remaining automatic imitation effect following counter-mirror training might be the SSC effect resulting from the overlap between distractors and prompts. Thus, it was difficult to determine whether the difference between our results and findings from Heyes et al. was due to different experimental designs or due to the nature of actions used in two studies. However, the primary interest of sensorimotor training studies has always been whether different types of sensorimotor training lead to opposite changes in automatic imitation, and answers from the current study and from previous training studies, including Heyes et al., have been consistently positive. Thus, it could be concluded that observation-execution links underlying both manual and orofacial actions can be modulated through sensorimotor training, but the extent of training effects on different actions requires further investigation.

#### ***4.4.3 Mirror training and the simulation theory of speech perception***

As mentioned in Section 4.1.1 in this chapter, Swaminathan et al. (2013) found higher cortical excitability measured from lip muscles during the viewing of sentence spoken in a known language (i.e., English) than during the viewing of sentence spoken in an unknown language (i.e., Hebrew). Following the associative hypothesis, the authors proposed that the difference in the articulatory excitability between two conditions was due to different strength of perception-production links underlying known and unknown languages. In speech science, the Simulation Theory of Speech Perception has proposed that perception-induced motor activation facilitates prediction of the incoming signals, which in turn supports speech comprehension (Gambi & Pickering, 2013; Pickering & Garrod, 2013). Critically, according to this stimulation account, greater motor involvement is proposed to occur when the observer is familiar with perceived speech, but which type of experience (e.g., sensory, motor, or sensorimotor) is crucially involved remains unclear. Swaminathan et al. supported the simulation hypothesis that more experience with perceived speech leads to greater perception-induced motor activation. The current finding that automatic imitation of speech increased

following mirror training further suggested that it is sensorimotor learning that facilitates perception-induced motor activation. The question of whether greater perception-induced motor activation leads to better speech comprehension is beyond the scope of the current study. However, in another training study, overt imitation of accented speech was found to improve participants' comprehension of that speech relative to the baseline condition, but such improvement was not observed in either the listening-only condition or the repeating-without-imitation condition (Adank, Hagoort, & Bekkering, 2010). Hence, results from Adank et al. further supported the notion that sensorimotor learning, rather than sensory and/or motor learning alone, leads to enhanced perception-induced motor activation which then facilitates speech comprehension. Interestingly, however, Schmitz et al. (2018) stimulated the lip motor cortex with TMS and found that listening to non-native vowels elicited higher articulatory excitability than native-like vowels, contradictory to what was found in Swaminathan et al. where sentence articulations were presented visually. Future research that controls linguistic levels and stimulus modalities is needed to investigate this inconsistency. Moreover, follow-up research could extend current findings by examining whether sensorimotor training modulates automatic imitation of auditory speech. Behavioural research could also examine automatic imitation of non-native speech and investigate the role of sensorimotor learning in establishing the perception-production link underlying second language (L2) processing.

#### ***4.4.4 Second language research and imitation research***

Importantly, if sensorimotor learning is critically engaged in the creation and facilitation of perception-production links underlying speech, this sensorimotor training paradigm could possibly be adapted and applied to L2 learning research. Previous studies have found that L2 learners' perception and production abilities do not always develop in parallel (Flege, MacKay, & Meador, 1999; Peperkamp & Bouchon, 2011); that is, improvement in one's L2 perception does not predict the improvement in his/her L2 production. More importantly, though perceptual training effects in L2 learning were found to transfer to the production system, the degrees of improvement in the two systems were uncorrelated (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). Similarly, in a L2 production training study, production training gave rise to improved perception, but the correlation between perception and production improvements was not significant (Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015). Furthermore, in another L2 production training study, no improvement in perception was observed following production training (Hattori, 2010). These studies together seem to suggest that L2 perception and production systems may be dissociated and that either sensory or motor training is insufficient to build imitative perception-production links underlying foreign speech sounds. Follow-up studies could examine and compare sensory, motor, and

sensorimotor training effects on learners' abilities to perceive and produce L2 speech sounds and test whether sensorimotor training leads to correlated improvements in the two systems. However, these are suggestions based on the current study that used visual speech stimuli, and vision is not the dominant sensory modality that people use during vocal communication. To further explore the role of sensorimotor learning in the perception-production link underlying speech, future research should first extend the current findings to the auditory modality by examining whether sensorimotor experience modulates audio-motor links underlying speech actions.

With its multimodal nature and hierarchical structure, speech provides a rich resource for imitation research in cognitive psychology. At the same time, however, one should be careful when comparing findings from studies using different speech stimuli (e.g., syllable vs. sentence). To further understand automatic imitation of speech, systematic investigation that controls for linguistic levels and stimulus modalities is needed. Additionally, investigation that compares different measures of observation-induced motor activation is also required. In the past research, automatic imitation of manual gestures was reversed in TMS studies using neurophysiological measures (Catmur et al., 2007; Cavallo, Heyes, Becchio, Bird, & Catmur, 2014). Yet, automatic imitation was only eliminated, but never reversed, in behavioural studies (Catmur & Heyes, 2017; Heyes et al., 2005). Since there is now substantial evidence that sensorimotor links underlying manual gestures are flexible (Catmur, 2013; R. Cook et al., 2014), follow-up research should further investigate the degree of flexibility of these links and to what extent they can be modulated through training, and the same questions need to be answered for orofacial gestures including speech.

#### **4.5 Conclusion**

In conclusion, results from the current study demonstrate that automatic imitation of visually perceived speech actions is enhanced following mirror training and reduced following counter-mirror training, suggesting that sensorimotor learning plays a critical role in linking speech perception and production systems and that the link between these two systems remains flexible in adulthood. Therefore, the current findings support the associative hypothesis that the matching properties of the perception-production link underlying both manual and orofacial movements are dependent on sensorimotor experience. The current study set out to explore commonalities between speech actions and manual gestures, two groups of movements that are qualitatively different in terms of effectors used for production (articulators versus hands) and dominant sensory modalities involved in perception (auditory versus visual). Despite of these differences, findings of the current study indicate that the same associative mechanism is involved in modulating the imitative capacity underlying both types

of movements and therefore advance the associative account of imitation by verifying and extending its predictions to a different effector system than the one commonly examined in the previous literature.

Crucially, compared to what has been observed in the hand literature, current results demonstrate that automatic imitation of speech actions seems to be more resilient to counter-mirror training than automatic imitation of manual gestures. This is suggested by the observations that, in Heyes et al. (2005), automatic imitation of hand opening/closing movements was eliminated following counter-mirror training; whereas, in the current study, automatic imitation of speech actions was only slightly reduced after counter-mirror training (960 trials) that was considerably longer than the training in Heyes et al. (432 trials). This indicates that, while the same associative mechanism is likely to be involved in imitation of both speech actions and manual gestures, the effectiveness of associative sensorimotor experience in modulating the perception-production link underlying each type of movements may depend on various factors. The Discussion section (Section 4.4) in this chapter has proposed several research avenues for elucidating possible qualitative differences between sensorimotor coding for manual and speech actions. Different from manual gestures (e.g., hand opening/closing and finger tapping/lifting) used in the previous research, speech actions are multimodal orofacial movements commonly used during social interactions to convey communicative intentions and meanings. Hence, the multimodal and communicative nature of speech actions and the social environment where these actions are commonly used may account for current findings. In other words, the effectiveness of associative sensorimotor experience in modulating the imitative capacity underlying certain types of movements is likely to result from the interaction between the movement nature and the context where these movements are commonly used. Thus, though slightly different from previous findings, current results are also consistent with the associative account of imitation which emphasises social factors that affect the development of humans' imitative capacity. Accordingly, research of speech imitation may provide a novel perspective to studying imitation and thus complementing findings of imitation in cognitive psychology.

Furthermore, as mentioned above, current findings suggested that sensorimotor learning plays a critical role in linking speech perception and production systems and that the link between these two systems remains flexible in adulthood. Discussed in Section 2.2.3, sensorimotor adaptation studies have reached a similar conclusion that remapping between perception and production systems occurs after receiving altered auditory feedback. In other words, both findings from the current study and from sensorimotor adaptation studies indicate a certain degree of separation between speech perception and production systems and therefore, a certain degree of flexibility in the perception-production link, allowing for mapping between sensory and motor representations of either the same or different speech

actions. This has important implications for L2 research which, as discussed in Section 4.4.4, has observed that improvement in one modality (e.g., perception) after training does not always lead to improvement in the other modality (e.g., production) and thus L2 perception and production systems seem to be somehow dissociated as well. Borrowing the idea of sensorimotor training from cognitive psychology, the current study suggests a perspective that, while both perception and production training is important in refining learners' ability in each modality, sensorimotor training with correctly produced and perceived speech sounds may play a crucial role in linking the two systems and preventing learners from incorrectly associating a non-native sound with a similar native sound. Future L2 studies could further develop the sensorimotor training paradigm used in the current study and investigate whether sensorimotor training facilitates L2 learning.

Finally, it is also important to note that current results do not allow us to completely exclude the possibility of an innate mechanism governing imitative, rather than counter- or non-imitative, sensorimotor associations underlying the imitation of orofacial movements. Heyes (2011) also acknowledges that results from sensorimotor training studies in principle "do not exclude a role for genetic pre-specification in establishing the long-term sensorimotor connections that generate automatic imitation" (p. 478). Nevertheless, current results are in line with the main associative hypothesis that it is sensorimotor experience, but not sensory or motor experience alone, that configures observation-execution links (Heyes, 2010), since the only difference between the two groups in the current study is the relationship between observed and executed movements per trial during training.

Overall, the current research adds to the work of sensorimotor training studies in showing that associative sensorimotor learning also modulates automatic imitation of visually perceived speech actions that, different from visually transparent manual gestures, are visually opaque. In doing so, the current work contributes to research questions concerning the flexibility of the imitative mechanism underlying various types of movements and to the growing body of research on the link between speech perception and production systems.

## **Chapter 5 Sensorimotor training effects on automatic imitation of non-native speech**

### **5.1 Introduction**

As discussed in Section 2.1.3, two different accounts have been proposed to explain the origin of humans' imitative capacity, i.e., the ability to quickly map observed actions onto their motor repertoire. According to the genetic account, the matching properties of the observation-execution link underlying the imitative capacity are innate and genetically predisposed, though sensory and/or motor experience may facilitate the development of this capacity (Ferrari et al., 2003; Meltzoff, 2002; Meltzoff & Decety, 2003; Rizzolatti et al., 1996). In contrast, according to the associative account, such matching properties are forged by sensorimotor experience that consists of concurrent observation and execution of the same actions, and this sensorimotor experience can be acquired through direct self-observation and/or through social interactions (Catmur et al., 2009; Heyes, 2001, 2005, 2010; Ray & Heyes, 2011). To examine the dispute between the genetic and the associative accounts, a series of sensorimotor training studies has been conducted to investigate whether varying the relationship between concurrently observed and executed actions modulates people's imitative capacity which is measured behaviourally using automatic imitation (see Catmur, 2013 for review). As discussed in Section 2.1.4 and Section 4.1.2, training studies have observed enhanced automatic imitation following mirror training during which participants observe and execute the same manual gestures in each trial and reduced automatic imitation following counter-mirror training during which participants observe and execute different manual gestures in each trial. Findings from these training studies suggest that sensorimotor experience modulates the matching properties of the observation-execution link underlying manual gestures, thus favouring the associative account over the genetic account. Moreover, the training study reported in Chapter 4 further demonstrated that mirror and counter-mirror training also led to divergent changes in automatic imitation of visually perceived speech actions. Findings reported in Chapter 4 not only supported but also advanced the associative account by verifying and extending its predictions to a different effector system than the one commonly examined in previous literature (i.e., orofacial versus manual).

Importantly, the associative account also proposes that the same associative mechanism is involved in imitation of both familiar and novel movements that were or were not part of the observer's motor repertoire (Heyes, 2005). Sensorimotor training studies so far, including the study reported in Chapter 4, have only examined whether associative sensorimotor experience modulates the imitative capacity underlying familiar actions that participants have performed and perceived throughout their life. However, it remains

unknown whether and to what extent associative sensorimotor experience modulates people's imitative capacity underlying novel actions. As discussed in Section 5.1.1 below, neuroimaging studies have examined how short-term learning experience modulates responses in the human mirror system during the observation of newly learned novel dance sequences (Cross, Hamilton, & Grafton, 2006; Cross, Hamilton, Kraemer, Kelley, & Grafton, 2009; Cross, Kraemer, Hamilton, Kelley, & Grafton, 2009). These studies have found greater mirror responses during the observation of dance sequences learned through sensorimotor training than during the observation of dance sequences learned through sensory training and thus seem to support the associative hypothesis. However, as further discussed below, these dance training studies share the same problem with studies showing expertise effects mentioned in Section 4.1.1; that is, motor and sensorimotor experience has been confounded in all these studies and thus their results cannot unequivocally support either the associative or the genetic account.

Furthermore, as discussed in Section 2.1.4, another neuroimaging study by Wiggett et al. (2012) examined responses in the human mirror system to the observation of newly learned sequences of simple hand movements following different types of training (visual, motor, or visuomotor). Wiggett et al. found that the observation of visuomotor-trained movement sequences led to greater activation in the human mirror system, as compared to the observation of either motorically or visually trained movement sequences. However, findings from Wiggett et al. cannot unambiguously support either the associative or the genetic account because their results can be explained in different ways. The observation of visuomotor-trained movement sequences elicited greater mirror activation probably because repetitive co-activation of sensory and motor representations of the same movement sequences during visuomotor training led to the formation of the perception-production matching link underlying these movement sequences; yet, such co-activation was absent during either visual or motor training. Alternatively, it was also possible that visuomotor training led to greater mirror activation because, compared to either visual or motor training, there was always additional motor output or visual input during visuomotor training, which probably facilitated the further development of the human mirror system. Moreover, because both dance training studies and Wiggett et al. have employed novel movements that consist of familiar elements arranged in a new sequence, mirror activation by the observation of newly learned movement sequences could be confounded with mirror activation by the observation of familiar sequence components. Therefore, whether associative sensorimotor experience plays a key role in modulating people's imitative capacity underlying novel actions requires further investigation.

As mentioned above, both dance training studies and Wiggett et al. (2012) have employed novel movements that consist of familiar components arranged in a new sequence. However, Heyes (2016) has also clarified that actions can vary on many dimensions and thus

can be defined as novel in various ways. For instance, compared to a familiar action, a novel action could be performed faster, with a greater force, with familiar elements arranged in a new sequence, or by using different effectors to execute a familiar action (i.e., an unfamiliar effector-movement combination). To test whether the associative mechanism is involved in imitation of novel actions, the current study examined whether the imitative capacity underlying novel non-native speech actions is subject to change through sensorimotor training. The novel non-native speech actions employed in the current study consist of unfamiliar effector-movement combinations that native British English speakers find difficult to produce.

In the following sections, I will first discuss dance training studies that provide relatively naturalistic examples showing how short-term training experience leads to change in mirror responses to the observation of newly learned dance sequences. I will then further explain how speech constitutes a distinctive group of movements and offers a way to investigate the role of sensorimotor experience in modulating the imitative capacity underlying novel actions. Lastly, I will introduce the current study that aimed to elucidate whether the imitative capacity underlying novel actions is subject to change through sensorimotor training and at what level training-induced modification occurs in the perception-production link.

### ***5.1.1 Imitation learning***

Expertise effects discussed in Section 4.1.1 suggest that responses in the human mirror system change as a result of experience. Specifically, greater mirror responses are shown in professionals during the observation of movements within their area of expertise, relative to the observation of movements outside their area of expertise. However, by comparing overlearned movements that participants have practiced for years and novel movements that they have never performed or rarely seen, expertise effects tell little about how the observation-execution matching link emerges in the first place when motor practice is still needed to perfect movement performance. This section discusses neuroimaging studies that have investigated whether and how short-term training modulates mirror responses to the observation of newly learned novel movements (Cross et al., 2006; Cross, Hamilton, et al., 2009; Cross, Kraemer, et al., 2009). As discussed in detail below, these short-term training studies have demonstrated that sensorimotor training leads to greater mirror responses than sensory training. More importantly, however, these training studies also share the same problem with studies showing expertise effects; that is, motor and sensorimotor experience has been confounded in all these studies and thus their findings cannot unequivocally support the associative or the genetic account of imitation.

Cross et al. (2006) first investigated whether and how short-term learning experience modulates responses in the human mirror system during the observation and imagery of newly learned dance sequences. Expert dancers were scanned repeatedly over a five-week period during which they learned novel dance sequences by following a model dancer in a studio. Brain activity was recorded weekly as dancers observed and imagined performing different dance sequences, half of which were rehearsed (i.e., trained sequences) and half of which were only seen during scanning (i.e., untrained sequences). Greater mirror responses were found during the observation of trained sequences, as compared to the observation of untrained sequences. Additionally, dancers' self-rated motor ability to perform trained sequences was positively correlated with mirror responses to the observation of these sequences (as compared to the observation of untrained sequences), indicating that the observation-execution matching link underlying novel movements was strengthened as performers' motor competence improved through training.

Similar results have also been observed in studies that recruited nonexpert dancers (Cross, Hamilton, et al., 2009; Cross, Kraemer, et al., 2009). In these two studies, participants received daily dance training for one week and were scanned before and after training. While physically practicing novel dance sequences during training, participants watched videos of a dancer performing the same dance sequences. Greater mirror responses were found during the observation of physically trained dance sequences as compared to the observation of untrained dance sequences that were only seen during scanning, and behavioural results suggested that participants' performance improved after training. More importantly, Cross, Kraemer, et al. included a watched condition where participants watched videos of the dancer performing novel dance sequences but never physically performed these sequences themselves (i.e., sensory training). Greater mirror responses were found during the observation of watched dance sequences as compared to the observation of untrained dance sequences, and greater mirror responses were also found during the observation of physically trained dance sequences as compared to the observation of watched dance sequences. These findings suggested that sensory experience alone in the watched condition also contributed to the formation/strengthening of the observation-execution matching link underlying novel movements, though to a lesser degree than when physical practice was also involved. Additionally, behavioural results revealed comparable improvement in performance after training in the trained and the watched conditions. Though there was some inconsistency between neuroimaging and behavioural results, both seemed to suggest that sensory experience alone also facilitates the development of the imitative capacity underlying novel movements and its proposed neural substrate, the human mirror system. However, as argued below, further exploration of the associative hypothesis of imitation learning suggests that findings from Cross, Kraemer, et al. could also be explained by the associative account.

As mentioned in Section 4.4.2, the associative account proposes that observing a novel movement initiates a sequence learning process (i.e., observational learning) during which sensory representations of the sequence components are activated and “horizontal” links are consequently formed between these sensory representations (Heyes, 2005). Critically, these horizontal links are assumed not to be specific to imitation, but the formation of these links allows an imitation-specific process to occur. Novel movements consist of familiar elements each of which has its own observation-execution matching link (also called a “vertical” link). As a result, successive activation of sensory representations of the sequence components leads to activation of the corresponding motor representations in the same order through the vertical links, which is suggested to essentially provide the potential for the observer to copy and overtly perform the observed novel movement. Importantly, repetitive activation of motor representations in the new order leads to the formation of horizontal links between these motor representations, and the linkage of these representations constitutes motor learning that eventually results in a new motor unit. The associative account suggests that these processes constitute imitation learning, which occurs when motor learning is facilitated by visual input from the other person’s behaviours.

Novel dance sequences employed in Cross, Kraemer, et al. (2009) consisted of simple foot stepping movements arranged in complex orders. Thus, according to the associative account, their findings of enhanced mirror responses and improved behavioural performance following sensory training could result from repetitive observation-induced motor activation of the sequence components. In other words, similar to sensorimotor training, there was also repetitive co-activation between sensory and motor representations of the same movements during sensory training, which then probably contributed to the formation of the observation-execution matching link underlying these novel movements. Crucially, Cross, Kraemer, et al. also found that sensorimotor training led to greater mirror responses to the observation of newly learned dance sequences than sensory training. One major difference between sensorimotor and sensory training in Cross, Kraemer, et al. was that motor representations in the former were triggered internally by participants themselves who physically performed these movements, whereas motor representations in the latter were triggered externally by movement stimuli. This difference might be the reason why sensorimotor training led to greater mirror responses than sensory training, though both were suggested to involve co-activation of sensory and motor representations of the same movements. Accordingly, in Cross, Kraemer, et al., it was either physical practice or the imitative relationship between observed and physically executed movements during sensorimotor training that further contributed to the formation of the observation-execution link underlying novel movements. Wiggett et al. (2012) observed greater mirror responses to the observation visuomotor-trained novel hand movement sequences than to the observation of motorically trained hand

movement sequences. Thus, their results further suggested that it was probably the imitative relationship between observed and physically executed movements that critically contributed to the formation of the observation-execution link. However, as mentioned earlier, when compared to either sensory or motor training, there was always additional sensory input or motor output during sensorimotor training, which might further facilitate the formation of the observation-execution link. Thus, further investigation is needed to elucidate whether the imitative relationship between observed and physically executed movements plays a critical role in determining the matching properties of the observation-execution link underlying novel actions. Following prior sensorimotor training studies, the current study manipulated the imitative relationship between observed and executed novel actions during training while keeping the amount of sensory input and motor output the same across different groups.

As mentioned earlier, expertise effects and short-term dance training studies by Cross and colleagues provide naturalistic examples showing that responses in the mirror system change as a result of experience. Dance training studies in particular demonstrate examples showing how short-term learning experience contributes to the development of the observation-execution matching link underlying novel movements. Novel movements that consist of familiar elements arranged in a new sequence, such as novel dance sequences, have been employed to investigate the mechanism underlying imitation learning. During imitation learning of novel movement sequences, learners unavoidably need stimuli to remind them of the sequence before they eventually remember it in either a motoric approach or other approaches. In other words, this imitation learning process is also likely to depend on working memory and other executive processes in addition to the perceptuomotor translational process (Catmur et al., 2009). The reminder stimuli used in previous dance training studies have been visual stimuli of a model dancer leading the dance in a studio or videos of a dancer performing the same dance sequences. Thus, motor and sensorimotor experience has always been confounded with each other in these studies. Verbal instructions for how to perform a novel movement sequence might be a solution, but verbal instructions without giving semantic meaning to each sequence component would be long and complicated or otherwise the semantic meaning of each component would be another confounding variable. Therefore, novel movements consisting of familiar elements arranged in a new order do not seem to be ideal candidates for testing the hypothesis that the associative mechanism is involved in imitation of novel movements.

As mentioned earlier, Heyes (2016) has clarified that actions can vary on many dimensions and thus can be defined as novel in various ways. As discussed earlier and further argued below, novel speech actions (i.e., non-native speech actions) may consist of unfamiliar articulator-movement combinations, rather than unfamiliar consequences of familiar movements. It means that reminder stimuli might not be necessary during the learning of novel

speech actions; in other words, learners might be able to practice producing novel speech actions while receiving either imitative or non-imitative movement stimuli. Accordingly, speech actions may provide a way to test whether the associative mechanism is involved in imitation of novel actions. The following section discusses how speech constitutes a distinctive group of movements from both production and perception perspectives and defines novel speech actions used in the current study that aimed to examine whether associative sensorimotor experience modulates the imitative capacity underlying novel actions.

### 5.1.2 *Novel speech actions*

To explore the commonalities between different effector systems, the training study reported in Chapter 4 investigated whether the flexibility observed for the observation-execution link underlying manual gestures extends to the perception-production link underlying orofacial movements. In the training study reported in Chapter 4, speech actions were employed as an orofacial counterpart to manual gestures and thus similarities between these two types of movements were emphasised (i.e., people use both in daily communication and tend to imitate both during perception). However, in addition to belonging to a different effector system, speech actions also constitute a distinctive group of movements from both production and perception perspectives. In this section, I will discuss some distinctive features of speech actions from both production and perception perspectives and potential difficulties in learning to produce and perceive novel speech actions used in a foreign language.

Speech actions consist of sub-movements executed by different muscle groups and require precise temporal control (Borden et al., 1994). Thus, the distinction between two speech actions could result from different articulators performing the same sub-movements or different sub-movements performed by the same articulator. More specifically, the production of two speech sounds may differ in terms of *places of articulation*, i.e., main effectors involved in obstructing or affecting airflow during production; for instance, /ba/ and /da/ are produced respectively by putting lips together or having the tongue tip in contact with the alveolar ridge located just behind the upper teeth to stop and then quickly let air pass through. The production of two sounds may differ in terms of *manners of articulation*, i.e., how articulators interact with each other while air is passing through; for instance, /za/ and /da/ are produced by having the tongue tip in contact with the alveolar ridge to stop and then slowly or quickly let air pass through respectively. The production of two sounds may also differ in terms of *voicing*, i.e., whether the vocal fold vibration is involved during consonant production; for instance, /ba/ and /pa/ are produced by putting lips together and then quickly releasing air simultaneously with the onset of vocal fold vibration or 10 – 35 ms before the onset of vocal fold vibration

respectively. These are three most common dimensions (also called phonetic features in speech science) used to describe the production of a consonant.

More recently, studies have used direct multi-electrode cortical recordings to investigate how articulatory movements are encoded in the human brain (e.g., Bouchard, Mesgarani, Johnson, & Chang, 2013; Chartier, Anumanchipalli, Johnson, & Chang, 2018). Chartier et al., for instance, used direct cortical recordings from sensorimotor cortex while participants spoke natural sentences. Kinematic trajectories of different articulators (e.g., lip, tongue, jaw, larynx) were inferred from participants' acoustic recordings using deep neural networks and were then related to neural recordings from sensorimotor cortex. Each electrode was found to encode a diversity of articulatory kinematic trajectories, indicating coordinated articulator movements toward a specific vocal tract shape. More importantly, broad groupings of electrodes were found to be sensitive to different places of articulation and, within these groupings, manners of articulation and voicing patterns were encoded differently. These findings demonstrated how phonetic features of consonant production are encoded in sensorimotor cortex, with places of articulation being hierarchically dominant.

Behaviourally, Roon & Gafos (2015) employed different syllable pairs for distractors and responses in their speech SRC tasks and observed automatic imitation effects in which perceiving a syllable articulation that shares some phonetic features with the required articulatory response facilitates participants' performance, relative to perceiving a syllable articulation that shares fewer features with the response. In Experiment 1, Roon & Gafos instructed participants to respond to prompts by saying /ka/ or /ga/ in one block and to ignore the auditory distractor /pa/ or /ba/. The two consonants in each pair (e.g., /pa/ and /ba/) share the same place and manner of articulation (bilabial and plosive) but differ in voicing (voiced and voiceless respectively). More importantly, /b/ and /g/ are voiced consonants and hence identical in voicing, and /p/ and /k/ are voiceless consonants and thus also identical in voicing. Accordingly, compatible trials occurred when distractor and response syllables were identical in voicing (e.g., a /ga/ response and a /ba/ distractor), and incompatible trials occurred when the two syllables differed in voicing (e.g., a /ga/ response and a /pa/ distractor). In Experiment 2, Roon & Gafos instructed participants to respond to prompts by saying /pa/ or /ka/ and to ignore the auditory distractor /ba/ or /ga/. Thus, in this experiment, compatible trials occurred when consonants in distractor and response syllables shared the same places of articulation (e.g., a /pa/ response and a /ba/ distractor), and incompatible trials occurred when the two consonants differed in places of articulation (e.g., a /pa/ response and a /ga/ distractor). Automatic imitation effects were observed in both experiments, suggesting that perception-induced motor activation may occur at the feature level (e.g., voicing or place of articulation). Similarly, Ghaffarvand Mokari et al. (2020) observed that participants' performance was

modulated by the degree of similarity between perceived and produced vowels in each trial, also suggesting that people might covertly ‘imitate’ components of observed actions.

More importantly, also in Roon & Gafos (2015), automatic imitation was found to be more robust when consonants in distractor and response syllable pairs share the same places of articulation in Experiment 2 than when the two consonants share the same voicing patterns in Experiment 1. As mentioned above, Chartier et al. (2018) used direct cortical recording to investigate how phonetic features of consonant production are encoded in sensorimotor cortex and found that the place of articulation was hierarchically dominant over other phonetic features. Thus, findings from Roon & Gafos that automatic imitation effects were more robust for the place of articulation than for the voicing pattern was probably because observing a syllable that shares the same place of articulation with the required response led to greater activation in the motor area where the response is encoded than observing a syllable that shares the same voicing pattern. One question thus arises as to whether sensorimotor training with one speech action would generalise to other speech actions that share some similarities, i.e., phonetic features, with the trained action, and the current study will address this issue as well.

Crucially, in order to imitate a novel action, one has to first recognise that action (i.e., know which muscle does what movement) and yet, being able to recognise a novel speech action sometimes could be challenging due to insufficient spatial information. When comparing signed and spoken languages, Gentilucci & Corballis (2006) have summarised that signed languages are mainly communicated through the visuo-manual medium which conveys more efficient information of spatial relations between different body parts and thus allows for a greater degree of parallel transmission; in contrast, with many articulators remaining invisible to perceivers, spoken languages are mainly communicated through the auditory-vocal medium which lacks an effective spatial dimension and relies more on serial transmission. To perform an action essentially means to move one or more body parts in space and thus, spatial information would be most useful when learning to perform a novel action through imitation. Spatial information can be conveyed through verbal instructions or sensory signals, and speech actions are at a disadvantage in terms of providing spatial information through sensory signals. Therefore, perceiving a novel speech action could be challenging sometimes due to insufficient spatial information.

Furthermore, another challenge to being able to recognise novel speech actions could be due to perceivers’ insensitivity to the difference between two speech sounds. Each language has a unique set of speech sounds whose differences are meaningful in that language. For example, “big”, “dig”, and “gig” are different English words and thus the differences across /b/, /d/, and /g/ are meaningful in English. To investigate how people perceive speech sounds, Liberman, Harris, Hoffman, & Griffith (1957) conducted one identification task and one discrimination task with synthetic speech stimuli that varied in acoustically equal steps along

a continuum covering English consonants /b/, /d/, and /g/. In the identification task, Liberman et al. asked participants to identify each received stimulus as /b/, /d/, or /g/. Multiple stimuli next to one another along the continuum were identified as belonging to the same categories, and only few ambiguous ones marked the category boundaries on the continuum. In the discrimination task, participants were better at discriminating sounds from different categories they had identified than at discriminating sounds from the same categories. The concept of *categorical perception* has thus emerged, indicating that speech stimuli along a single continuum are not perceived continuously but rather identified as belonging to discrete units. Importantly, people become worse at discriminating some sound contrasts when the differences are not meaningful in their native languages. The classic example would be native Japanese speakers demonstrating poor discrimination between American /r/ and /l/ because the distinction between these two sounds are not meaningful in Japanese (e.g., Goto, 1971; Miyawaki et al., 1975). Notably, there are also examples where people show good discrimination between non-native speech sounds, e.g., native American English speakers were found to be good at discriminating Zulu click contrasts that do not exist in English (Best, McRoberts, & Sithole, 1988).

Speech actions constitute a distinctive group of movements and prior research has provided a lot of insights into the perception and production of this group of movements. Of the primary interest of the current study, research findings in speech science have offered a systematic way to describe and manipulate action stimuli and suggested potential generalisation of sensorimotor training effects to the imitative capacity underlying actions that share some similarities with the trained actions. One aim of the current study was to investigate whether and how sensorimotor experience modulates the imitative capacity underlying novel actions. To shorten the current study, I selected a pair of non-native speech actions that native British English speakers are able to recognize without perceptual training and gave participants verbal instructions of how to perform these actions. The other aim of the current study was to investigate whether the sensorimotor training effects generalise to the imitative capacity underlying familiar actions that share some similarities with the trained novel actions. I selected a pair of native speech actions whose consonants share the same places of articulation with the selected non-native counterparts. This phonetic feature was chosen because Roon & Gafos (2015) found more robust automatic imitation effects for the place of articulation than for voicing and therefore I expected greater generalisation of training effects in this dimension if sensorimotor training effects do generalise. The following section will introduce the current study in more detail.

### 5.1.3 *The current study*

As mentioned in Section 2.3, the two objectives of the current study were:

- **Objective 4:** to determine whether sensorimotor experience is critically involved in establishing the perception-production matching link underlying novel speech actions.
- **Objective 5:** to explore at what level training-induced modification occurs in the perception-production link.

The two novel non-native speech actions selected in the current study were both consonant-vowel syllables whose consonants do not exist in British English. One action consisted of a voiced bilabial trill (written as [ʙ] in the International Phonetic Alphabet [IPA]) and an /a/, and I will use /Ba/ to represent this sound in the following text; the other action consisted of a voiced alveolar lateral fricative (written as [ɮ] in the IPA) and an /a/, and I will use /La/ to represent this sound in the following text. To shorten the current study, these two non-native speech actions were selected because they are both visually and auditorily salient enough that native British English speakers are able to recognize without perceptual training. In other words, these two non-native speech sounds are perceptually distinct from any British English sounds and thus will not fall into any sound categories in British English. The two native speech actions were /ba/ and /la/. Specifically, /b/ is a voiced bilabial plosive that shares the same place of articulation and voicing pattern with /B/ but has a different manner of articulation; /l/ is a voiced alveolar lateral approximant that shares the same place of articulation and voicing pattern with /L/ but has a different manner of articulation.

In the current study, participants were first given verbal instructions of how to produce each action for both non-native and native speech sounds. Their automatic imitation effects for both non-native and native speech actions were then measured in separate SRC tasks before they received either counter-mirror (i.e., say /Ba/ when the speaker says /La/ and vice versa) or mirror (i.e., say /Ba/ when the speaker says /Ba/ and likewise for /La/) training with non-native speech actions only. In the SRC task used to measure automatic imitation of non-native speech actions, participants were instructed to respond to printed symbol prompts by saying either /Ba/ or /La/ and to ignore the speaker who was simultaneously saying either /Ba/ or /La/ in distractor videos presented audiovisually. Different from the study reported in Chapter 4, audiovisual distractors were selected in the current study to maximise the amount of sensory information from distractors so as to elicit larger automatic imitation effects. In the speech SRC task used to measure automatic imitation of native speech actions, participants had to respond to prompts by saying either /ba/ or /la/ and to ignore the speaker who was

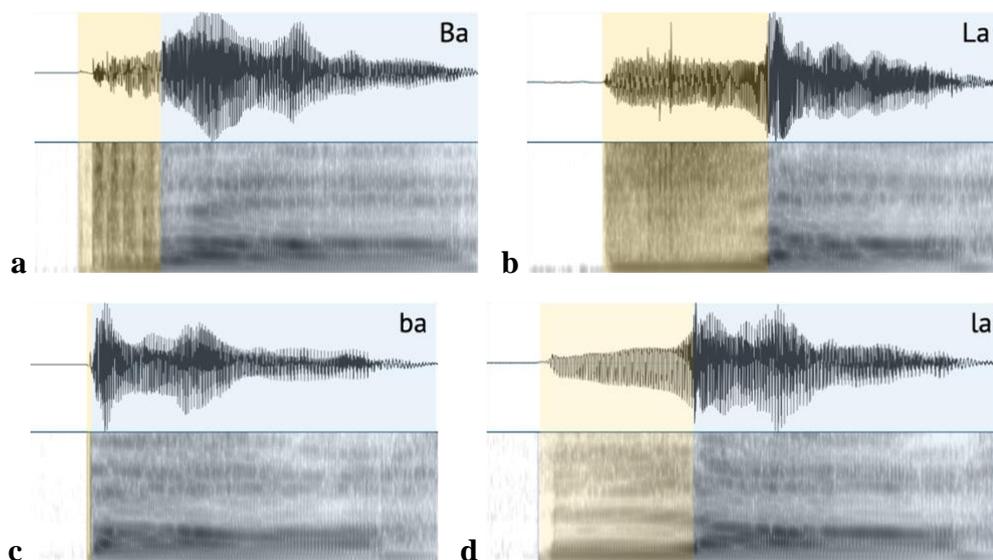
saying either /ba/ or /la/ in distractor videos that were also presented audiovisually. Four symbols employed as prompts were “£”, “%”, “&”, and “#”, and the symbol-syllable pairs were counterbalanced across participants. The SRC task used in the current study was decided based on three pilot studies which were designed to assess the effects of using syllable versus symbol prompts, to ensure that RTs were measured reliably, to determine optimal SOAs, and to evaluate the potential of using simple response tasks. Pilot studies will be reported first in the following section. Participants were assumed to have less sensorimotor experience of concurrently perceiving and producing either non-native speech action, as compared to native speech actions, and thus automatic imitation of non-native speech actions was expected to be smaller than automatic imitation of native speech actions (i.e., a native bias).

Approximately 24 hours after training, automatic imitation effects of both non-native and native speech actions were measured again using the same SRC tasks. Automatic imitation effects were measured before and after training to ensure whether the final results were confounded with any pre-training differences between two groups. For Objective 4, following the associative hypothesis, non-native automatic imitation effects were expected to increase following mirror training and decrease following counter-mirror training. Alternatively, following the genetic hypothesis, non-native automatic imitation effects were expected to increase following either mirror or counter-mirror training. Additionally, for Objective 5, if the modification occurs at the individual action level, non-native automatic imitation effects were expected to increase following mirror training and decrease following counter-mirror training, and native automatic imitation effects were not expected to change following either training. Alternatively, if the modification occurs at the feature level, both non-native and native automatic imitation effects were expected to increase following mirror training and decrease following counter-mirror training. Also, according to findings from studies reported in Chapter 3, automatic imitation of speech varies when prompts are presented at different SOAs and so three SOAs were included to examine whether training effects would interact with automatic imitation at different SOAs.

## **5.2 Pilot study**

The current study was the first to look at automatic imitation of non-native speech actions and compare it with its native counterparts. Three pilot studies were conducted to assess different SRC task designs. Each pilot study tested four participants to obtain a rough picture of whether and how each task design would elicit automatic imitation of both non-native and native speech actions. Participants were first given verbal instructions of how to produce each speech action, native or non-native. The same audiovisual distractor videos of the speaker saying /Ba/, /La/, /ba/, and /la/ were used in all pilot studies. RTs were first measured online using the

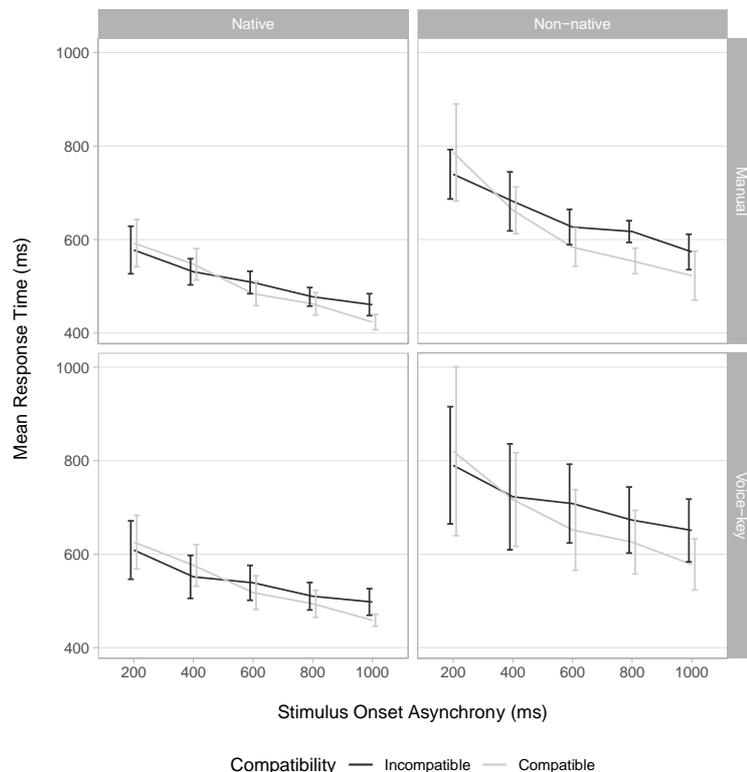
voice-key and then measured manually offline in Praat. When measured using the voice-key, RTs marked the time intervals between prompt onset and the time points when the system detected an audio input at .2 of Presentation's total range. When measured manually in Praat, RTs marked the time intervals between prompt onset and consonant production onset. Each response recording was presented in Praat with its waveform and spectrogram and annotated manually. Waveforms and spectrograms of the speaker's /Ba/, /La/, /ba/, and /la/ recordings were given as examples in Figure 5.1. The onset of /Ba/ production was marked when air started vibrating between lips. The onset of /La/ production was marked when air was slowly released from two sides of the tongue, creating turbulence. The onset of /ba/ production was marked when air was quickly released between lips and immediately followed by vowel articulation. The onset /la/ production was marked when air was slowly released from two sides of the tongue, creating slight turbulence. The four syllables required for responses in the current study are produced with different manners of articulation and particularly, the production of /B/, /L/, and /l/ starts with frication turbulence, which might be difficult for the voice-key to capture consistently. Thus, RTs were measured again manually in Praat to examine the reliability of RTs measured using the voice-key. Detailed descriptions of stimuli and data processing were given in the Method section (Section 5.3). In all pilot studies, incorrect responses, missed trials, anticipatory responses with RTs < 100 ms, and outliers with RTs more than three standard deviations away from corresponding condition means per syllable response were classified as errors and excluded from RT results. Results will be discussed based on figures as the sample size was too small to run formal analyses.



**Figure 5.1** Waveforms and spectrograms of the speaker's recordings of /Ba/ in **a**, /La/ in **b**, /ba/ in **c**, and /la/ in **d**. Yellow rectangles make the consonant part of each syllable, and blue rectangles make the vowel part of each syllable.

### Pilot 1

Pilot 1 tested four participants (4 female,  $M_{age} = 22.00$ ,  $SD_{age} = 2.58$ ,  $range_{age} = 19-25$ ), and each participant completed one SRC task for non-native speech and one SRC task for native speech. In the SRC task used to measure automatic imitation of non-native speech actions, participants were instructed to respond to the printed syllable *Ba* or *La* by saying either /Ba/ or /La/ respectively and to ignore the speaker who was saying either /Ba/ or /La/ in a distractor video. In the SRC task used to measure automatic imitation of native speech actions, participants were instructed to respond to the printed syllable *ba* or *la* by saying either /ba/ or /la/ respectively and to ignore the speaker who was saying either /ba/ or /la/ in a distractor video. As discussed in Section 4.1.3, syllable prompts, instead of symbol prompts, were selected to avoid learning effects in which participants become familiar with arbitrarily assigned symbol-syllable pairs (e.g., say /ba/ when seeing the prompt “#”), which would then bring systematic change in automatic imitation from pre-training to post-training sessions in both groups. The order of non-native and native SRC tasks was counterbalanced across participants. Five SOAs were employed to locate larger automatic imitation effects, and they were 200, 400, 600, 800, and 1,000 ms.



**Figure 5.2** Mean RTs  $\pm$  SEs in each experimental condition. Four panels represent the native (top-left) and non-native (top-right) sessions with manually measured response times and the native (bottom-left) and non-native (bottom-right) sessions with response times measured using the voice-key.

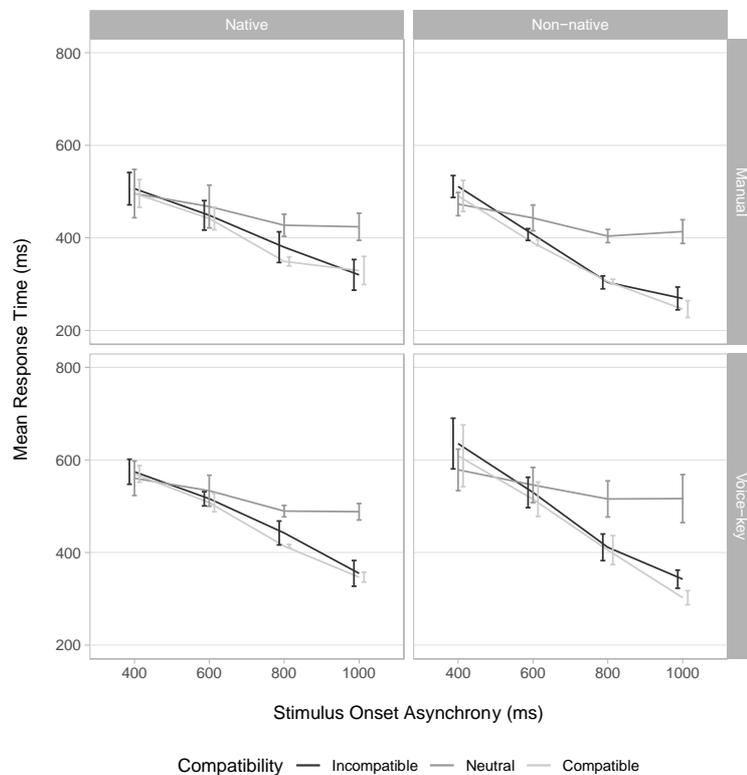
Based on the observation of Figure 5.2, RTs appear to be slower for non-native speech actions than for native speech actions, suggesting that the former were probably harder for participants to produce. However, some participants reported that they tended to associate the syllable prompts *Ba* and *La* with the native speech actions /ba/ and /la/ first before linking the prompts to the corresponding non-native speech actions /Ba/ and /La/. Accordingly, relatively slower RTs in the non-native session could result from the initial association between syllable prompts and native speech actions and/or participants' relative inexperience in producing non-native speech actions. The two causes of delay in RTs in the non-native session were difficult to dissociate, and thus using syllable prompts seemed to be problematic because they would bring systematic change in RTs in the non-native session but not in the native session.

Moreover, there seem to be automatic imitation effects for both native and non-native speech actions and interestingly the latter appears to be larger. Automatic imitation effects were expected to be smaller for non-native speech than for native speech (i.e., a native bias) due to less sensorimotor experience with the non-native actions. However, these preliminary results suggested the opposite pattern, i.e., a reversed native bias, or a non-native bias. I will discuss this finding in detail in Section 5.5.1 below. Furthermore, automatic imitation effects do not seem to occur at the two shortest SOAs. Lastly, voice-key measured RTs seem to be slower and have larger variances than manually measured RTs, particularly for non-native responses. As mentioned earlier, four syllable responses used in the current study started with consonants that are produced with different manners of articulation. Particularly, the production of /B/, /L/, and /l/ starts with frication turbulence (see Figure 5.1), which might be difficult for the voice-key to capture consistently and thus undermine the reliability of voice-key measured RTs. These findings confirmed my speculations, and manually measured RTs will be used in the current study.

## ***Pilot 2***

Pilot 2 tested four participants (1 female,  $M_{age} = 28.00$ ,  $SD_{age} = 3.16$ ,  $range_{age} = 24-31$ ). Based on findings from Pilot 1, a simple response task was employed where responses were pre-specified before each block and thus no prompt was needed. The SRC task used in Heyes et al. (2005) was adapted in this pilot study. Each participant completed four blocks of SRC tasks. The same response was required in each block and was pre-specified before the block started. When the pre-specified response was /Ba/, participants were instructed to respond to a fixation cross by saying /Ba/ and to ignore the speaker's /Ba/ or /La/ articulation in distractor videos. When the pre-specified response was /La/, participants were instructed to respond to the same fixation cross by saying /La/ and to ignore the speaker's /Ba/ or /La/ articulation. Data were collapsed across these two non-native blocks. Compatible trials occurred when participants

and the speaker were saying the same syllables, and incompatible trials occurred when participants and the speaker were saying different syllables. Similarly, when the pre-specified response was /ba/ (or /la/ in another block), participants were instructed to respond to the same fixation cross by saying /ba/ (or /la/ in another block) and to ignore the speaker's /ba/ or /la/ articulation in distractor videos. In addition, a neutral condition was included with the speaker remaining static throughout the distractor video. Since results from Pilot 1 suggested a larger automatic imitation effect for non-native speech actions than for native speech actions, the current pilot study included a neutral condition to examine whether the larger effect for non-native speech actions was due to more facilitation in compatible trials or more interference in incompatible trials, as compared to the effect for native speech actions. The four longest SOAs (400, 600, 800, and 1,000 ms) used in Pilot 1 were employed in this pilot study.



**Figure 5.3** Mean RTs  $\pm$  SEs in each experimental condition. Four panels represent the native (top-left) and non-native (top-right) sessions with manually measured response times and the native (bottom-left) and non-native (bottom-right) sessions with response times measured using the voice-key.

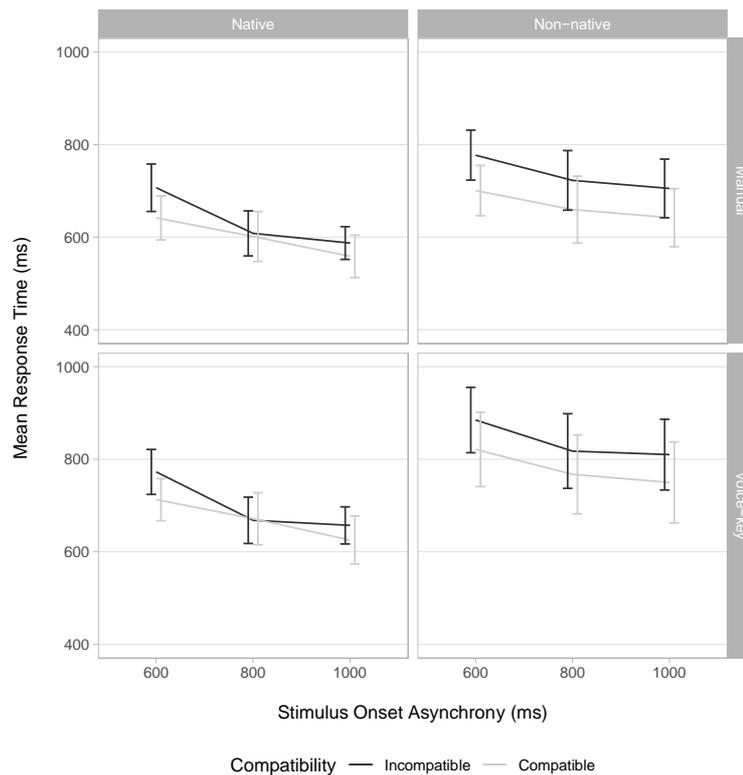
Based on the observation of Figure 5.3, RTs appear to be similar for native and non-native speech actions, suggesting that slower responses for non-native speech actions observed in Pilot 1 was probably due to the association between syllable prompts and native speech actions. However, the simple response task used in this pilot study also gave participants more time to prepare for a response, e.g., moving articulators to the initial position,

before they saw the fixation cross and this preparation time was not included in RTs. Accordingly, simple response tasks were not suitable for measuring automatic imitation of native and non-native speech actions in the current because the time spent in action preparation was missing in RT data.

Automatic imitation effects for both native and non-native speech actions seem to be small, as compared to the effects observed in Pilot 1. It was probably also because participants in the simple response task were better prepared for a response and thus were more resilient to the influence of distractor videos. Most interestingly, RTs are slower in the neutral condition than in either the compatible or the incompatible condition. It was probably because participants became more alert when seeing dynamic stimuli and the speaker's compatible /a/ articulation might have facilitated all their responses that also included /a/ in the end. Therefore, a neutral condition in speech SRC tasks did not seem to be informative in terms of elucidating whether the larger automatic imitation effect for non-native speech actions than for native speech actions in Pilot 1 was due to more facilitation in compatible trials or more interference in incompatible trials.

### ***Pilot 3***

Pilot 3 tested four participants (3 female,  $M_{age} = 26.50$ ,  $SD_{age} = 5.26$ ,  $range_{age} = 22-34$ ). Based on findings from Pilot 1 and 2, it seemed that using syllable prompts would bring systematic change between the native and the non-native sessions and that using simple response tasks with no prompts presented would omit important information in RT data. Therefore, symbol prompts £, %, &, and # were used in this pilot study, and different symbol-syllable pairs (e.g., when seeing £, say /Ba/) were assigned to each participant. In the SRC task used to measure automatic imitation of non-native speech, participants were instructed to respond to the assigned symbol prompt by saying either /Ba/ or /La/ and to ignore the speaker's /Ba/ or /La/ articulation in distractor videos. Similarly, in the task used to measure automatic imitation of native speech, participants were instructed to respond to the assigned symbol prompt by saying either /ba/ or /la/ and to ignore the speaker's /ba/ or /la/ articulation. The three longest SOAs (600, 800, and 1,000 ms) used Pilot 1 were adopted in this pilot study.



**Figure 5.4** Mean RTs  $\pm$  SEs in each experimental condition. Four panels represent the native (top-left) and non-native (top-right) sessions with manually measured response times and the native (bottom-left) and non-native (bottom-right) sessions with response times measured using the voice-key.

Based on the observation of Figure 5.4, RTs appear to be slower for non-native speech actions than for native speech actions. Since symbol-syllable pairs were arbitrary for both non-native and native speech, the finding of slower RTs for non-native speech suggested that non-native speech actions were harder for participants to produce, or say, participants needed longer time to prepare for a non-native response. Similar to findings from Pilot 1, the automatic imitation effect appears to be larger for non-native speech actions than for native speech actions, showing the opposite to what was expected. I will further discuss this finding in Section 5.5.1 below.

### **Discussion**

According to findings from the above three pilot studies, the current study adopted the SRC task used in Pilot 3 where participants had to respond to symbol prompts. The increased familiarity with arbitrarily assigned symbol-syllable pairs would still be a concern, but more practice trials would be given before each pre-training session to mitigate the familiarity effect. Additionally, the same SOAs (600, 800, and 1,000 ms) used in Pilot 3 were also adopted in the current study for they seemed to elicit relatively large automatic imitation effects for both native and non-native speech actions. Furthermore, in the current study, RTs from testing

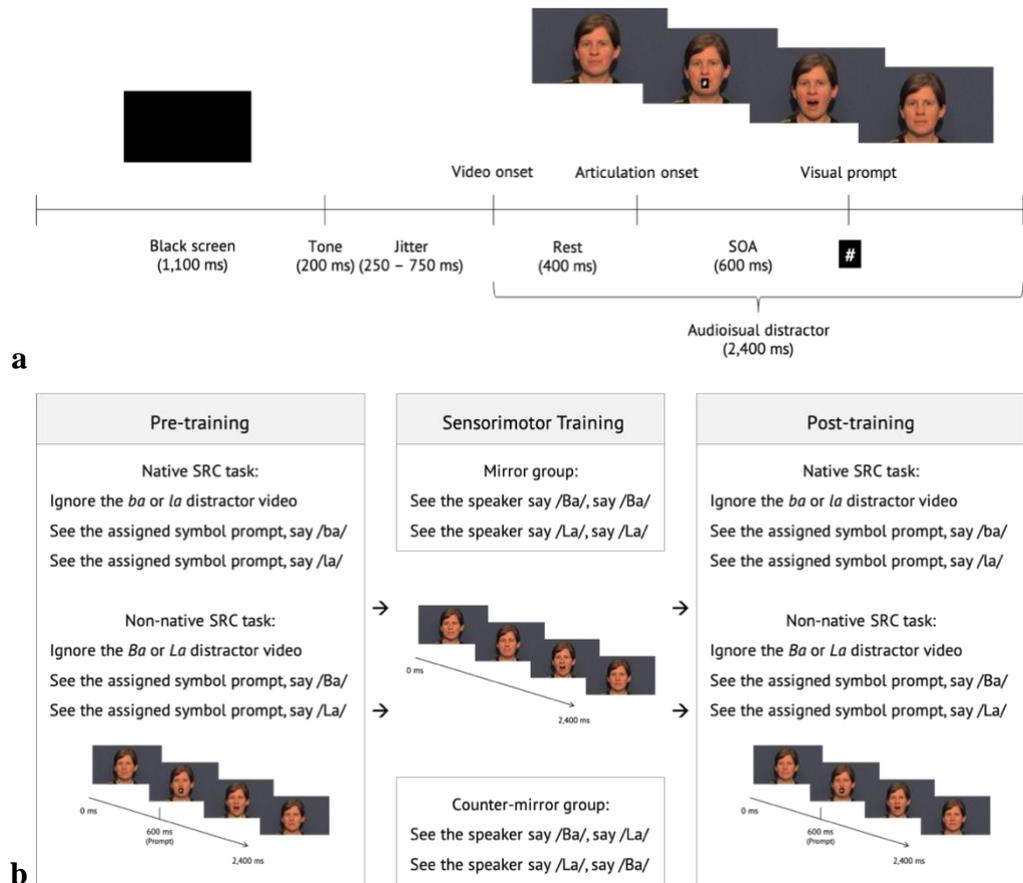
sessions would be measured manually to obtain more reliable data. However, since it was more important to ascertain that participants were doing the correct training, RTs from training sessions were measured using the voice-key.

### **5.3 Method**

#### ***Participants***

An a-priori power analysis was conducted using G\*Power 3.1 (Faul et al., 2009) with an effect size  $\eta_p^2 = .145$ , obtained from the study reported in Chapter 4. Sixty-eight participants were needed to detect significant sensorimotor training effects on automatic imitation with a power of .80 and an  $\alpha$  of .05. Before the university closure due to the COVID-19 pandemic, 53 participants were recruited, but one was excluded for saying /ba/ instead of /Ba/ in the non-native session, one for not finishing the post-training session, and three for performing at chance level in testing sessions. Another participant was excluded for her/his error rate was more than four standard deviations away from the group mean. The final analysis included 24 participants in the mirror group (17 female,  $M_{age} = 23.17$ ,  $SD_{age} = 4.14$ ,  $range_{age} = 18-33$ ) and 23 in the counter-mirror group (17 female,  $M_{age} = 22.78$ ,  $SD_{age} = 4.21$ ,  $range_{age} = 18-35$ ). All were native British English speakers with self-reported normal or corrected-to-normal vision, normal hearing and no speech or language disorders, or other neurological disorders. Participants received £20 or course credit. The University Research Ethics Committee approved the procedures and all participants gave written informed consent.

## Material and stimuli



**Figure 5.5 a** Schematic timeline of a trial presented in either the pre-training or post-training session. **b** The experimental procedure with pre-training, training, and post-training sessions. Participants were randomly assigned to either mirror or counter-mirror group. Both pre- and post-training sessions consist of one native and one non-native SRC task, and the order of two tasks was counterbalanced across participants.

Stimuli (see Figure 5.5) included audiovisual instruction videos of a speaker introducing how to produce /Ba/, /La/, /ba/, and /la/, audiovisual distractor videos of the same speaker saying /Ba/, /La/, /ba/, and /la/, and printed symbol prompts £, %, &, and #. Videos (25 frame per second) were filmed with a Canon Legria video camera, cut and edited in iMovie on an Apple iMac, and stored in MP4 format with a resolution of 1,920 x 1,080. A female native British English speaker was shown in the videos from her neckline upward in colour on a blue background, and her nose was approximately centred on the screen. Within each instruction video, the speaker first produced the target speech action once, then introduced how to produce this action step by step and produced the same action twice in the end. Verbal instructions for each speech action are given below.

- After the speaker pronounced the non-native sound /Ba/ once, she stated “*To produce this sound, bring your lips together to block airflow, release the air slowly, letting it pass between your lips, as if to blow raspberries, and say /a/*”.

- After the speaker pronounced the non-native sound /La/ once, she stated “*To produce this sound, move your lips apart slightly, place the tip of your tongue behind your upper teeth to block airflow, raise the sides of your tongue, release the air slowly, causing turbulence and letting it pass by the sides of your tongue, and say /a/*”.
- After the speaker pronounced the native sound /ba/ once, she stated “*To produce this sound, bring your lips together to block airflow, let the air out in one burst, and say /a/*”.
- After the speaker pronounced the native sound /la/ once, she stated “*To produce this sound, move your lips apart slightly, place the tip of your tongue behind your upper teeth to block airflow, release the air slowly, letting it pass by the sides of your tongue, and say /a/*”.

Distractor videos were further edited using MATLAB and then stored in AVI format. Each distractor video lasted 2,400 ms (i.e., 60 frames), starting and ending with the speaker’s mouth closed in a resting configuration. Prior to articulation onset, the speaker remained in a resting configuration for 400 ms. In the *Ba* video, the consonant production lasted 1,200 ms, the /a/ articulation commenced at 1,600 ms, the vertical extent of the mouth configuration reached its maximum at 1,800 ms, and the mouth began closing at 2,120 ms. In the *La* video, the consonant production lasted 1,160 ms, the /a/ articulation commenced at 1,560 ms, the mouth’s vertical extent reached its maximum at 1,760 ms, and the mouth began closing at 2,040 ms. In the *ba* video, the consonant production lasted 1,120 ms, the /a/ articulation commenced at 1,520 ms, the mouth’s vertical extent reached its maximum at 1,680 ms, and the mouth began closing at 1,960 ms. In the *la* video, the consonant production lasted 1,040 ms, the /a/ articulation commenced at 1,440 ms, the mouth’s vertical extent reached its maximum at 1,600 ms, and the mouth began closing at 1,920 ms. The duration of each event (i.e., the speaker’s static face, consonant, and vowel) in each video is given in Table 5.1. Compared to syllable articulations occurring under a natural circumstance, the speaker’s articulation for each syllable was exaggerated in distractor stimuli to better convey the articulatory information of each syllable. Participants’ viewing was unrestrained at a distance of 60 cm from the screen, and the speaker’s face was shown in 22.16° x 16.50° of visual angle. In all videos, the speaker’s mouth was shown in 1.78° x 5.24° of visual angle at its resting configuration and in 4.17° x 4.63° during the vowel articulation.

Spoken syllables were recorded using a RØDE NT1-A Condenser Microphone and a Focusrite Scarlett 2i4 USB Computer Audio Interface pre-amplifier plugged into the sound card input of a Dell PC at 44.1kHz with 16 bits. Syllable recordings were then edited and root-mean-square normalised using Praat. The corresponding video and auditory recording were

presented simultaneously in each trial. Thus, to align auditory recordings with videos when presented simultaneously in Presentation, 1,470-ms silence was added before the speaker's /Ba/ recording which itself lasted 761 ms, 1,135-ms salience was added before /La/ that lasted 1,035 ms, 1,520-ms silence was added before /ba/ that lasted 687 ms, and 1,178-ms silence was added to /la/ that lasted 806 ms (see Table 5.1). Auditory stimuli were delivered to participants through Beyerdynamic DT 297 PV MK II / 80 ohms headset at a comfortable level.

**Table 5.1** Consonant and vowel durations in ms of the speaker's articulation of each syllable in both video (V) and audio (A) recordings.

Syllable	Static face	Consonant	Vowel			Total
			Mouth opening	Full opening	Mouth closing	
V /Ba/	400	1,200	200	320	280	800
/La/	400	1,160	200	280	360	840
/ba/	400	1,120	160	280	440	880
/la/	400	1,040	160	320	480	960
Syllable	Silence	Consonant	Vowel			
A /Ba/	1,470	132				629
/La/	1,135	433				602
/ba/	1,520	6				681
/la/	1,178	262				544

Printed symbol prompts £, %, &, and # were presented in white Helvetica font and in size 36 in Presentation. With participants' viewing unrestrained at a distance of 60 cm from the screen, the prompts were shown in 0.95° x 1.15° of visual angle. To make participants focus on the speaker's lip area, the prompts were positioned extending from the speaker's bottom lip to her top lip. The prompt was presented at one of three SOAs (600, 800, or 1,000 ms relative to articulation onset; see Figure 5.5) in each trial. SOAs were equally spaced and all occurred during the consonant production in each distractor video. The experiment was performed using Presentation.

### ***Design and procedure***

The experiment included two testing sessions (pre- versus post-training) and one training session (see Figure 5.5) and took place in a soundproofed, light-controlled booth. Within each testing session, participants completed two SRC tasks, one with non-native speech actions and the other with native speech actions. For the non-native SRC task, written instructions were presented on PowerPoint slides. Participants first watched two instruction videos of the speaker introducing how to produce the two non-native speech actions. Participants could choose to watch the videos as many times as they wanted and were asked to produce the action

at least five times after watching each video. SRC task instructions were then presented on a following slide. To avoid naming either non-native speech action, task instructions were given in a combination of text and videos. For example, a sentence saying *Please speak out the following syllable, when you see the symbol “£”* was presented on the left side of the screen with the video of the speaker saying /Ba/ placed underneath, and a sentence saying *Please speak out the following syllable, when you see the symbol “#”* was simultaneously presented on the right side of the screen with the video of the speaker saying /La/ placed underneath. The SRC experiment was then performed using Presentation, and instructions were presented again asking participants to respond to symbol prompts by producing the assigned non-native speech actions and to ignore the distractor videos of the speaker saying /Ba/ or /La/. The procedure for the native SRC task was identical to that for the non-native SRC task, except that participants watched instruction videos of how to produce /ba/ and /la/ and were asked to produce /ba/ and /la/ in response to the remaining two symbol prompts (% and &) and that the distractor videos of the speaker saying /ba/ or /la/ were presented. Participants were randomly assigned with either the native or the non-native SRC task first. There were 24 combinations of symbol-syllable pairs, and participants were randomly assigned with one combination.

Each trial started with a black screen for 1,100 ms, which was then followed by a 200-ms tone with a frequency of 500 Hz at 70 dB of sound pressure level played through Beyerdynamic DT 297 PV MK II / 80 ohms headset. The screen then remained black for one of five jittered durations (250, 375, 500, 652, or 750 ms) that were included to reduce the trial onset’s predictability. A distractor video was then presented. The speaker’s articulation started at 400 ms after the video onset and a printed symbol prompt was presented at one of three SOAs after the articulation onset. The screen turned black at the end of the video to signal the next trial onset. Each SRC task session included six blocks with 30 trials each (180 trials in total). Thus, 60 trial types (2 Prompts x 2 Distractors x 3 SOAs x 5 Jitters) were repeated three times and presented in a randomised order. Twenty practice trials were randomly chosen and given before the first block, and participants could take a break between blocks. Each SRC task session lasted about 20 min, and the whole pre-training session with one native and one non-native SRC tasks lasted about 40 min. The post-training session was identical to the pre-training session, except that the post-training session took place a day later, no instruction videos were presented, and only five practice trials were given.

Training took place immediately after participants finished both native and non-native SRC tasks, and participants were then randomly assigned to one of the two training groups. Participants in the counter-mirror group had to say /Ba/ as soon as they saw the speaker saying /La/ and vice versa, and participants in the mirror group had to say /Ba/ as soon as they saw the speaker saying /Ba/ and likewise for /La/. No native speech action was required to produce or was presented. No prompt was presented during training, and participants were only

responding to the speaker. Each trial started with a black screen for 1,100 ms, which was then followed by the same 200-ms tone used in testing sessions. The screen remained black for one of five jittered durations (250, 375, 500, 652, or 750 ms) and was then followed by a video of the speaker saying either /Ba/ or /La/. There were 20 blocks with 30 trials each (i.e., 600 trials in total) in each training session, and 10 trial types (2 Videos x 5 Jitters) were repeated in a randomised order. Five practice trials were randomly chosen and given before the first block, and participants could choose to take a break between blocks. Each training session lasted about one hour. In total, the whole experience lasted about two hours on the first day and 40 min on the second day.

### ***Data recording and analysis***

Responses were recorded using the Beyerdynamic DT 297 PV MK II / 80 ohms headset microphone and a Focusrite Scarlett 2i4 USB Computer Audio Interface pre-amplifier plugged into the sound card input of a Dell PC at 44.1 kHz with 16 bits. Audio recording started at the video onset and lasted for 3,000 ms. The voice-key was triggered when the system detected an audio input at .2 of Presentation's total range. RTs were measured relative to prompt onset in testing sessions and were measured relative to video onset in training sessions. If no response was detected, a warning saying *NO RESPONSE* would be presented for 500 ms. In addition, a warning saying *RESPONSE TOO EARLY* would appear if participants responded within 50 ms of the prompt onset, and a warning saying *RESPONSE TOO LATE* would appear if participants responded after 1,500 ms of the prompt onset.

As discussed in the Pilot study section (Section 5.2), manually measured RTs were more reliable than voice-key measured RTs. Therefore, for further analysis, RTs from testing sessions were measured again manually using Praat and were defined as time intervals between prompt onset and consonant production onset (see Section 5.2 for detailed descriptions). Voice-key measured RTs from training sessions were used in statistical analysis for it was more important to ascertain that participants were doing the correct training and RTs were less of interest in the current study. For both testing and training sessions, errors included incorrect responses, missed trials, anticipatory responses with RTs < 100 ms, and outliers with RTs more than three standard deviations away from corresponding condition means per syllable response, and were excluded from the RT analysis. For non-native responses, participants were told to try their best to produce the turbulence for either non-native speech action. Hence, whether a non-native response was correct depended on whether there was turbulence created at the articulation onset. A natural log-transformation was applied to RTs for statistical analyses, but figures present back-transformed RTs.

ERs and RTs collected from testing sessions were subjected to separate repeated-measures ANOVA with test (pre- versus post-training), nativeness (native versus non-native), compatibility (compatible versus incompatible), and SOA (600, 800, or 1,000 ms, also referred to as SOA1, 2, and 3) as within-subjects variables and with training (mirror versus counter-mirror) as a between-subjects variable. Expected statistical results for each prediction for Objective 4 and 5 are listed below:

- **Objective 4:**
- **Prediction 11:** following the associative hypothesis, non-native automatic imitation effects increase following mirror training and decrease following counter-mirror training – the interaction between training, test, and compatibility should be significant for non-native actions.
- **Prediction 12:** following the genetic hypothesis, non-native automatic imitation effects increase following either mirror or counter-mirror training – the interaction between training, test, and compatibility should not be significant for non-native actions, but the interaction between test and compatibility was significant for non-native actions.
- **Objective 5:**
- **Prediction 13:** if the modification occurs at the individual action level, non-native automatic imitation effects increase following mirror training and decrease following counter-mirror training, and native automatic imitation effects do not change following either training – the interaction between training, test, nativeness, and compatibility should be significant.
- **Prediction 14:** if the modification occurs at the feature level, both non-native and native automatic imitation effects increase following mirror training and decrease following counter-mirror training – the interaction between training, test, nativeness, and compatibility should not be significant, but the interaction between training, test, and compatibility should be significant.

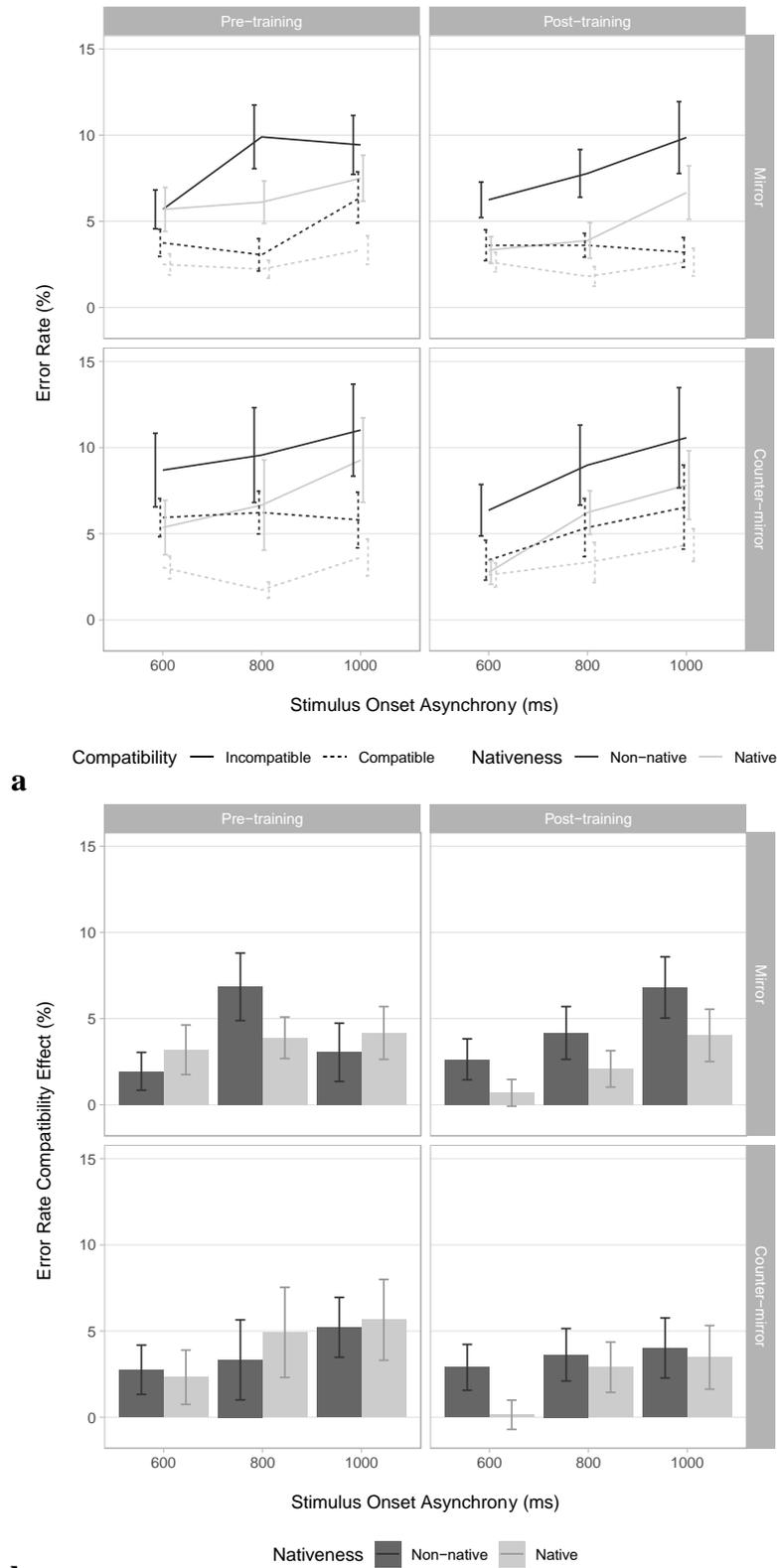
Both ERs and RTs collected from training sessions were divided into four blocks and then subjected to separate repeated-measures ANOVA with block (4 blocks) as a within-subjects variable and with training as a between-subjects variable. The significance level was set to  $p < .05$ . Greenhouse-Geisser correction for non-sphericity and Bonferroni-correction for multiple comparisons were applied whenever appropriate.

When selecting consonants, the current study had controlled for places of articulation between non-native and native pairs, but not for manners of articulation. In fact, all four consonants are produced with different manners of articulation. When using different pairs of responses in separated blocks, Galantucci et al. (2009) only observed automatic imitation effects for the voiced response pair /ba/ and /da/ but not for either the voiceless pair /pa/ and /ta/ or the nasal pair /ma/ and /na/. Their findings suggested that voicing patterns and manners of articulation of response syllables might play a role in modulating automatic imitation. To further examine whether manners of articulation also affected automatic imitation effects in the current study, ERs and RTs collected from testing sessions were also subjected to separated repeated measures ANOVA with test, nativeness, compatibility, and syllable (ba, la, Ba, or La) as within-subjects variables and with training as a between-subjects variable. The significance level was set to  $p < .05$ . Greenhouse-Geisser correction for non-sphericity and Bonferroni-correction for multiple comparisons were applied whenever appropriate.

## 5.4 Results

### 5.4.1 Testing sessions

#### Error rate analysis



**Figure 5.6 a** ERs  $\pm$  SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training (bottom-right) sessions in the counter-mirror group. **b** ER compatibility effects  $\pm$  SEs at each SOA in each testing session for each training group.

**Table 5.2** Five-way ANOVA summary for ERs, as a function of training, test, nativeness, compatibility, and SOA.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	Epsilon	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Training	1	45	1.00	297.18	18671.79	0.72	.402	.016
Test	1	45	1.00	176.51	2343.30	3.39	.072	.070
Test x Training	1	45	1.00	1.76	2343.30	0.03	.855	.001
Nativeness	1	45	1.00	1523.41	7807.25	8.78	.005	.163
Nativeness x Training	1	45	1.00	28.91	7807.25	0.17	.685	.004
Compatibility	1	45	1.00	3516.07	5256.26	30.10	< .001	.401
Compatibility x Training	1	45	1.00	2.44	5256.26	0.02	.886	.000
SOA	1.39	62.41	0.69	970.84	3847.96	11.35	< .001	.201
SOA x Training	1.39	62.41	0.69	26.23	3847.96	0.31	.656	.007
Test x Nativeness	1	45	1.00	0.49	2102.28	0.01	.919	.000
Test x Nativeness x Training	1	45	1.00	15.71	2102.28	0.34	.565	.007
Test x Compatibility	1	45	1.00	47.05	1322.67	1.60	.212	.034
Test x Compatibility x Training	1	45	1.00	9.80	1322.67	0.33	.567	.007
Nativeness x Compatibility	1	45	1.00	47.12	1527.26	1.39	.245	.030
Nativeness x Compatibility x Training	1	45	1.00	12.05	1527.26	0.36	.554	.008
Test x Nativeness x Compatibility	1	45	1.00	68.57	1895.26	1.63	.209	.035
Test x Nativeness x Compatibility x Training	1	45	1.00	0.18	1895.26	0.00	.948	.000
Test x SOA	2	90	1.00	24.35	2265.27	0.48	.618	.011
Test x SOA x Training	2	90	1.00	97.98	2265.27	1.95	.149	.041
Nativeness x SOA	2	90	0.99	35.29	1439.10	1.10	.336	.024
Nativeness x SOA x Training	2	90	0.99	21.75	1439.10	0.68	.509	.015
Test x Nativeness x SOA	1.71	76.79	0.85	4.29	1530.68	0.13	.850	.003
Test x Nativeness x SOA x Training	1.71	76.79	0.85	41.07	1530.68	1.21	.300	.026
Compatibility x SOA	2	90	0.97	315.57	1378.40	10.30	< .001	.186
Compatibility x SOA x Training	2	90	0.97	5.02	1378.40	0.16	.849	.004
Test x Compatibility x SOA	2	90	0.97	31.50	1980.75	0.72	.492	.016
Test x Compatibility x SOA x Training	2	90	0.97	73.19	1980.75	1.66	.195	.036
Nativeness x Compatibility x SOA	2	90	0.92	5.09	1882.97	0.12	.886	.003
Nativeness x Compatibility x SOA x Training	2	90	0.92	51.86	1882.97	1.24	.294	.027
Test x Nativeness x Compatibility x SOA	2	90	0.82	14.06	1352.97	0.47	.628	.010
Test x Nativeness x Compatibility x SOA x Training	2	90	0.82	28.47	1352.97	0.95	.392	.021

Note.  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.

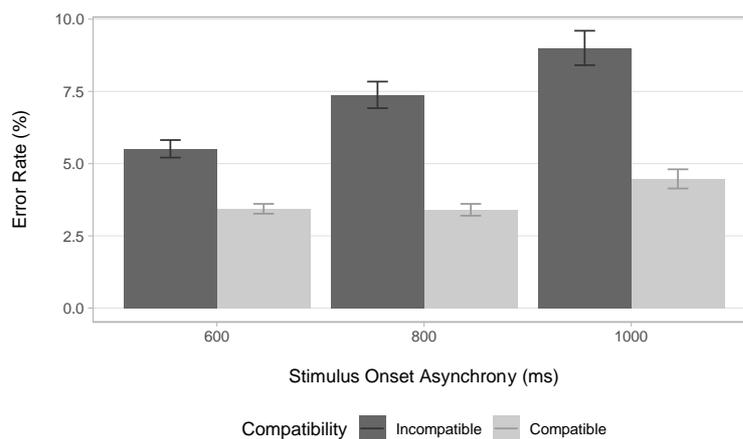
$SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

On average, participants made 5.6% errors (incorrect and missed responses: 5.3%; anticipatory responses: 0.1%; outliers: 0.2%). ER analyses (see Figure 5.6, Table 5.2, and Table 5.3) revealed a main effect of nativeness, with a higher ER for non-native responses than for native responses. The ER compatibility effect revealed a higher ER for incompatible trials than for compatible trials ( $\Delta = 3.53$ ). Follow-up  $t$  tests for the main effect of SOA revealed higher ERs for longer SOAs (all  $p < .01$ ).

**Table 5.3** ERs, SEs, and 95% CIs for each level within nativeness, compatibility, and SOA.

	Level	Mean	SE	95% CI
Nativeness	Non-native	6.71	0.91	[4.88, 8.53]
	Native	4.38	0.47	[3.43, 5.33]
Compatibility	Incompatible	7.31	0.88	[5.54, 9.08]
	Compatible	3.78	0.41	[2.94, 4.61]
SOA (ms)	600	4.48	0.41	[3.65, 5.32]
	800	5.41	0.61	[4.17, 6.64]
	1000	6.74	0.89	[4.96, 8.53]

The two-way interaction between compatibility and SOA (see Figure 5.7 and Table 5.4) was also significant, and follow-up  $t$  tests revealed significant ER compatibility effects at all SOAs (all  $p < .001$ ). Further comparisons of ER compatibility effects across SOAs revealed a smaller ER compatibility effect at SOA1 than at SOA2-3 (both  $p < .003$ ).



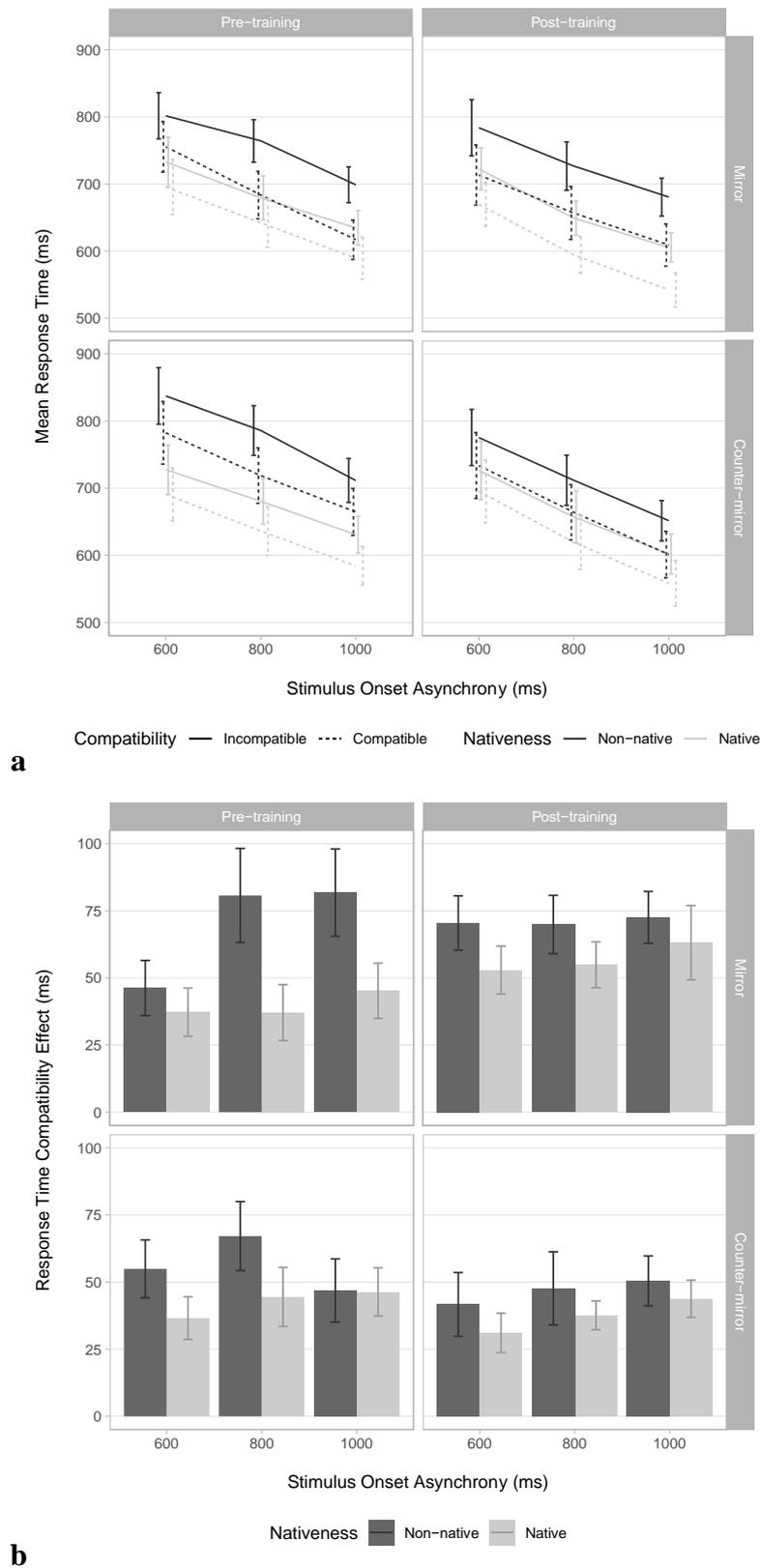
**Figure 5.7** ERs  $\pm$  SEs at each SOA for each compatibility condition.

**Table 5.4** ERs, SEs, and 95% CIs for each level within the interaction between compatibility and SOA.

Compatibility	SOA (ms)	Mean	SE	95% CI
Incompatible	600	5.52	0.62	[4.28, 6.76]
	800	9.02	1.21	[6.57, 11.46]
	1,000	3.42	0.40	[2.61, 4.23]
Compatible	600	7.39	0.94	[5.51, 9.28]
	800	3.45	0.34	[2.76, 4.13]
	1,000	4.47	0.67	[3.13, 5.81]

To further examine Prediction 11 and 12 given in Section 5.3, a four-way repeated-measures ANOVA between training, test, compatibility, and SOA was conducted for non-native actions. The ER compatibility effect ( $\Delta = 3.94$ ) was significant,  $F(1, 45) = 26.59$ ,  $p < .001$ ,  $\eta^2_p = .371$ . The main effect of SOA was significant,  $F(1, 45) = 7.76$ ,  $p = .003$ ,  $\eta^2_p = .147$ , and follow-up  $t$  tests revealed a smaller ER at SOA1 than at SOA2-3 (both  $p < .04$ ). The two-way interaction between compatibility and SOA was significant,  $F(2, 90) = 3.50$ ,  $p = .034$ ,  $\eta^2_p = .072$ , and follow-up  $t$  tests revealed a smaller ER compatibility effect at SOA1 than at SOA2-3 (both  $p < .007$ ).

**Response time analysis**



**Figure 5.8 a** Mean RTs  $\pm$  SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training

(bottom-right) sessions in the counter-mirror group. **b** RT compatibility effects  $\pm$  SEs for at each SOA in each testing session for each training group.

**Table 5.5** Five-way ANOVA summary for log-transformed RTs, as a function of training, test, nativeness, compatibility, and SOA.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	<i>Epsilon</i>	$SS_{Num}$	$SS_{Den}$	<i>F</i>	<i>p</i>	$\eta^2_p$
Training	1	45	1.00	0.02	45.08	0.02	.894	0.000
Test	1	45	1.00	0.78	3.33	10.50	.002	0.189
Test x Training	1	45	1.00	0.02	3.33	0.28	.602	0.006
Nativeness	1	45	1.00	2.43	4.82	22.65	< .001	0.335
Nativeness x Training	1	45	1.00	0.01	4.82	0.10	.749	0.002
Compatibility	1	45	1.00	2.09	0.80	117.88	< .001	0.724
Compatibility x Training	1	45	1.00	0.03	0.80	1.88	.177	0.040
SOA	1.18	52.97	0.59	5.26	0.92	256.09	< .001	0.851
SOA x Training	1.18	52.97	0.59	0.01	0.92	0.31	.615	0.007
Test x Nativeness	1	45	1.00	0.06	1.27	2.03	.162	0.043
Test x Nativeness x Training	1	45	1.00	0.06	1.27	2.10	.155	0.045
Test x Compatibility	1	45	1.00	0.00	0.11	1.49	.229	0.032
Test x Compatibility x Training	1	45	1.00	0.01	0.11	3.23	.079	0.067
Nativeness x Compatibility	1	45	1.00	0.03	0.23	5.05	.030	0.101
Nativeness x Compatibility x Training	1	45	1.00	0.00	0.23	0.66	.420	0.014
Test x Nativeness x Compatibility	1	45	1.00	0.00	0.17	0.55	.464	0.012
Test x Nativeness x Compatibility x Training	1	45	1.00	0.00	0.17	0.57	.453	0.013
Test x SOA	2	90	0.95	0.03	0.20	6.60	.002	0.128
Test x SOA x Training	2	90	0.95	0.01	0.20	1.87	.161	0.040
Nativeness x SOA	1.72	77.56	0.86	0.01	0.22	1.83	.172	0.039
Nativeness x SOA x Training	1.72	77.56	0.86	0.00	0.22	0.44	.619	0.010
Test x Nativeness x SOA	2	90	0.89	0.03	0.20	6.20	.003	0.121
Test x Nativeness x SOA x Training	2	90	0.89	0.01	0.20	1.46	.238	0.031
Compatibility x SOA	2	90	0.91	0.04	0.18	10.26	< .001	0.186
Compatibility x SOA x Training	2	90	0.91	0.00	0.18	0.84	.437	0.018
Test x Compatibility x SOA	1.68	75.58	0.84	0.00	0.12	1.13	.321	0.024
Test x Compatibility x SOA x Training	1.68	75.58	0.84	0.00	0.12	1.50	.230	0.032
Nativeness x Compatibility x SOA	2	90	0.95	0.00	0.17	0.92	.402	0.020
Nativeness x Compatibility x SOA x Training	2	90	0.95	0.00	0.17	0.77	.468	0.017
Test x Nativeness x Compatibility x SOA	2	90	0.94	0.00	0.18	0.95	.393	0.021
Test x Nativeness x Compatibility x SOA x Training	2	90	0.94	0.00	0.18	1.09	.339	0.024

*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, *p*-values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.

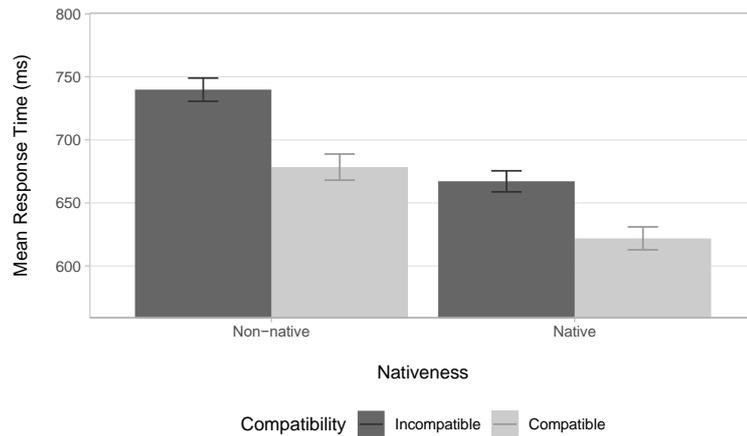
$SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

After errors were removed, log-transformed RT analyses (see Figure 5.8, Table 5.5, and Table 5.6) revealed a main effect of test, with slower RTs for the pre-training session than for the post-training session. The main effect of nativeness revealed slower RTs for non-native responses than for native responses. The main RT compatibility effect revealed slower RTs for incompatible trials than for compatible trials ( $\Delta = 57$  ms). Follow-up  $t$  tests for the main effect of SOA revealed faster RTs for longer SOAs (all  $p < .001$ ).

**Table 5.6** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within test, nativeness, compatibility, and SOA.

	Level	Mean	RT(ms)	SE	95% CI
Test	Pre-training	6.52	677	0.03	[6.46, 6.58]
	Post-training	6.47	643	0.03	[6.40, 6.53]
Nativeness	Non-native	6.54	691	0.03	[6.47, 6.60]
	Native	6.45	630	0.03	[6.38, 6.51]
Compatibility	Incompatible	6.54	689	0.03	[6.48, 6.59]
	Compatible	6.45	632	0.03	[6.38, 6.51]
SOA (ms)	600	6.58	718	0.03	[6.51, 6.64]
	800	6.49	659	0.03	[6.43, 6.55]
	1000	6.41	607	0.03	[6.35, 6.47]

The two-way interaction between nativeness and compatibility (see Figure 5.9 and Table 5.7) was also significant, and follow-up  $t$  tests revealed a larger RT compatibility effect for non-native responses ( $\Delta = 66$  ms,  $t(46) = 9.60$ ,  $p < .001$ ) than for native responses ( $\Delta = 49$  ms,  $t(46) = 9.50$ ,  $p < .001$ ). To analyse this interaction in a different approach, follow-up  $t$  tests revealed a larger RT difference between non-native and native responses for incompatible trials ( $\Delta = 70$  ms,  $t(46) = 5.25$ ,  $p < .001$ ) than for compatible trials ( $\Delta = 53$  ms,  $t(46) = 4.14$ ,  $p < .001$ ).

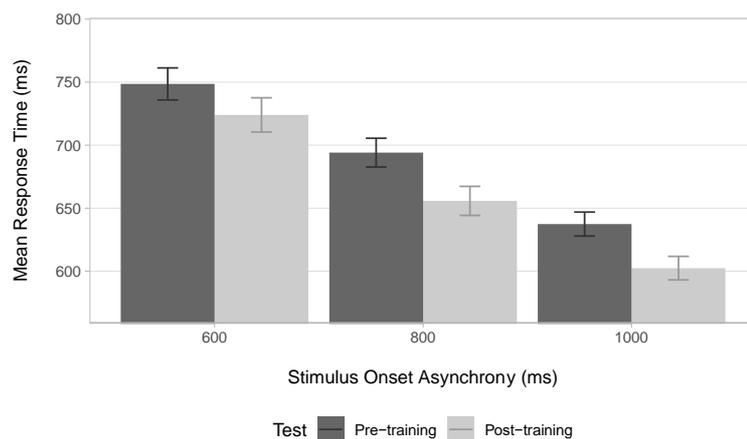


**Figure 5.9** Mean RTs ± SEs for each compatibility condition in each nativeness session.

**Table 5.7** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between nativeness and compatibility.

Nativeness	Compatibility	Mean	RT(ms)	SE	95% CI
Non-native	Incompatible	6.59	725	0.03	[6.53, 6.65]
	Compatible	6.49	659	0.04	[6.42, 6.56]
Native	Incompatible	6.48	655	0.03	[6.43, 6.54]
	Compatible	6.41	606	0.03	[6.34, 6.47]

The two-way interaction between test and SOA (see Figure 5.10 and Table 5.8) was significant, and follow-up *t* tests revealed significant reductions in RTs at all SOAs (all  $p < .04$ ). Further comparisons of RT reductions across SOAs revealed a smaller reduction at SOA1 than at SOA2-3 (both  $p < .01$ ).

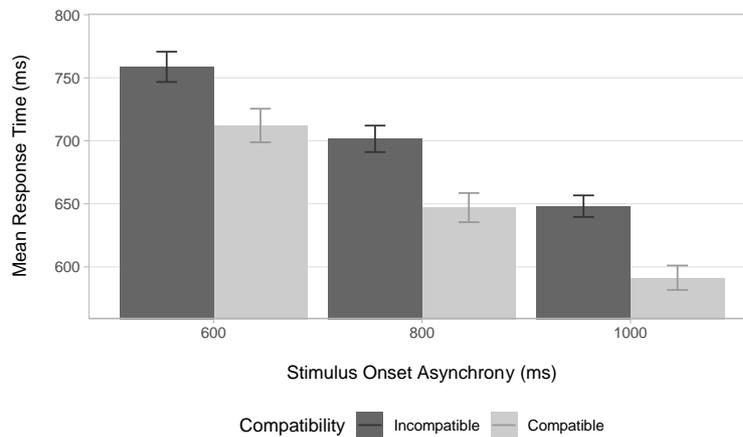


**Figure 5.10** Mean RTs ± SEs at each SOA for each testing session.

**Table 5.8** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between test and SOA.

Test	SOA (ms)	Mean	RT(ms)	SE	95% CI
Pre-training	600	6.60	731	0.03	[6.53, 6.66]
	800	6.44	626	0.03	[6.38, 6.50]
	1,000	6.46	640	0.03	[6.40, 6.53]
Post-training	600	6.52	679	0.03	[6.46, 6.58]
	800	6.56	704	0.03	[6.49, 6.63]
	1,000	6.38	589	0.03	[6.32, 6.44]

The two-way interaction between compatibility and SOA (see Figure 5.11 and Table 5.9) was significant, and follow-up *t* tests revealed significant RT compatibility effects at all SOAs (all  $p < .001$ ). Further comparisons of RT compatibility effects across SOAs revealed a smaller RT compatibility effect at SOA1 than at SOA2-3 (both  $p < .003$ ).



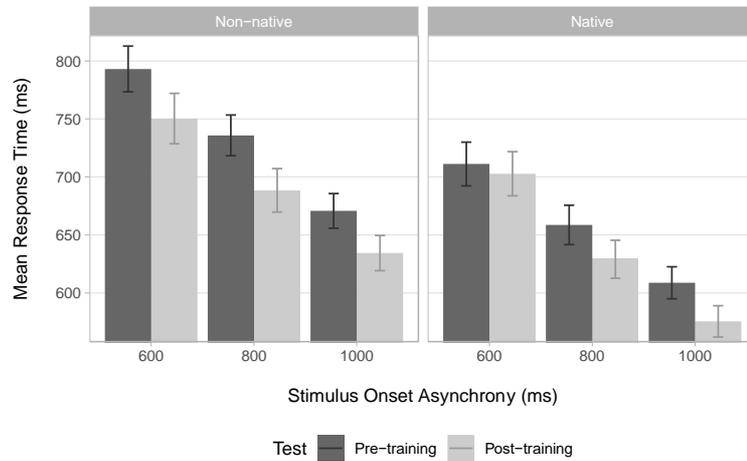
**Figure 5.11** Mean RTs ± SEs at each SOA for each compatibility condition.

**Table 5.9** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between compatibility and SOA.

Compatibility	SOA (ms)	Mean	RT(ms)	SE	95% CI
Incompatible	600	6.61	743	0.03	[6.553, 6.67]
	800	6.46	638	0.03	[6.406, 6.511]
	1,000	6.45	630	0.03	[6.38, 6.513]
Compatible	600	6.54	689	0.03	[6.478, 6.591]
	800	6.54	693	0.03	[6.473, 6.609]
	1,000	6.36	578	0.03	[6.295, 6.423]

The three-way interaction between test, nativeness, and SOA (see Figure 5.12 and Table 5.10) was significant, and follow-up *t* tests revealed significant reductions in RTs after training at all SOAs for non-native responses and at SOA2-3 for native responses (all  $p < .04$ ). To further analyse this interaction in a different approach, follow-up *t* tests revealed that RTs

were slower for non-native responses than for native responses at all SOAs in both pre-training and post-training sessions (all  $p < .02$ ) and that the RT difference between two nativeness sessions decreased significantly after training at SOA1 ( $t(46) = 2.48, p = .017$ ).



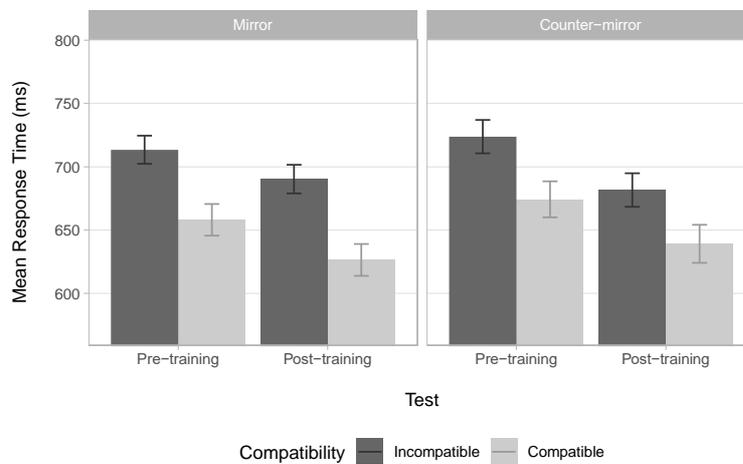
**Figure 5.12** Mean RTs  $\pm$  SEs at each SOA in each testing session for each nativeness session.

**Table 5.10** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between test, nativeness, and SOA.

Nativeness	Test	SOA (ms)	Mean	RT(ms)	SE	95% CI
Non-native	Pre-training	600	6.65	774	0.03	[6.59, 6.72]
		800	6.58	718	0.03	[6.51, 6.64]
		1,000	6.49	657	0.03	[6.43, 6.55]
	Post-training	600	6.59	725	0.04	[6.51, 6.66]
		800	6.50	667	0.04	[6.43, 6.58]
		1,000	6.43	618	0.03	[6.36, 6.49]
Native	Pre-training	600	6.54	692	0.03	[6.47, 6.61]
		800	6.46	642	0.03	[6.40, 6.53]
		1,000	6.39	596	0.03	[6.33, 6.45]
	Post-training	600	6.53	684	0.03	[6.46, 6.60]
		800	6.42	613	0.03	[6.35, 6.48]
		1,000	6.33	563	0.03	[6.27, 6.40]

The three-way interaction between training, test, and compatibility (see Figure 5.13 and Table 5.11) was not significant ( $p = .079$ ), but it indicated a trend that mirror and counter-mirror training seemed to affect compatibility effects differently. Since the current study was underpowered due to an inadequate sample size, further analyses were conducted for this non-significant interaction. Follow-up  $t$  tests first revealed significant RT compatibility effects in both pre- and post-training sessions in both groups. Further comparisons of RT compatibility effects between pre- and post-training sessions for each group revealed that the RT compatibility effect slightly increased from 59 ms to 68 ms after mirror training ( $\Delta = 9$  ms,

$t(23) = 1.92, p = .068$ ) but did not change after counter-mirror training ( $\Delta = -6$  ms,  $t(22) = -.48, p = .640$ , though numerically it decreased from 53 ms to 47 ms).



**Figure 5.13** Mean RTs ± SEs for each compatibility condition in each testing session for each group.

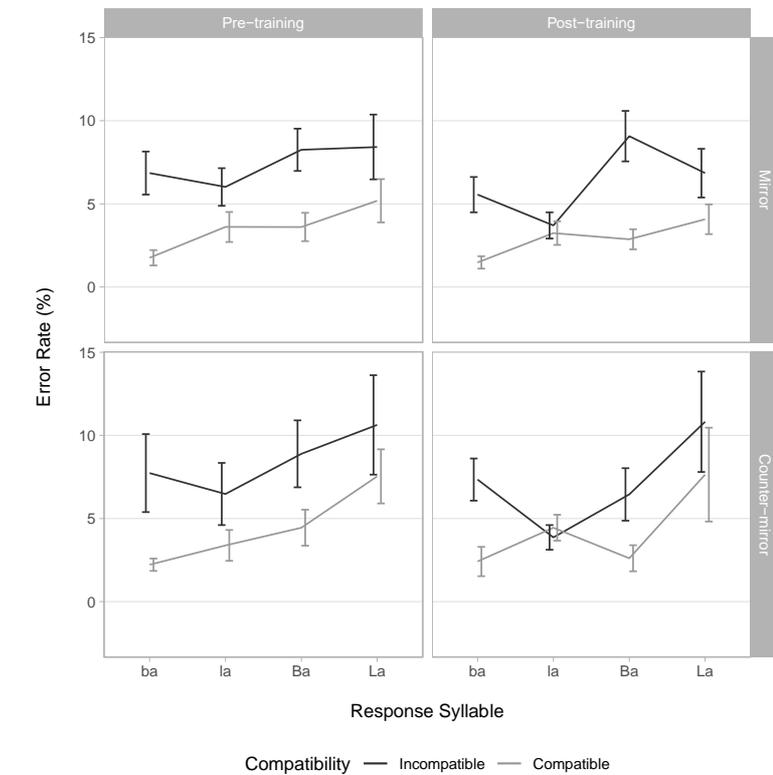
**Table 5.11** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between training, test, and compatibility.

Training	Test	Compatibility	Mean	RT(ms)	SE	95% CI
Mirror	Pre-training	Incompatible	6.55	702	0.04	[6.48, 6.63]
		Compatible	6.47	643	0.05	[6.38, 6.56]
	Post-training	Incompatible	6.52	678	0.04	[6.44, 6.60]
		Compatible	6.41	610	0.05	[6.33, 6.51]
Counter-mirror	Pre-training	Incompatible	6.57	710	0.04	[6.49, 6.64]
		Compatible	6.49	657	0.05	[6.40, 6.58]
	Post-training	Incompatible	6.50	666	0.04	[6.42, 6.59]
		Compatible	6.43	620	0.05	[6.33, 6.53]

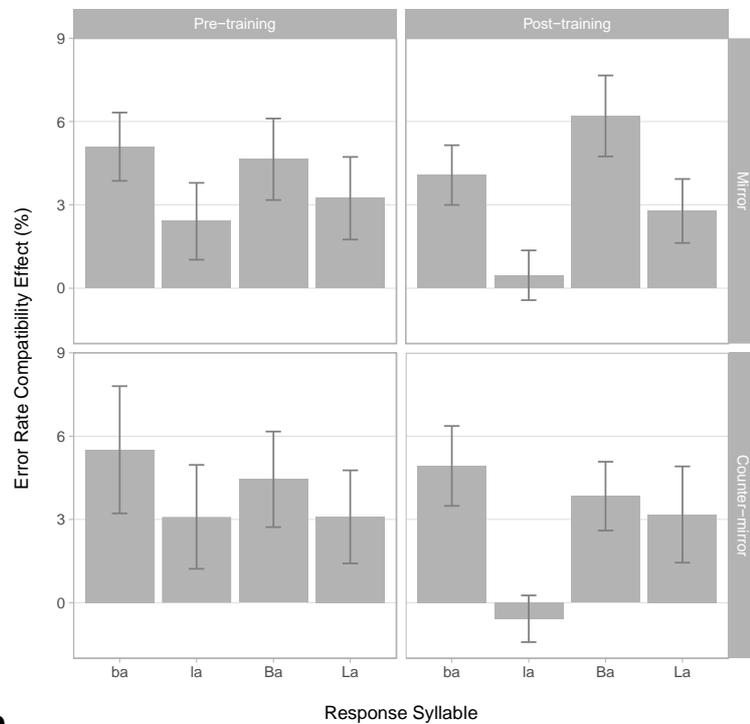
To further examine Prediction 11 and 12 given in Section 5.3, a four-way repeated-measures ANOVA between training, test, compatibility, and SOA was conducted for non-native actions. The main effect of test,  $F(1, 45) = 12.06, p < .001, \eta^2_p = .211$ , revealed a faster RT ( $\Delta = 46$  ms) after training. The RT compatibility effect ( $\Delta = 66$  ms) was significant,  $F(1, 45) = 96.20, p < .001, \eta^2_p = .681$ . The main effect of SOA was significant,  $F(1.23, 55.22) = 180.30, p < .001, \eta^2_p = .800$ , and follow-up  $t$  tests revealed faster RTs for longer SOAs (all  $p < .001$ ). The two-way interaction between compatibility and SOA was significant,  $F(2, 90) = 4.90, p = .010, \eta^2_p = .098$ , and follow-up  $t$  tests revealed a smaller RT compatibility effect at SOA1 than at SOA2-3 (both  $p < .007$ ).

## 5.4.2 Testing sessions – syllable effects

### Error rate analysis



**a**



**b**

**Figure 5.14 a** ERs  $\pm$  SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training

(bottom-right) sessions in the counter-mirror group. **b** ER compatibility effects  $\pm$  SEs for each syllable response in each testing session for each training group.

**Table 5.12** Four-way ANOVA summary for ERs, as a function of training, test, compatibility, and syllable.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	Epsilon	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Training	1	45	1.00	195.76	12466.36	0.71	.405	.015
Test	1	45	1.00	116.07	1562.07	3.34	.074	.069
Test x Training	1	45	1.00	0.95	1562.07	0.03	.869	.001
Compatibility	1	45	1.00	2334.43	3497.22	30.04	< .001	.400
Compatibility x Training	1	45	1.00	1.44	3497.22	0.02	.892	.000
Syllable	1.57	70.71	0.52	1347.98	10018.83	6.06	.007	.119
Syllable x Training	1.57	70.71	0.52	295.15	10018.83	1.33	.268	.029
Test x Compatibility	1	45	1.00	32.31	879.98	1.65	.205	.035
Test x Compatibility x Training	1	45	1.00	6.23	879.98	0.32	.575	.007
Test x Syllable	2.29	103.23	0.77	13.93	3299.40	0.19	.855	.004
Test x Syllable x Training	2.29	103.23	0.77	90.22	3299.40	1.23	.299	.027
Compatibility x Syllable	3	135	0.97	395.95	2657.17	6.71	< .001	.130
Compatibility x Syllable x Training	3	135	0.97	23.14	2657.17	0.39	.759	.009
Test x Compatibility x Syllable	2.20	99.17	0.74	70.88	2137.54	1.49	.229	.032
Test x Compatibility x Syllable x Training	2.20	99.17	0.74	17.76	2137.54	0.37	.709	.008

*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

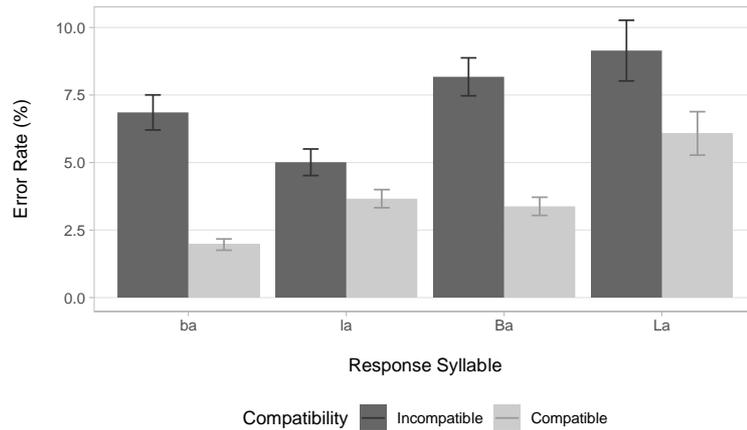
ER analyses (see Figure 5.14, Table 5.12, and Table 5.13) revealed an ER compatibility effect, with a higher ER for incompatible trials than for compatible trials ( $\Delta = 3.53$ ). Follow-up  $t$  tests for the main effect of syllable revealed a higher ER for La responses than for ba responses ( $p = .049$ ).

**Table 5.13** ERs, SEs, and 95% CIs for each level within compatibility and syllable.

	Level	Mean	SE	95% CI
Compatibility	Incompatible	7.31	0.88	[5.54, 9.08]
	Compatible	3.78	0.42	[2.95, 4.62]
Syllable	ba	4.42	0.53	[3.34, 5.50]
	la	4.34	0.47	[3.39, 5.30]
	Ba	5.78	0.65	[4.48, 7.07]
	La	7.65	1.32	[4.99, 10.30]

The two-way interaction between compatibility and syllable (see Figure 5.15 and Table 5.14) was significant, and follow-up  $t$  tests revealed significant ER compatibility effects

for ba, Ba, and La responses (all  $p < .002$ ) but not for la responses. Further comparisons of ER compatibility effects across all syllable responses revealed a smaller ER compatibility effect for la responses than for ba and Ba responses. To analyse the data the other way around, further comparisons of ERs between ba and la responses and between Ba and La responses for each compatibility level revealed a higher ER for incompatible ba responses than for incompatible la responses, a lower ER for compatible ba responses than for compatible la responses, and a lower ER for compatible Ba responses than for compatible La responses.

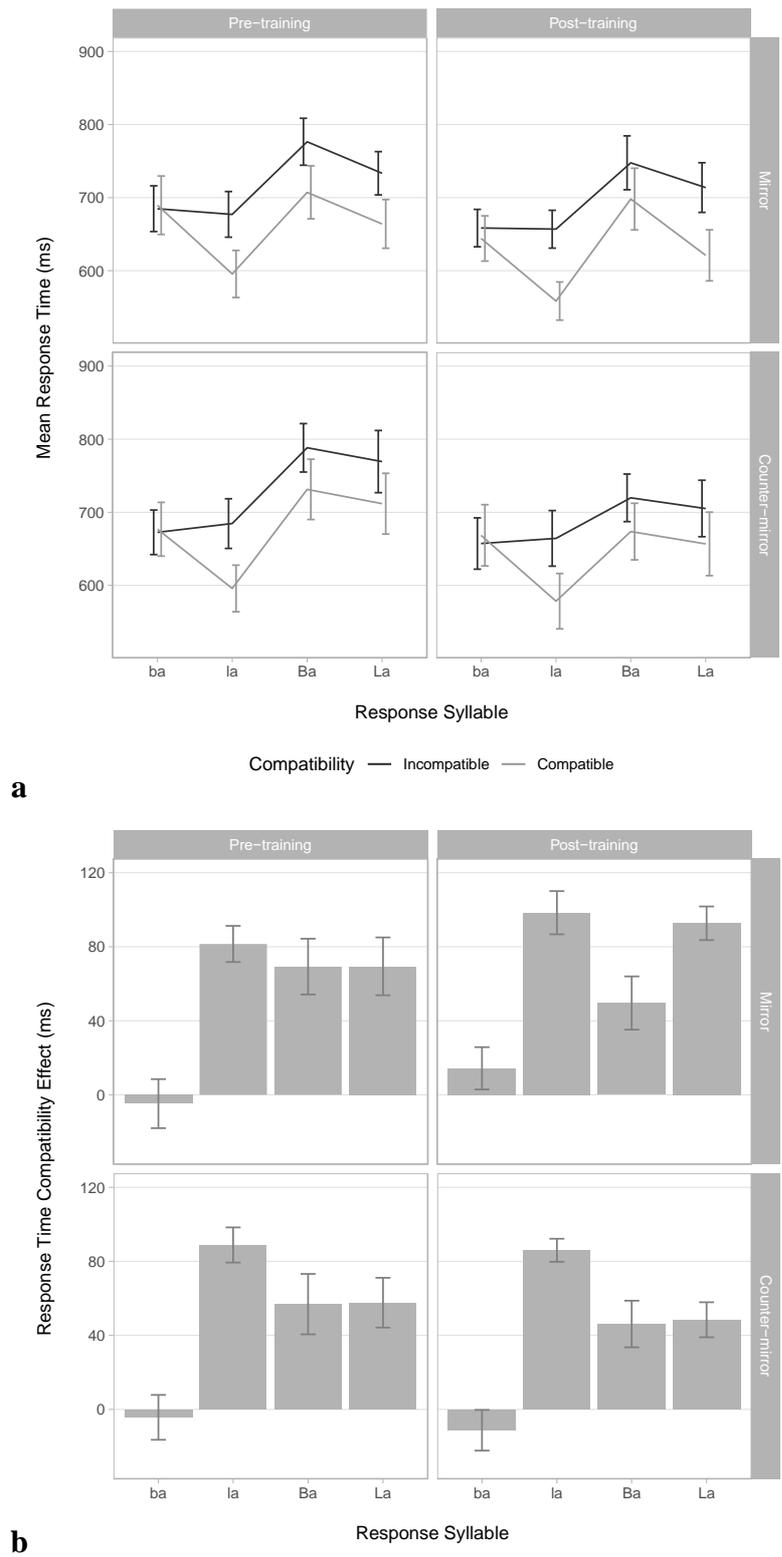


**Figure 5.15** ERs  $\pm$  SEs for each compatibility condition for each response syllable.

**Table 5.14** ERs, SEs, and 95% CIs for each level within the interaction between compatibility and syllable.

Compatibility	Syllable	Mean	SE	95% CI
Incompatible	ba	6.87	0.93	[5.00, 8.74]
	la	5.02	0.71	[3.59, 6.44]
	Ba	8.17	1.01	[6.14, 10.19]
	La	9.18	1.60	[5.96, 12.40]
Compatible	ba	1.97	0.30	[1.37, 2.57]
	la	3.67	0.48	[2.70, 4.64]
	Ba	3.38	0.48	[2.41, 4.35]
	La	6.11	1.14	[3.82, 8.40]

**Response time analysis**



**Figure 5.16 a** Mean RTs  $\pm$  SEs in each experimental condition. Four panels represent the pre- (top-left) and post-training (top-right) sessions in the mirror group and the pre- (bottom-left) and post-training

(bottom-right) sessions in the counter-mirror group. **b** RT compatibility effects  $\pm$  SEs for each syllable response in each testing session for each training group.

**Table 5.15** Four-way ANOVA summary for log-transformed RTs, as a function of training, test, compatibility, and syllable.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	<i>Epsilon</i>	$SS_{Num}$	$SS_{Den}$	<i>F</i>	<i>p</i>	$\eta^2_p$
Training	1	45	1.00	0.01	29.82	0.02	.883	.000
Test	1	45	1.00	0.51	2.19	10.48	.002	.189
Test x Training	1	45	1.00	0.01	2.19	0.24	.624	.005
Compatibility	1	45	1.00	1.42	0.53	120.54	< .001	.728
Compatibility x Training	1	45	1.00	0.02	0.53	1.78	.189	.038
Syllable	1.68	75.39	0.56	2.35	4.29	24.67	< .001	.354
Syllable x Training	1.68	75.39	0.56	0.04	4.29	0.45	.606	.010
Test x Compatibility	1	45	1.00	0.00	0.07	1.73	.195	.037
Test x Compatibility x Training	1	45	1.00	0.01	0.07	3.36	.074	.069
Test x Syllable	1.48	66.50	0.49	0.04	1.06	1.71	.195	.037
Test x Syllable x Training	1.48	66.50	0.49	0.04	1.06	1.82	.179	.039
Compatibility x Syllable	3	135	0.94	0.50	0.57	39.16	< .001	.465
Compatibility x Syllable x Training	3	135	0.94	0.01	0.57	0.88	.455	.019
Test x Compatibility x Syllable	2.54	114.27	0.85	0.01	0.27	1.73	.173	.037
Test x Compatibility x Syllable x Training	2.54	114.27	0.85	0.01	0.27	0.78	.487	.017

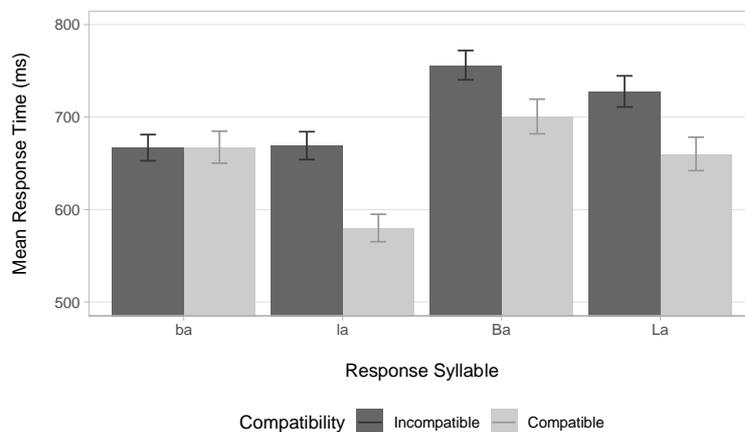
*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, *p*-values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

Log-transformed RT analyses (see Figure 5.16, Table 5.15, and Table 5.16) revealed a main effect of test, with slower RTs for the pre-training session than for the post-training session. The main RT compatibility effect revealed slower RTs for incompatible trials than for compatible trials ( $\Delta = 57$  ms). Follow-up *t* tests for the main effect of syllable revealed faster RTs for la responses than for other three syllable responses and slower RTs for Ba responses than for other three syllable responses (all  $p < .003$ ).

**Table 5.16** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within test, compatibility, and syllable.

	Level	Mean	RT (ms)	SE	95% CI
Test	Pre-training	6.52	678	0.03	[6.46, 6.58]
	Post-training	6.47	644	0.03	[6.40, 6.53]
Compatibility	Incompatible	6.54	690	0.03	[6.48, 6.59]
	Compatible	6.45	632	0.03	[6.38, 6.51]
Syllable	ba	6.48	652	0.03	[6.42, 6.54]
	la	6.41	609	0.03	[6.35, 6.47]
	Ba	6.57	710	0.03	[6.50, 6.63]
	La	6.51	675	0.03	[6.45, 6.58]

The two-way interaction between compatibility and syllable (see Figure 5.17 and Table 5.17) was significant, and follow-up *t* tests revealed significant RT compatibility effects for la, Ba, and La responses (all  $p < .002$ ) but not for ba responses. Further comparisons of RT compatibility effects across all syllable responses revealed a larger RT compatibility effects for la responses than for other three syllable responses and a smaller RT compatibility effects for ba responses than for other three syllable responses (all  $p < .03$ ). To analyse the data the other way around, further comparisons of RTs between ba and la responses and between Ba and La responses for each compatibility level revealed a faster RT for compatible la responses than for compatible ba responses and a faster RT for La responses than for Ba responses at each compatibility level.



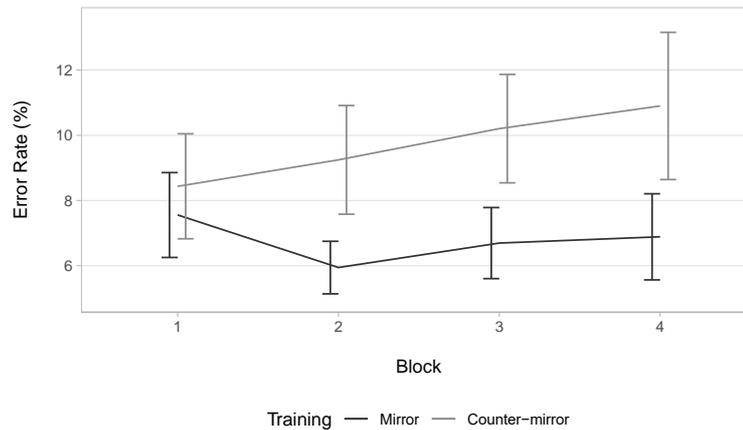
**Figure 5.17** Mean RTs ± SEs for each compatibility condition for each response syllable.

**Table 5.17** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level within the interaction between compatibility and syllable.

Syllable	Compatibility	Syllable	Mean	RT (ms)	SE	95% CI
ba	Incompatible	ba	6.48	655	0.03	[6.45, 6.58]
	Compatible	ba	6.48	650	0.03	[6.50, 6.63]
la	Incompatible	la	6.49	655	0.03	[6.43, 6.54]
	Compatible	la	6.34	565	0.03	[6.41, 6.54]
Ba	Incompatible	Ba	6.61	742	0.03	[6.43, 6.55]
	Compatible	Ba	6.52	680	0.03	[6.27, 6.40]
La	Incompatible	La	6.57	711	0.03	[6.55, 6.67]
	Compatible	La	6.46	640	0.04	[6.45, 6.59]

### 5.4.3 Training sessions

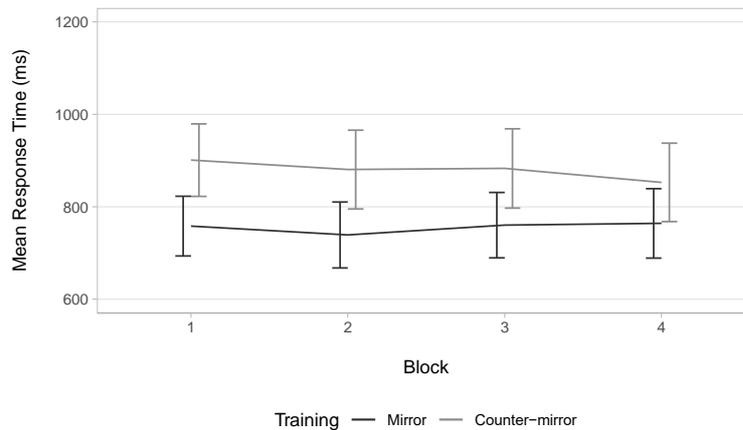
#### Error rate analysis



**Figure 5.18** ERs  $\pm$  SEs for each block in each group.

On average, participants made 8.3% errors (incorrect responses: 3.3%; missed trials: 1.7%; anticipatory responses: 2.4%; outliers: 0.9%). ER analyses revealed no significant main effect or interaction, though based on the observation of Figure 5.18, there seems to be a lower ER for the mirror group than for the counter-mirror group.

## Response time analysis



**Figure 5.19** Mean RTs  $\pm$  SEs for each block in each group.

RT analyses revealed no significant main effect or interaction, though based on the observation of Figure 5.19, RTs seem to be faster for the mirror group than for the counter-mirror group.

## 5.5 Discussion

The two objectives of the current study were:

- **Objective 4:** to determine whether sensorimotor experience is critically involved in establishing the perception-production matching link underlying novel speech actions.
- **Objective 5:** to explore at what level training-induced modification occurs in the perception-production link.

Measured using a speech SRC task, automatic imitation effects were used to index perception-induced motor activation with a larger effect indicating greater motor activation. Greater motor activation then indicated a stronger perception-production matching link presumably underlying the imitative capacity. A between-subjects training study was conducted, in which participants were randomly assigned to either a mirror or a counter-mirror training group. Automatic imitation effects of both non-native and native speech actions were measured before and after training. As discussed in Section 2.1.3, the associative account of imitation proposes that the matching properties of the perception-production link underlying the imitative capacity are acquired through sensorimotor experience; whereas the genetic account

proposes that such matching properties are innate, though sensory or motor experience may facilitate the development of the matching.

ER and RT compatibility effects (i.e., automatic imitation) were observed for both non-native and native responses, and interestingly the RT compatibility effect was larger for non-native responses than for native responses. Further analyses revealed no RT compatibility effect for ba responses, which probably reduced the overall compatibility effect for native responses. In the following sections, I will discuss these findings in more detail. Of primary interest, statistical analyses revealed a trend ( $p = .079$  for the interaction between training, test, and compatibility) that the RT compatibility effect across both non-native and native responses slightly increased following mirror training ( $\Delta = 9$  ms,  $t(23) = 1.92$ ,  $p = .068$ ) but did not change following counter-mirror training ( $\Delta = -6$  ms,  $t(22) = -.48$ ,  $p = .640$ ). This finding suggested that sensorimotor training with novel actions modulates the imitative capacity underlying both novel and familiar actions that are produced using the same effectors. This finding indirectly supported Prediction 11 given in Section 5.3, though further analysis for only non-native responses supported neither Prediction 11 nor 12, which could be due to an insufficient sample size. More importantly, consistent with Prediction 14 given in Section 5.3, this finding further suggested that training-induced modification seemed to occur at the feature level in the perception-production link. In the following sections, I will discuss this finding in accordance with the associative hypothesis. However, readers should be aware that, since the current study was underpowered due to an insufficient sample size, it was too early to decisively conclude whether sensorimotor training did modulate automatic imitation effects across both non-native and native speech actions, or this finding was likely to be false positive.

### ***5.5.1 Automatic imitation of non-native speech actions***

Participants in the current study were assumed to have had less sensorimotor experience of concurrently perceiving and producing the selected non-native speech actions, as compared to their native counterparts. Thus, automatic imitation of non-native speech actions was predicted to be smaller than that of native speech actions. However, current results revealed a larger RT compatibility effect for non-native responses than for native responses, demonstrating the opposite to what was predicted. Further analyses revealed a larger RT compatibility effect for la responses (90 ms) than for either Ba (62 ms) or La (71 ms); yet, no RT compatibility effect was observed for ba responses, which probably reduced the overall compatibility effect for native responses. Thus, it was difficult to conclude whether automatic imitation of non-native speech actions observed in the current study tended to be larger or smaller than that of native speech actions. However, it was still unexpected to observe a 70-ms automatic imitation effect for non-native responses in the current study (e.g., as compared

to the 40-ms automatic imitation effect observed in the study reported in Chapter 4) for initially participants were assumed to have had little sensorimotor experience with either non-native action. Some participants did not even know how to produce the /La/ sound after they first saw and heard the speaker's /La/ articulation. The overall higher ER and slower RT for non-native responses than for native responses also suggested participants' relative inexperience in producing the selected non-native actions. What then contributed to or led to the non-native automatic imitation effect observed in the current study?

### ***Effector compatibility***

One possibility was that the effector compatibility gave rise to the non-native automatic imitation effect in the current study. As mentioned in Section 2.1.1, automatic imitation has been found to be effector-specific (e.g., Bach & Tipper, 2007; Catmur et al., 2009; Gillmeister et al., 2008). In another study by Leighton & Heyes (2010), they first presented both pictures of a hand and a mouth remaining in a neutral condition and then turned one effector into an open or closed posture with the other still remaining in the initial position. In the meantime, participants had to ignore the hand and the mouth shown on the screen (i.e., distractor stimuli) and respond to the assigned letter prompts by opening or closing their hand or mouth. Thus, there were two types of compatibility effects: effector and movement. Leighton & Heyes observed an effector compatibility effect that occurred regardless of whether distractor and response movements were compatible or incompatible, though it was smaller for movement-incompatible trials. The non-native automatic imitation effect in the current study could at least partially result from the effector compatibility. Future studies could include both pairs of non-native and native speech distractors that share the same places of articulation and ask participants to make non-native responses only. If the effector compatibility fully accounts for the non-native automatic imitation, compatibility effects should be the same regardless of whether the perceived speech actions are non-native or native.

### ***Native and non-native speech interaction***

From a language/speech perspective, another possibility was that the non-native automatic imitation occurred in the current study because non-native speech actions /Ba/ and /La/ 'borrowed' the perception-production link from closely related native speech actions. The Perceptual Assimilation Model of non-native speech perception proposes that non-native speech sounds are perceived in terms of their similarities and differences, as compared to listeners' native sound categories, and that listeners perceptually assimilate non-native sounds to native categories whenever possible, based on detected commonalities in phonetic features (Best, 1995; Best & Tyler, 2007). Non-native speech production is also influenced by speakers'

native language, and second language learners tend to speak with foreign accents (Piske, MacKay, & Flege, 2001). In psychology, *stimulus generalisation* is a common feature of associative learning (Leighton & Heyes, 2010; McLaren & Mackintosh, 2002; Pearce, 1987), which occurs when the response-stimulus link associated with the stimulus AX is activated by the stimulus BX that shares the same X element with AX. Accordingly, in the case of speech, perceiving a non-native action might activate the perception-production link underlying the native action that shares one or more phonetic features with the perceived non-native action. The Perceptual Assimilation Model also proposes that listeners directly perceive articulatory gestures underlying received speech signals and, through perceptual learning, come to extract articulatory invariants in non-native speech stimuli. Following the Perceptual Assimilation Model, speech stimuli are essentially action stimuli that both perceivers and speakers can produce (correctly or incorrectly). Thus, it would not be surprising if people associate the production of a non-native action with the production of an articulatorily related native action (i.e., response generalisation) or even develop the motor command for non-native speech based on the already existing motor representation of native speech.

Participants in the current study not only saw and heard either non-native speech action, but also were given verbal instructions of how to produce either action. In other words, participants obtained enough information that could help them figure out how each non-native speech action is similar to and/or different from their native speech sounds in terms of articulatory gestures, and they might consequently associate each non-native action with the closest native category. To investigate how American English speakers assimilated Zulu consonants to their native sound categories, Best et al. (2001) asked participants to transliterate syllables they heard using English orthography. Half of participants labelled the voiced alveolar lateral fricative (i.e., /L/ in the current study) as /l/ or /lh/, and others labelled it as /z/, /zl/, /zhl/, /thl/, /th/ with /z/ or /v/, or /szt/. This finding demonstrated how people are able to detect some phonetic features of perceived non-native speech sounds and assimilate them to their native sound categories. Participants in the current study might adopt similar strategies for both perception and production, though they were not asked to transliterate either non-native sound using English orthography. According to their response recordings, it seemed that some participants tended to say /zla/, /zha/, or /szta/ when they were supposed to say /La/ and say /bwa/ or /bra/ when supposed to say /Ba/. These observations suggested how participants possibly ‘assimilated’ /B/ and /L/ to their native sound categories.

The Perceptual Assimilation Model further proposes that non-native speech sounds may be perceived and assimilated to listeners’ native sound categories in one of three ways. A non-native speech sound could be assimilated and *categorised* as a native sound and perceived as a good, an acceptable, or a deviant exemplar of the category. A non-native speech sound could also be assimilated to listeners’ native sound system but as *uncategorised* when

it is perceived as speech-like but fails to resemble any native category. In addition, a non-native speech sound could be *non-assimilable* when it is heard as a non-speech sound and thus cannot be assimilated to the perceiver's native sound system. Since the current study did not ask participants to identify or judge each non-native speech sound in relation to their native sounds, it was difficult to conclude whether and how participants assimilated either non-native sound to their native categories. Future studies could first establish the degree of assimilability for each non-native speech sound and then examine whether and how it is correlated with the corresponding automatic imitation effect. If non-native speech actions tend to borrow the perception-production link from closely related native speech actions, non-native actions identified with a higher degree of assimilability would be expected to elicit a larger automatic imitation effect. Additionally, for non-native speech sounds that are assimilated to a single native category, future studies could also investigate how automatic imitation of these non-native speech sounds changes as perceivers' discrimination ability improves.

### ***Inhibitory control of imitation***

So far, discussions of the non-native automatic imitation effect observed in the current study have been concerned about possibilities of how observing non-native speech actions led to motor activation, which then interacted with participants' articulatory responses. However, an increase in automatic imitation could result from different sources: more facilitation (i.e., faster responses) in compatible trials due to greater motor activation by action perception (Press et al., 2007), or more interference (i.e., slower responses) in incompatible trials due to less successful inhibition of unwanted motor activation by action perception (Sowden & Catmur, 2015). Thus, it was possible that the robust non-native automatic imitation effect in the current study resulted from participants' weak inhibitory control to suppress unwanted motor activation by the observation of non-native speech actions. The 'novel' actions used in the current study were not completely new. Participants had chance to see and practice the non-native speech actions until they were confident to perform these actions in SRC tasks and they might have encountered and/or performed similar actions even before the study. It was possible that the perception-production matching link underlying novel actions could easily be built through limited sensorimotor experience, and the link would then be further strengthened while people learn to perfect their motor skills and use these actions during social interaction. However, it remains unclear whether people's ability to control their imitative tendency for newly learned actions develops automatically in parallel as the corresponding perception-production link is being built through sensorimotor learning.

People tend to imitate their interaction partners, but also modulate the degree to which they imitate depending on various factors so as to maintain high-quality social interaction

(Chartrand & Lakin, 2013; Heyes, 2011; Wang & Hamilton, 2012). In other words, people do not imitate all the time; more precisely, people do not overtly imitate every action they see during social interaction and thus have to learn to inhibit unwanted motor activation by action observation and prevent it from being executed overtly. Developmental research has found that young children have weak inhibitory control and struggle to suppress unwanted imitative behaviours (Diamond & Taylor, 1996; Simpson & Riggs, 2011), suggesting that the tendency to imitate observed actions emerges first before children acquire the ability to control this tendency and they probably acquire this ability during social interaction. That is, when interacting with others, children learn about when to imitate more is rewarding and when to imitate less is appropriate. Accordingly, it was possible that the non-native automatic imitation effect observed in the current study partially also resulted from participants' weak inhibitory control to suppress unwanted motor activation triggered by the perception of the speaker's non-native speech articulation.

A top-down control process has been proposed to modulate imitation and its proposed neural substrate, the human mirror system, and brain areas engaged in this control process include the inferior frontal cortex (IFC), the temporoparietal junction (TPJ), and the medial prefrontal cortex (mPFC; Brass, Ruby, & Spengler, 2009; Catmur et al., 2009; Hogeveen et al., 2015; Santiesteban et al., 2012; Sowden & Catmur, 2015). Studies using brain stimulation techniques have provided causal evidence that the IFC and TPJ are involved in imitation control (Hogeveen et al., 2015; Santiesteban et al., 2012; Sowden & Catmur, 2015). Using repetitive TMS to disrupt the brain area functioning, Sowden & Catmur observed an enhanced automatic imitation (i.e., imitative compatibility) effect after applying TMS to the right TPJ, relative to the condition where TMS was applied to a control site. Additionally, in the same study, the spatial compatibility effect was not affected by whether applying TMS to the TPJ or to the control site. Sowden & Catmur proposed that repetitive TMS disrupted the TPJ's ability to control the tendency to imitate and thus automatic imitation was enhanced. In another study, Hogeveen et al. found a reduced automatic imitation effect after using transcranial direct current stimulation (tDCS) to enhance cortical excitability in either the right IFC or right TPJ, but not after applying tDCS to a control site. The authors proposed that excitation of the TPJ resulted in an increased ability to inhibit imitation and thus automatic imitation was reduced. Future studies could examine whether stimulating brain areas involved in imitation control would lead to different change in automatic imitation of native and non-native actions. If participants' inhibitory control to suppress unwanted imitative behaviours was weaker for non-native actions than for native actions, excitatory stimulation would be expected to lead to larger reduction in non-native automatic imitation than in native automatic imitation and inhibitory stimulation would lead to smaller enhancement in non-native automatic imitation than in native automatic imitation.

### *Syllable selection for speech automatic imitation*

Another interesting finding in the current study concerned about the automatic imitation effect for native responses. Further analyses revealed a larger RT compatibility effect for la responses than for ba, Ba, or La responses, and the RT compatibility effect for ba responses was not significant. Conversely, ER analyses revealed larger ER compatibility effects for ba and Ba responses than for la responses, and the ER compatibility effect for la responses was not significant. In other words, for the native pair selected in the current study, ba responses tended to elicit ER compatibility effects and la responses tended to elicit RT compatibility effects. Yet, these patterns were not observed for the non-native speech pair whose consonants' places of articulation matched those of the native pair. The /ba/ and /da/ pair used by Kerzel & Bekkering (2000) and used in the study reported in Chapter 4 did not elicit such patterns either, though /d/ and /l/ share the same places of articulation. Therefore, it seemed that consonants' manners of articulation might have played a role in causing different patterns of RT and ER compatibility effects for native responses in the current study. Though produced with different manners of articulation, both /B/ and /L/ are produced by narrowing the air path at some place in the vocal tract and then causing the air flow to become turbulent. In contrast, while /l/ is also produced by narrowing the air path at some place in the vocal tract and then causing slight turbulence, /b/ is produced by first completely obstructing the air flow with lips and then generating a quick burst to release air flow. Hence, /b/ is produced much faster than /l/, /B/, or /L/. It was possible that the fast production of /b/ was more resilient to the compatibility manipulation in terms of RTs and was less likely to be affected by incompatible /la/ distractors. In contrast, the slow production of /l/ was more resilient to the compatibility manipulation in terms of ERs and was less likely to be affected by incompatible /ba/ distractors. The current study was the first to employ a pair of speech actions that differed in both places and manners of articulation, and it could be the manner of articulation or the combination of the two that caused different patterns of ER and RT compatibility effects for native responses. Follow-up studies aiming to tackle this issue could control for both places and manners of articulation to investigate the effect of each phonetic feature in speech SRC tasks in a factorial design study.

#### *5.5.2 Sensorimotor training with non-native speech actions*

Of primary interest in the current study, statistical analyses revealed a trend that the RT compatibility effect across both non-native and native responses slightly increased following mirror training but did not change following counter-mirror training. This finding suggested that sensorimotor training with novel actions modulates the imitative capacity underlying both

novel and familiar actions that are produced using the same effectors. In this section, I will discuss this finding in accordance with the associative hypothesis. However, as mentioned earlier, readers should be aware that, since the current study was underpowered due to an insufficient sample size, it was too early to conclude whether sensorimotor training did modulate automatic imitation effects across both non-native and native speech actions, or this finding was likely to be false positive.

### ***Inhibitory control of imitation***

Importantly, while statistical analyses suggested a trend that mirror and counter-mirror training differently affected automatic imitation across both non-native and native speech actions; observations of Figure 5.8 suggested that the non-native automatic imitation seemed to decrease after training in both groups. It was possible that, while mirror and counter-mirror training respectively enhanced and reduced automatic imitation of non-native actions, there was another process driving this automatic imitation to decrease in both groups. As proposed in the previous section, the robust non-native automatic imitation effect observed in the current study might result from participants' weak inhibitory control to suppress unwanted imitative behaviours. It was possible that, after completing the non-native SRC task in the pre-training session, participants might have learned to some extent to control their imitative tendency for non-native speech actions and thus the corresponding automatic imitation effects decreased after training in both groups. Accordingly, the change in the non-native automatic imitation could be a product of interaction between sensorimotor training and increased inhibitory control. That is, while increased inhibitory control led to a reduction in the non-native automatic imitation effect after training in both groups, counter-mirror training led to a further reduction in the corresponding effect but mirror training conversely led to an enhancement in the corresponding effect. Consequently, the overall decrease in the non-native automatic imitation was smaller after mirror training than after counter-mirror training. Again, readers should be aware that these were speculations based on the observations from Figure 5.8, while statistically the corresponding result was far from being significant ( $p = .453$  for the interaction between training, test, nativeness, and compatibility). As mentioned repeatedly in this chapter, the current study was underpowered, and more participants would be needed before reaching a statistically valid conclusion.

### ***Effector compatibility***

Interestingly, mirror and counter-mirror training in the current study were also suggested to differently affect automatic imitation of native speech actions that were not included in the training session. This finding indicated that the effects of sensorimotor training with novel

actions generalise to familiar actions that are produced using the same effectors as the trained actions. In the previous section, it was proposed that the non-native automatic imitation observed in the current study might at least partially result from the effector compatibility. Sensorimotor training in the current study might in fact modulate the mechanism mediating the effector compatibility, and thus mirror and counter-mirror training effects were observed in automatic imitation of both non-native and native speech actions that are produced using the same effectors. From a speech perspective, there might exist perception-production links underlying each phonetic feature (e.g., place and manner of articulation and voicing), and all are subject to change through sensorimotor training. As discussed in Section 5.1.2, Roon & Gafos (2015) observed an automatic imitation effect in which perceiving syllable articulations that share some phonetic features (e.g., voicing or place of articulation) with the required articulatory responses facilitated participants' performance, relative to perceiving syllable articulations that share fewer features with the responses. Current findings were in line with and further complemented findings from Roon & Gafos, suggesting that observation-induced motor activation might occur at the feature level and that the perception-production link underlying each phonetic feature might be flexible to change through sensorimotor training. The current study only controlled for places of articulation between non-native and native pairs and demonstrated sensorimotor training effects at this feature dimension, as indicated by the finding that the effects of sensorimotor training with non-native actions generalised to native actions that share the same places of articulation. Future studies could further investigate whether sensorimotor training effects also occur at other feature dimensions, such as manners of articulation or voicing patterns.

### ***Native and non-native speech interaction***

Furthermore, as discussed in the previous section, non-native speech perception and production is largely influenced by one's native language, a phenomenon also called *forward transfer* (V. Cook, 2003; Kartushina, Frauenfelder, & Golestani, 2016). That is, people tend to perceptually assimilate non-native sounds to their native sound categories and speak non-native languages with foreign accents. Accordingly, it was also proposed that the non-native automatic imitation effect occurred in the current study probably because non-native speech actions /Ba/ and /La/ 'borrowed' the perception-production link from closely related native actions. In more naturalistic situations, people might associate /B/ and /L/ with other British English consonants rather than /b/ and /l/. However, the introduction of /ba/, /la/, /Ba/, and /La/ only in the current study might implicitly cause participants to associate /B/ and /L/ with /b/ and /l/ respectively, and thus they were likely to borrow the perception-production link underlying /ba/ and /la/. Consequently, what was being altered during sensorimotor training

with non-native actions in the current study might be the perception-production link underlying native actions /ba/ and /la/, though neither native action was used or presented in the training session. In other words, sensorimotor training with non-native speech actions affected the perception-production link underlying closely related native speech actions. A question then arises as to whether it was the perception-production link underlying untrained native speech actions or the link underlying individual phonetic features that was being altered during sensorimotor training with non-native speech actions in the current study.

Research on second language learning has also shown that native speech perception and production could also be affected by a later-acquired second language, a phenomenon called *backward transfer* (V. Cook, 2003; Kartushina et al., 2016). Change in first language processing due to exposure to a second language at the phonetic level has been called *phonetic drift*, which occurs when native speech production drifts toward or away from the phonetic properties of non-native speech (see Chang, 2019 for review). As summarised by Chang (2019), how native speech production drifts in relation to non-native speech depends on factors including similarities between native and non-native speech sounds (Flege, 1987), speakers' tendency to imitate (Sancier & Fowler, 1997), and the recency effect on memory (C. B. Chang, 2013). More importantly, it has been suggested that phonetic drift may occur at two levels: the individual sound and the natural class of sounds. Chang (2012) measured voice onset time (VOT) of English stop consonants spoken by native American English speakers who participated in a six-week Korean learning course. VOT measures the time interval between the release of a stop consonant and the onset of vocal fold vibration, and VOT is longer for English voiceless stops /p t k/ than for English voiced stops /b d g/ (Lisker & Abramson, 1964). Interestingly, for native English speakers, VOT of Korean aspirated stops /p<sup>h</sup> t<sup>h</sup> k<sup>h</sup>/ is perceptibly longer than those of English voiceless stops /p t k/, except the /t<sup>h</sup>/ sound produced by female. After six weeks of Korean classes taught by two female Korean instructors, participants in Chang (2012) were found to produce longer VOT for all native voiceless stops /p t k/, but the increase in VOT of /t/ was relatively smaller. In other words, participants' production of their native English voiceless stops drifted toward Korean aspirated stops after the six-week Korean learning course, even for the /t/ sound whose VOT was already close to that of Korean /t<sup>h</sup>/. Chang (2012) suggested that due to perceptibly overall longer VOT for Korean aspirated stops than for English voiceless stops, the phonetic drift first occurred at the class level where all voiceless stops drifted toward aspirated stops. Additionally, due to relatively close VOT between /t/ and /t<sup>h</sup>/, the phonetic drift also occurred at the individual sound level where the amount of drift was moderated for /t/. The change at the class level suggested by Chang (2012) essentially indicated a change in the phonetic feature, voicing. Findings in the current study could also be considered as a demonstration of how experience with non-native speech affects native speech processing, and it was possible

that training-induced modification of the imitative capacity might occur at two levels as well: individual speech actions and phonetic features.

Following the associative hypothesis, Section 5.1.1 introduced how the associative mechanism is involved in imitation learning of novel actions that consist of familiar elements arranged in a new sequence. Findings of the current study further suggested how the associative mechanism is potentially involved in imitation of novel actions that result from unfamiliar effector-movement combinations. As proposed above, training-induced modification of the imitative capacity might occur at two levels. To further confirm whether sensorimotor training modulates the perception-production link underlying individual phonetic features, future studies could examine the effects of sensorimotor training with a pair of native speech actions on automatic imitation of different pairs of native actions that share one or more phonetic features with the trained actions. If training-induced modification occurs to the automatic imitation of untrained native action pairs, it would indicate the involvement of the associative mechanism at the feature/elemental level and research along this line might help further understand the organisation of the motor cortex, which has remained controversial (Graziano, 2016). However, if training-induced modification does not occur to the automatic imitation of untrained native actions, it would indicate that non-native speech actions/sounds in fact ‘borrowed’ the perception-production matching link from their native counterparts. See arguments below.

Sensorimotor training with non-native speech actions might modulate the perception-production link underlying closely related native speech actions because the non-native actions borrowed the native perception-production link. The borrowing might occur due to stimulus and response generalisation, with the former being a common feature of associative learning and the latter being proposed based on the fact that speech stimuli are essentially action stimuli that both perceivers and speakers produce during communication. Future studies could manipulate the degree of assimilation of non-native sound to native categories and investigate whether sensorimotor training with non-native actions varying in the degree of assimilation would lead to different change in automatic imitation of related native actions. As discussed in Section 4.4.4, findings from second language learning studies seem to suggest some kind of dissociation between perception and production systems underlying non-native speech processing. Such dissociation might be due to the borrowing of the native perception-production link, through which perceiving a non-native sound always leads to the activation of the native perception-production link. Consequently, it might lead to ambiguous activation in motor representations of both closely related non-native and native actions. Such ambiguous motor activation might prevent the establishment of the non-native perception-production link, especially with perception and production training given separately. More importantly, the Simulation Theory of speech perception discussed in Section 4.4.3 has

proposed that perception-induced motor activation facilitates prediction of the incoming signals, which in turn supports speech comprehension (Gambi & Pickering, 2013; Pickering & Garrod, 2013). Accordingly, if there is ambiguous activation in motor representations of both closely related non-native and native actions, it would interfere with non-native speech perception. Future studies could try giving participants mirror training with one native speech sound and one non-native speech that resemble the native one; this type of sensorimotor training may help establish not only the excitatory links between sensory and motor representations of the same non-native speech action and also the inhibitory links between sensory and motor representations of different non-native and native speech actions. At the same time, sensorimotor training would also need to be accompanied by perception and production training, both of which could help better target and fine-tune each end of the non-native perception-production link.

## **5.6 Conclusion**

In conclusion, results from the current study demonstrate that mirror and counter-mirror training with non-native speech actions modulates automatic imitation of both non-native and native speech actions that are produced using the same articulators. Together with findings from the study reported in Chapter 4, current results support the associative hypothesis that the same associative mechanism is involved in imitation of both novel and familiar actions that were or were not part of the observers' motor repertoire. More importantly, current findings raise concerns about further specification of the associative hypothesis firstly regarding the development of inhibitory control during the initial establishment of the perception-production matching link and secondly regarding the level at which the perception-production link is established.

As discussed in Section 5.5.1, people do not always overtly imitate observed actions but rather modulate the degree to which they imitate so as to maintain high-quality social interaction. Thus, inhibitory control of unwanted motor activation is crucial. Previous automatic imitation studies and sensorimotor training studies, including the ones reported in Chapter 3 and Chapter 4, have only looked at familiar actions that participants have used and encountered throughout their life. Therefore, how the development of inhibitory control interacts with the initial establishment of the perception-production matching link has remained unknown and in fact been ignored to some extent in the automatic imitation literature. Current findings demonstrate robust automatic imitation effects for non-native speech actions that participants at first found difficult to produce and seemingly decreased non-native automatic imitation after either mirror or counter-mirror training. These findings suggest that the imitative capacity emerges before, but not in parallel with, the development of inhibitory

control of imitation, which is consistent with findings from the developmental research (Diamond & Taylor, 1996; Simpson & Riggs, 2011). Further research on imitation of novel actions may help illustrate how people learn to acquire the ability to control their imitative tendency, and automatic imitation of novel actions necessarily provides behavioural manifestation of this learning process.

Interestingly, current results also demonstrate that sensorimotor training with non-native speech actions modulates automatic imitation of native speech actions whose consonants' places of articulation match those of the trained actions. Using the same pair of actions for both distractors and responses, previous studies have mostly examined automatic imitation effects at the individual action level. Leighton & Heyes (2010) discussed in Section 5.5.1 was an exception; they employed both mouth and hand opening/closing movements for distractors and responses and observed an effector compatibility effect which was further modulated by movement compatibility. Additionally, as discussed in Section 5.1.2, Roon & Gafos (2015) observed automatic imitation effects at the feature level (i.e., place of articulation and voicing). In line with Leighton & Heyes and Roon & Gafos, the current study also suggests that observation-induced motor activation may occur at the feature level and further indicates that the perception-production link underlying each phonetic feature may be flexible to change through sensorimotor training. Therefore, the current study not only supports the associative hypothesis of imitation, but also proposes further specification for elucidating at what level observation-induced motor activation occurs and similarly at what level training-induced modification occurs in the perception-production link.

Furthermore, the current finding that sensorimotor training with non-native speech actions modulates automatic imitation of native speech actions demonstrates an example of backward transfer in which training with non-native speech actions influences the perception and/or production of closely related native speech actions (C. B. Chang, 2019). Consistent with previous second language learning research, the current study also indicates a close relationship between native and non-native speech processing and potentially bi-directional influences between the two. Accordingly, based on the associative hypothesis of imitation, it is important to build not only excitatory links between sensory and motor representations of the same non-native speech actions, but also inhibitory links between sensory and motor representations of different non-native and native speech actions. Therefore, the current study suggests a way in which cognitive research on imitation may help further research in the speech domain.

Overall, the current research adds to the work of sensorimotor training studies and second language research in showing that associative sensorimotor training with novel speech actions modulates automatic imitation of both novel and familiar speech actions that are produced using the same articulators. In doing so, the current work contributes to research

questions concerning the role of sensorimotor experience in modulating the imitative mechanism underlying familiar versus novel actions, to accumulating research on the link between speech perception and production systems, and also to the growing body of research on the interaction between native and non-native speech processing.

## Chapter 6 General discussion

### 6.1 Summary of current research findings

The overarching aims of this thesis were 1) to establish the role of sensorimotor experience in modulating the imitative capacity underlying communicative orofacial movements, namely speech actions, that are either familiar or novel to perceivers and 2) to elucidate at what level experience-induced modification occurs in the perception-production link presumably underlying the imitative capacity. To achieve these aims, five experiments were conducted, each of which targeted one or two objectives.

Chapter 3 first established that automatic imitation effects are confounded with, but cannot be reduced to, SSC effects when distractors are associated with both prompts and responses in a speech SRC task (Objective 1 and 2). These findings were in line with the dual-route hypothesis that the perception of speech distractors triggers both perceptual and motor processes and interferes with participants' performance when there is S-S and/or S-R overlap respectively. Critically, as discussed in Section 3.5, even when linguistic prompts are used in a speech SRC task, as in the case of SSC+SRC effects, current findings also suggested the involvement of perception-induced motor activation which, however, is likely to interact with a perceptual process triggered by the same speech distractors. Hence, future studies using speech-related prompts need to consider the possibilities of whether and how their experimental manipulations may modulate the effects of S-S and/or S-R overlap.

Chapter 4 determined that sensorimotor experience is critically involved in establishing the perception-production matching link underlying visually perceived speech actions (Objective 3). Consistent with the associative account of imitation, these findings suggested that the flexibility observed for the observation-execution link underlying manual gestures extends to the perception-production link underlying orofacial movements such as speech. However, compared to manual gestures, automatic imitation of speech was more susceptible to mirror training, but relatively resilient to counter-mirror training. As discussed in Section 4.5, the multimodal and communicative nature of speech actions and the social environment where these actions are commonly used might explain these findings.

Chapter 5 suggested that sensorimotor experience is critically involved in establishing the perception-production matching link underlying novel speech actions (Objective 4) and that training-induced modification occurs at the feature level in the perception-production link (Objective 5). In line with the associative account of imitation, these findings suggested that the associative mechanism is involved in imitation of both familiar and novel actions that were or were not part of the observers' motor repertoire. More importantly, as discussed in Section 5.6, current findings raised concerns about further specification of the associative hypothesis

firstly regarding the development of inhibitory control during the initial establishment of the perception-production matching link and secondly regarding the level at which the perception-production link is established. However, as mentioned earlier, since this study was underpowered due to an insufficient sample size, findings and further implications should be read with care.

In sum, studies reported in this thesis further confirmed that automatic imitation effects measure a motor process triggered by the perception of speech distractors and suggested that the associative mechanism involved in imitation of familiar manual gestures is also involved in imitation of visually opaque orofacial movements and imitation of novel actions that were not part of the observers' motor repertoire.

## **6.2 Limitations, implications, and future directions**

### ***6.2.1 Compatibility effects and speech processing***

#### ***Limitations***

There were limitations of studies reported in Chapter 3. One limitation was that the manual and vocal responses used in Experiments 2 in Chapter 3 were probably not comparable in terms of movement complexity. As discussed in Section 3.3.3, there seemed to be an asymmetry in which SRC effects were more prominent for ER data, whereas SSC effects were more prominent for RT data. This asymmetry between ER and RT data could make comparisons of time courses between different types of compatibility effects more difficult. Future studies could employ relatively more complex manual gestures that relies on coordination by various effectors (e.g., multiple fingers or even wrist and arm movements) and then investigate whether such an asymmetry would still exist for SSC and SRC effects. In addition, eye-tracking analyses for Experiment 2 and 3 had to exclude quite many participants due to data loss. It could be due to massive head movements or even body movements because participants in these two experiments were not restrained from moving around. This flexibility to move was maintained to make sure that participants could make vocal responses in a relatively natural state, but the disadvantage was that the flexibility could cause quite a lot of data loss. Future studies could consider using wearable eye-tracking device to prevent data loss and also maintain participants' flexibility to move.

#### ***Implications and future directions***

As mentioned in Section 3.5, Kornblum et al. have described research on SSC and SRC effects as "fragmented" (1999, p. 688) because previous research has viewed different compatibility effects as unrelated empirical phenomena and examined these effects in isolation. In cognitive

and experimental psychology, SRC effects have been used to investigate the interaction between action and perception, that is, how perceiving a stimulus affects one's behavioural performance (Hommel, Müsseler, Aschersleben, & Prinz, 2001). SRC effects have been commonly examined in the Simon task in which participants' responses are modulated by the relative spatial correspondence between stimulus positions and response key positions (e.g., Simon & Rudell, 1967; Simon & Small, 1969). Compared to the Simon effect, automatic imitation effects observed with manual gestures and speech actions have provided relatively more dynamic and complex examples where action perception interacts with the perceivers' own action performance and thus introduced SRC paradigms into a more social context.

In addition, SSC effects have been used to investigate multisensory interaction during which perceivers have to properly integrate or segregate different incoming signals from the environment (Donohue et al., 2013). SSC effects have been examined in the Stroop task in which participants' manual responses are modulated by the relative semantic correspondence between printed colour words and ink colours of the words (e.g., Redding & Gerjets, 1977). The flanker effect also represents a type of SSC effects, in which participants have to respond to the target letter which is surrounded by either the same (e.g., "HHHHH") or different (e.g., "HSHHH") letters (e.g., Eriksen & Eriksen, 1974). In addition, the verbal version of the Stroop effect represents a kind of SSC+SRC effects in which, instead of pressing a response key, participants have to say ink colours of the presented colour words but to ignore the meaning of the words (e.g., Stroop, 1935). The SSC effects observed in Experiment 2 in Chapter 3 provided another possibility of how this kind of effects could be used to investigate audiovisual speech at the pre-lexical level, a multisensory process that is more often encountered in everyday social life.

Furthermore, findings reported in Chapter 3 suggested that SSC, SRC, and SSC+SRC effects measured using different speech SRC tasks could be explained within the Dimensional Overlap framework. In other words, these findings indicated that different compatibility effects could be examined using one single object, namely speech, which is multimodal by nature. Several dual-route models have been proposed to explain SSC and SRC effects and the interaction between the two (De Jong et al., 1994; Hommel, 1993, 1997; Kornblum et al., 1999). The Dimensional Overlap model is the most general that attempts to account for various compatibility effects in a unified theoretical framework and thus to understand the basic common cognitive processing structure shared by all these effects. Nevertheless, all these models predict interactions between SSC and SRC effects when both are present, which has been confirmed in previous studies combining the Simon task with the Stroop task or with the flanker task (e.g., Hommel, 1997; Kornblum et al., 1999). Importantly, in the tasks where the Simon task is combined with either the Stroop or the flanker task, prompts and responses overlap with different distractor dimensions (e.g., spatial S-R overlap versus semantic S-S

overlap). Hence, it is possible that the processing of different distractor dimensions starts at different time points and so does the automatic activation of the overlapping response and prompt, particularly when it is spatial processing versus higher-level semantic processing. In contrast, in the speech SSC+SRC effects observed in Experiment 2 in Chapter 3, prompts and responses overlapped with the same distractor dimension, which means that the automatic activation of the overlapping prompt and response should presumably occur simultaneously. Thus, SSC+SRC effects measured using speech SRC tasks seem to represent a different processing situation than the one previously examined. Further examination of speech SSC+SRC effects may provide some different insights into information processing and its interaction with perceivers' own action.

## ***6.2.2 The associative account of imitation and its future development***

### ***Limitations***

One major limitation of the studies reported in Chapter 4 and Chapter 5 concerned about the generalisability of current results to different speech actions, since both experiments were based on only a few speech actions. Particularly, in Chapter 5, it seemed that ba responses tended to elicit ER compatibility effects and la responses tended to elicit RT compatibility effects. Yet, these patterns were not observed for the non-native speech pair whose consonants' places of articulation matched those of the native pair. These findings suggested that consonants' manners of articulation might have played a role in causing different patterns of RT and ER compatibility effects for native responses in this study. As discussed in Section 5.5.1, though produced with different manners of articulation, both /B/ and /L/ are produced by narrowing the air path at some place in the vocal tract and then causing the air flow to become turbulent. In contrast, while /l/ is also produced by narrowing the air path at some place in the vocal tract and then causing slight turbulence, /b/ is produced by first completely obstructing the air flow with lips and then generating a quick burst to release air flow. In other words, /b/ is produced much faster than /l/, /B/, or /L/. It could be manner of articulation or the combination of manner and place of articulation that caused different patterns of ER and RT compatibility effects for native responses. As suggested in Section 5.5.1, follow-up studies aiming to tackle this issue could control for both places and manners of articulation to investigate the effect of each phonetic feature in speech SRC tasks in a factorial design study.

A further limitation of studies reported in Chapter 4 and Chapter 5 was that both studies were based on behavioural measures, automatic imitation effects, to indicate change in the perception-production link, while change in automatic imitation effects could result from different sources. As discussed in Section 5.5.1, a larger automatic imitation effect could result from different sources, either greater perception-induced motor activation in compatible

trials or less successful inhibition of unwanted perception-induced motor activation. For manual gestures and bodily movements, both neuroimaging and TMS studies have observed reversed motor activation patterns following counter-mirror training as compared to mirror training (Catmur et al., 2008, 2007), suggesting training-induced change at the neurophysiological level. To further determine that sensorimotor training with speech actions modulates the perception-production link, neuroimaging and brain stimulation studies are needed to demonstrate that sensorimotor training with speech actions leads to change at the neurophysiological level. For instance, as mentioned in Section 2.2.1, Fadiga et al. (2002) found that passively hearing tongue-related speech sounds facilitated tongue MEPs relative to hearing lip-related speech sounds, and it was the other way around for lip MEPs as observed by Nuttall et al. (2016). These studies have exhibited effector-specific sensorimotor mapping at the neurophysiological level for speech actions. Future studies could examine whether this effector-specific motor facilitation during speech perception could be reduced or even eliminated following counter-mirror training and enhanced following mirror training.

Lastly, as mentioned in Section 4.5, findings reported in Chapter 4 and also in Chapter 5 could not completely exclude the possibility of an innate mechanism governing imitative, rather than counter- or non-imitative, sensorimotor associations underlying the imitation of orofacial movements. As argued in Section 4.5, current findings were in line with the main associative hypothesis that it is sensorimotor experience, but not sensory or motor experience alone, that configures observation-execution links (Heyes, 2010); since the only difference between the two groups in both studies was the relationship between observed and executed movements per trial during training. Crucially, however, current findings also suggested that, compared to manual gestures, automatic imitation of speech actions was more resilient to counter-mirror training, and the decrease in automatic imitation following counter-mirror training did not reach significance when collapsed across SOAs. Thus, it seemed that the perception-production matching link underlying speech actions might be somewhat innate and less flexible than the link underlying manual gestures. However, considering the differences between speech actions and manual gestures, future research is needed to examine whether cognitive and social factors play a role in interfering with sensorimotor training. In the following section, I will discuss how cognitive and social factors may interact with sensorimotor training and possible follow-up studies to tackle these issues.

### ***Implications and future directions***

As summarised above, findings reported in Chapter 4 suggested that the associative mechanism involved in imitation of manual gestures is also likely to be involved in imitation of orofacial movements such as speech. More importantly, current findings also suggested

that, compared to manual gestures, automatic imitation of speech was relatively more resilient to counter-mirror training. Several factors that were not discussed in the associative account of imitation were proposed to potentially explain this difference. Though the dominant modality of speech perception is auditory, speech actions are multimodal and visual speech has been shown to have a strong influence on the processing of auditory speech (e.g., McGurk & MacDonald, 1976). Hence, research is needed to further clarify what happens to the speech audiomotor link when there is training-induced modification in the corresponding visuomotor link and how the audiomotor link possibly interferes with visuomotor training effects. In addition, as mentioned in Section 4.4.3, speech actions are relatively complex movements that rely on coordination among different muscle groups and require precise temporal control. Hence, questions then arise as to at which level the training-induced modification occurs in the perception-production link, e.g., individual actions versus specific action dimensions (e.g., phonetic features in speech). Findings reported in Chapter 5 also raised similar concerns and I will further discuss this point below.

Furthermore, as mentioned in Section 4.4.3, Möttönen et al. (2010) observed enhanced motor excitability in hand muscles after participants realised that they were watching sign languages, relative to the condition when they were unaware of the stimuli's communicative nature. Accordingly, Möttönen et al. suggested that, the processing of manual gestures in the motor cortex is not only determined by their physical features, but also affected by other cognitive factors such as its underlying communicative intentions. Syllable articulations employed in studies in this thesis are communicative orofacial movements used in daily conversation, e.g., /ba/ is a word in British English as for "bar". Thus, their underlying communicative intentions might interact with mirror and counter-mirror training effects in the study in Chapter 4. In addition, since speech actions are mostly used during social interactions, the degree of social engagement during sensorimotor training might play a role in modulating training effects. Hence, research is needed to further specify how cognitive and social factors may interfere with or facilitate sensorimotor training effects on the imitative capacity. For example, as suggested in Section 4.4.2, future research could first elucidate whether automatic imitation of non-communicative orofacial movements, such as lip and tongue protrusion, would be smaller than that of syllable articulations and then investigate whether the perception-production matching link underlying non-communicative orofacial movements is more susceptible to sensorimotor training. Likewise for manual gestures, studies could examine whether automatic imitation of manual gestures would increase if participants become aware of their communicative nature and whether the observation-execution link underlying communicative manual gestures is more resilient to counter-mirror training. Moreover, as suggested in Section 4.4.2, the degrees of social engagement may possibly influence sensorimotor training effects on the observation-execution links underlying manual

versus orofacial movements. Future studies could manipulate social factors of the training environment and examine whether and which social manipulations modulate sensorimotor training effects on automatic imitation of different types of actions. On the one hand, further specifications of these factors may explain why certain types of movements are more resilient to counter-mirror training that might undermine its imitative capacity, but not because the imitative capacity underlying these movements are genetically prespecified. On the other hand, further specifications of social and cognitive factors are relevant since the associative account itself also emphasises the role of social factors in contributing to the development of the imitative capacity.

Additionally, as summarised above, findings reported in Chapter 5 suggested that the same associative mechanism involved in imitation of familiar actions is also likely to be involved in imitation of novel actions that people find difficult to perform. More importantly, sensorimotor training effects on automatic imitation of novel actions seemed to interact with increased inhibitory control of unwanted perception-induced motor activation. Recent research has investigated inhibitory control of imitation and its relation to the self-other distinction mechanism that is also involved in other cognitive functions such as the theory of mind and perspective taking (e.g., Hogeveen et al., 2015; Santiesteban et al., 2012). However, how the development of inhibitory control of imitation interacts with the initial establishment of the observation-execution matching link underlying the imitative capacity has remained unclear for previous sensorimotor training studies mainly have used familiar manual gestures. Thus, research is needed to further specify the role of inhibitory control in modulating the development of the imitative capacity, which is not discussed in the associative account of imitation.

Lastly, findings reported in Chapter 5 also suggested that the effects of sensorimotor training with novel actions generalise to automatic imitation of familiar actions that share some similarities with the trained actions. As mentioned above, these findings posed questions regarding at which level the training-induced modification occurs in the perception-production link, e.g., individual actions versus specific action dimensions, which is not specified in the associative account of imitation. As discussed in Section 5.5.2, findings reported in Chapter 5 and findings from Roon & Gafos (2015) suggested that observation-induced motor activation might occur at the feature level and that the perception-production link underlying each phonetic feature might be flexible to change through sensorimotor training. Future studies could further investigate whether sensorimotor training effects also generalise for other phonetic features, such as manners of articulation and voicing patterns. These studies could help elucidate the organisation of speech actions within the motor cortex and also elucidate whether and how the perception-production link is organised at a finer grained feature level.

### ***6.2.3 Speech processing examined in the context of cognitive psychology***

#### ***Limitations***

With respect to speech processing, one limitation of training studies reported in Chapter 4 and Chapter 5 was that no tasks were implemented to examine potential changes in participants' perception and production following either mirror or counter-mirror training. Thus, it remains unknown whether sensorimotor training also modulates participants' perception and production of the trained speech actions, in addition to the imitative capacity. Future research could combine and compare the effects of perception training, production training, and sensorimotor training and employ different tasks that examine change in each modality and change in the link between these two modalities. Another limitation that has been mentioned earlier concerned about the generalisability of current results to different speech actions. As summarised by the Perceptual Assimilation Model discussed in Section 5.5.1, non-native speech actions/sounds could differ from native categories in different ways. Therefore, as suggested in Section 5.5.1, future studies could first establish the degree of assimilability for each non-native speech sound and then examine whether and how it is correlated with the corresponding automatic imitation effect. In addition, studies could also investigate whether the degree of assimilability affect sensorimotor training. For example, if non-native speech actions tend to borrow the perception-production link from closely related native speech actions, non-native actions identified with a higher degree of assimilability would be expected to elicit a larger automatic imitation effect and would be expected to be less susceptible to sensorimotor training.

#### ***Implications and future directions***

Crucially, current findings also have implications for speech research. As discussed in Section 3.5, findings reported in Chapter 3 suggested that SSC, SRC, and SSC+SRC effects could be used to investigate multimodal speech processing. Automatic imitation of speech (i.e., SRC effects) have already been used in various studies to investigate sensorimotor interaction during speech processing (e.g., Adank et al., 2018; Kerzel & Bekkering, 2000). Additionally, SSC effects resulting from the relative correspondence between auditory and visual speech could possibly be used to investigate multisensory speech perception, thus providing a different approach to study audiovisual speech at the pre-lexical level apart from the McGurk effect. Furthermore, SSC+SRC effects that presumably measure both perceptual and motor processes triggered by the same speech distractors may be used to behaviourally investigate the interaction between the ventral stream where speech recognition and comprehension

occurs and the dorsal stream where perceptuomotor translation occurs. As mentioned in Section 6.2.1, compatibility effects have long been used in experimental and cognitive psychology to investigate the interaction between perception and action and also multisensory integration. Therefore, it is not surprising that compatibility effects provide a feasible way to investigate multimodal speech processing as we receive speech stimuli and perform speech actions all the time during social interactions. Also, automatic imitation of speech has provided an example showing how compatibility effects could be used to investigate sensorimotor interaction during speech processing.

Furthermore, as discussed in Section 4.5, both findings reported in Chapter 4 and findings from previous sensorimotor adaptation studies discussed in Section 2.2.3 indicated a certain degree of separation between speech perception and production systems. In other words, it further indicates a certain degree of flexibility in the perception-production link, allowing for mapping between sensory and motor representations of either the same or different speech actions. Moreover, as discussed in Section 5.6, findings reported in Chapter 5 suggested a close relationship between native and non-native speech processing and, together with previous second language (L2) findings, indicated bi-directional influences between the two. Therefore, as proposed in Section 4.5 and 5.6, while both perception and production training are important in refining learners' ability in each modality, sensorimotor training with correctly produced and perceived speech sounds may additionally play a crucial role in linking the two systems and preventing learners from incorrectly associating a non-native sound with a similar native sound. That is, it is important to build not only excitatory links between sensory and motor representations of the same non-native speech actions, but also inhibitory links between sensory and motor representations of different non-native and native speech actions.

As discussed in Section 2.2.2, the Dual-Route model of speech processing proposes that the primary function of the dorsal stream is to support speech motor learning and motor control and that sensory feedback is crucial in guiding the tuning of speech gestures so that the sounds are accurately reproduced (Hickok et al., 2011). Thus, in line with the associative account of imitation, the perception-production matching link underlying speech actions is proposed to be established through sensorimotor experience during which people speak and immediately receive sensory feedback of their own speech. Current findings also suggested that associative sensorimotor experience plays a critical role in modulating the perception-production link underlying speech actions. However, one difference to be highlighted is that, during speaking, people first produce a speech action and then receive sensory feedback. In contrast, most sensorimotor training studies ask participants to produce an action after they perceive an action produced by others; in other words, the sequence of co-occurring motor and sensory events is reversed, as compared to speaking. Accordingly, it remains unclear

whether the perception-production link acquired under these two circumstances are the same. Studies using manual gestures or bodily movements have suggested that associations learned through sensorimotor experience are bidirectional (Elsner & Hommel, 2001; Wiggett et al., 2011). It means that, once associations/links between a sensory and a motor event are learned through sensorimotor experience, activation of one event will lead to activation of the other, no matter what the sequence of these two events is during learning. If the associative mechanism is also involved in establishing the perception-production link underlying speech actions, this link is likely to be bidirectional, meaning that the link acquired from the above two circumstances is likely to be the same. To further determine whether the associative mechanism is involved in imitation of speech actions, studies could reverse the sequence of motor and sensory events during mirror and counter-mirror training and examine whether automatic imitation effects would change following training.

As discussed in Section 2.2.3, to support efficient motor control, models of speech production, such as the DIVA model, have incorporated an internal forward control system that uses *effference copies* of issued motor commands to predict sensory consequences (Hickok et al., 2011; Tian & Poeppel, 2010; Tourville & Guenther, 2011). The associations between motor commands and predicted sensory consequences are learned by linking motor commands with received sensory feedback, and feedforward sensory predictions can then be used for self-monitoring and motor planning. During speaking, the predicted and the actual sensory consequences are compared, and any discrepancy between the two will be used to correct motor commands and update the forward model. In cognitive psychology, the Ideomotor Theory of action control also proposes that the action-perception link (or *shared representations*) is learned by associating an action with its subsequent effects (Greenwald, 1970; Massen & Prinz, 2009). Additionally, because of the bidirectionality of this link, perceiving an action or effects of this action also leads to activation in the corresponding motor commands in the perceiver, which is not explicitly discussed in the DIVA model of speech acquisition and production.

Moreover, the Ideomotor Theory tends to focus on manual or bodily movements that involve interactions with an external object (e.g., grabbing a mug) which can be used as a fixed physical reference to guide actions. It also proposes that an action can be associated with its perceived physical movement characteristics, its environmental effects, or its more abstract action goal/intention. The same action may be used to elicit different environmental effects or to achieve different goals, and thus the learned associations are suggested to be context specific. Accordingly, perceiving another person's action or its consequent effects may activate not only the associated motor representations, but also the context where this association is learned. In contrast, speech actions do not involve interactions with external objects, and the DIVA model proposes that the goal of speech production is to generate a

target sound. Sensorimotor adaptation studies discussed in Section 2.2.3 have observed changes in both perception and production, suggesting that target speech sounds remain rather variable and may be changed as a result of experience. In addition, categorical perception of speech discussed in Section 5.1.2 further suggests that, while being variable, each target sound also seems to cover a range of speech sounds whose acoustic patterns vary. More research is needed to both theoretically and empirically investigate the relation between the context-specificity proposed in the Ideomotor Theory and the target variability suggested in speech science. Comparisons of findings from these two research fields that look at motor control in quite different scenarios may advance each other and contribute to our understanding of how motor control is achieved through the perception-production link and how it is modulated through associative learning.

In addition to motor control, the forward mechanism is also essential in providing performers a sense of agency and allows them to distinguish between sensations resulting from their own actions and those triggered externally by environmental stimuli. A hallmark of the forward mechanism is action-induced sensory attenuation, which refers to the observations that self-generated, and thus predicted, stimuli tend to feel less salient and elicit smaller neurophysiological responses, relative to externally triggered, yet physically identical, stimuli (Hughes, Desantis, & Waszak, 2013). In the speech domain, two separate mechanisms have been proposed in the auditory regions to deal with predicted and unpredicted stimuli received during speaking (E. F. Chang, Niziolek, Knight, Nagarajan, & Houde, 2013). One mechanism is proposed to afford the speaker a sense of agency and allow them to distinguish between self-generated (and therefore well-predicted) speech sounds and sounds from external sources; the other is proposed to contribute to the detection of feedback errors, facilitate the correction of motor commands by signalling the motor areas, and thus help update the internal forward model. The updated forward model could then be used to facilitate the operation of the first mechanism. The predictive coding mechanism is also prevalent in perception in general, and similarly the Opposing Process Theory postulates two processes involved during perception (Press, Kok, & Yon, 2020). Specifically, an early process is proposed to support perception (particularly of noisy or ambiguous input) through predictions based on prior knowledge (i.e., models of the external world); a relatively late process is proposed to detect unpredicted, yet informative, input and possibly facilitate the updating of internal models of the world. It seems that similar predictive mechanisms are at play in different domains and are subject to change through learning. More research is needed to both theoretically and empirically explore the commonalities and differences between predictive mechanisms used in various domains and to investigate whether and how these predictive mechanisms are subject to change through learning, and if so, what type of learning is essential.

### 6.3 Conclusion

The overarching aims of this thesis are to establish the role of sensorimotor experience in modulating the imitative capacity underlying communicative orofacial movements, namely speech actions, that are either familiar or novel to perceivers and to elucidate at what level experience-induced modification occurs in the perception-production link presumably underlying the imitative capacity. Research work presented in this thesis has provided evidence that:

- Automatic imitation effects are confounded with – but cannot be reduced to – SSC effects, when distractors are associated with both prompts and responses in a speech SRC task (Objective 1 and 2). These findings suggest that, in line with the dual-route account of speech processing, the perception of speech distractors triggers both perceptual and motor processes and interferes with performance when distractors are associated with prompts and/or responses respectively. Moreover, these results suggest that, specifically, automatic imitation effects measure a motor process triggered by the perception of speech distractors.
- Sensorimotor training modulates automatic imitation of visually perceived speech actions (Objective 3). Thus, this finding suggests that, in line with the associative account of imitation, the same associative mechanism is involved in imitation of both visually transparent manual gestures and visually opaque orofacial movements.
- Sensorimotor training with novel speech actions modulates automatic imitation of both novel and familiar speech actions produced using the same articulators (Objective 4 and 5). Thus, these findings suggest 1) that, in line with the associative account of imitation, the same associative mechanism is involved in imitation of both familiar and novel actions that were or were not part of the observers' motor repertoire and 2) that training-induced modification occurs at the feature level in the perception-production link.

Overall, the current work contributes to research questions concerning the flexibility of the imitative mechanism underlying various types of movements and to research questions concerning the role of sensorimotor experience in modulating the imitative mechanism underlying familiar versus novel actions. Specifically, the current work not only verifies and extends the associative predictions of imitation to a different effector system, i.e., from manual gestures to speech actions, but also proposed several directions for future research to further develop the associative account of imitation, making it better accommodate to different

effector systems. Furthermore, the current work also contributes to accumulating research on the link between speech perception and production systems and to the growing body of research on the interaction between native and non-native speech processing. Lastly, the current work also demonstrates the possibility of employing various compatibility effects, which have been widely used but often examined separately in cognitive psychology, to investigate multimodal speech processing and therefore also suggests a way in which findings from the speech domain can be linked to findings of more general cognitive processing.

## Bibliography

- Ackermann, H., & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain and Language*, *89*(2), 320–328. [http://doi.org/10.1016/S0093-934X\(03\)00347-X](http://doi.org/10.1016/S0093-934X(03)00347-X)
- Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, *21*(12), 1903–1909. <http://doi.org/10.1177/0956797610389192>
- Adank, P., Nuttall, H., Bekkering, H., & Maegherman, G. (2018). Effects of stimulus response compatibility on covert imitation of vowels. *Attention, Perception, & Psychophysics*, 1–10. <http://doi.org/10.3758/s13414-018-1501-3>
- Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty Years After Hearing Lips and Seeing Voices: the McGurk Effect Revisited. *Multisensory Research*, *31*(1–2), 111–144. <http://doi.org/10.1163/22134808-00002565>
- Bach, P., & Tipper, S. P. (2007). Implicit action encoding influences personal-trait judgments. *Cognition*, *102*(2), 151–178. <http://doi.org/10.1016/j.cognition.2005.11.003>
- Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-guided TMS reveals that the STS is a Cortical Locus of the McGurk Effect. *Journal of Neuroscience*, *30*(7), 2414–2417. <http://doi.org/10.1523/JNEUROSCI.4865-09.2010.fMRI-guided>
- Bertenthal, B. I., Longo, M. R., & Kosobud, A. (2006). Imitative response tendencies following observation of intransitive actions. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 210–225. <http://doi.org/10.1037/0096-1523.32.2.210>
- Best, C. T. (1995). A Direct Realist View of Cross-Language Speech Perception. In W. Strange (Ed.), *Speech perception and language experience: Issues in cross-language research* (pp. 171–204). Baltimore: York Press.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794. <http://doi.org/10.1121/1.1332378>
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360. <http://doi.org/10.1037/0096-1523.14.3.345>
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 13–34). <http://doi.org/10.1075/llt.17.07bes>
- Birmingham, E., & Kingstone, A. (2009). Human social attention (pp. 309–320). [http://doi.org/10.1016/S0079-6123\(09\)17618-5](http://doi.org/10.1016/S0079-6123(09)17618-5)
- Bishop, D. V. M., Brown, B. B., & Robson, J. (1990). The Relationship Between Phoneme Discrimination, Speech Production, and Language Comprehension in Cerebral-Palsied Individuals. *Journal of Speech, Language, and Hearing Research*, *33*(2), 210–219. <http://doi.org/10.1044/jshr.3302.210>

- Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-Temporal Prediction Modulates the Perception of Self-Produced Stimuli. *Journal of Cognitive Neuroscience*, *11*(5), 551–559. <http://doi.org/10.1162/089892999563607>
- Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer. Retrieved from <http://www.praat.org/>
- Bonini, L., Maranesi, M., Livi, A., Fogassi, L., & Rizzolatti, G. (2014). Space-Dependent Representation of Objects and Other's Action in Monkey Ventral Premotor Grasping Neurons. *Journal of Neuroscience*, *34*(11), 4108–4119. <http://doi.org/10.1523/JNEUROSCI.4187-13.2014>
- Bonini, L., Rozzi, S., Serventi, F. U., Simone, L., Ferrari, P. F., & Fogassi, L. (2010). Ventral Premotor and Inferior Parietal Cortices Make Distinct Contribution to Action Organization and Intention Understanding. *Cerebral Cortex*, *20*(6), 1372–1385. <http://doi.org/10.1093/cercor/bhp200>
- Borden, G. J., Harris, K. S., & Raphael, L. J. (1994). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. (J. P. Butler, Ed.) (3rd ed.). Baltimore: Williams & Wilkins.
- Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*, *495*(7441), 327–332. <http://doi.org/10.1038/nature11911>
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English / r / and / l /: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310. <http://doi.org/10.1121/1.418276>
- Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, *106*(1–2), 3–22. [http://doi.org/10.1016/S0001-6918\(00\)00024-X](http://doi.org/10.1016/S0001-6918(00)00024-X)
- Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between Observed and Executed Finger Movements: Comparing Symbolic, Spatial, and Imitative Cues. *Brain and Cognition*, *44*(2), 124–143. <http://doi.org/10.1006/brcg.2000.1225>
- Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2359–2367. <http://doi.org/10.1098/rstb.2009.0066>
- British Society of Audiology. (2011). *Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking*. The British Society of Audiology. Retrieved from [http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA\\_RP\\_PTA\\_FINAL\\_24Sept11\\_MinorAmend06Feb12.pdf](http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA_RP_PTA_FINAL_24Sept11_MinorAmend06Feb12.pdf)
- Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., & Rizzolatti, G. (2004). Neural circuits underlying imitation learning of hand actions: An event-related fMRI study. *Neuron*, *42*(2), 323–334. [http://doi.org/10.1016/S0896-6273\(04\)00181-3](http://doi.org/10.1016/S0896-6273(04)00181-3)
- Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, *25*(5), 663–678. [http://doi.org/10.1207/s15516709cog2505\\_2](http://doi.org/10.1207/s15516709cog2505_2)
- Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and*

- Brain Sciences*, 21(5), 667–684. <http://doi.org/10.1017/S0140525X98001745>
- Callan, D. E., Tsytsarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., & Turner, R. (2006). Song and speech: Brain regions involved with perception and covert production. *NeuroImage*, 31(3), 1327–1342. <http://doi.org/10.1016/j.neuroimage.2006.01.036>
- Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport*, 10(12), 2619–2623. <http://doi.org/10.1097/00001756-199908200-00033>
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, 15(8), 1243–1249. <http://doi.org/10.1093/cercor/bhi007>
- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or Doing? Influence of Visual and Motor Familiarity in Action Observation. *Current Biology*, 16(19), 1905–1910. <http://doi.org/10.1016/j.cub.2006.07.065>
- Campbell, M. E. J., & Cunnington, R. (2017). More than an imitation game: Top-down modulation of the human mirror system. *Neuroscience and Biobehavioral Reviews*, 75, 195–202. <http://doi.org/10.1016/j.neubiorev.2017.01.035>
- Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences*, 100(9), 5497–5502. <http://doi.org/10.1073/pnas.0935845100>
- Casile, A., Caggiano, V., & Ferrari, P. F. (2011). The mirror neuron system: A fresh view. *Neuroscientist*, 17(5), 524–538. <http://doi.org/10.1177/1073858410392239>
- Cassery, E. D., & Pisoni, D. B. (2010). Speech perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(5), 629–647. <http://doi.org/10.1002/wcs.63>
- Catmur, C. (2013). Sensorimotor learning and the ontogeny of the mirror neuron system. *Neuroscience Letters*, 540, 21–27. <http://doi.org/10.1016/j.neulet.2012.10.001>
- Catmur, C., Gillmeister, H., Bird, G., Liepelt, R., Brass, M., & Heyes, C. (2008). Through the looking glass: Counter-mirror activation following incompatible sensorimotor learning. *European Journal of Neuroscience*, 28(6), 1208–1215. <http://doi.org/10.1111/j.1460-9568.2008.06419.x>
- Catmur, C., & Heyes, C. (2011). Time Course Analyses Confirm Independence of Imitative and Spatial Compatibility. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2), 409–421. <http://doi.org/10.1037/a0019325>
- Catmur, C., & Heyes, C. (2017). Mirroring “meaningful” actions: sensorimotor learning modulates imitation of goal-directed actions. *The Quarterly Journal of Experimental Psychology*, 0218(November), 1–38. <http://doi.org/10.1080/17470218.2017.1344257>
- Catmur, C., Walsh, V., & Heyes, C. (2007). Sensorimotor Learning Configures the Human Mirror System. *Current Biology*, 17(17), 1527–1531. <http://doi.org/10.1016/j.cub.2007.08.006>
- Catmur, C., Walsh, V., & Heyes, C. (2009). Associative sequence learning: the role of experience in the development of imitation and the mirror system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1528), 2369–2380. <http://doi.org/10.1098/rstb.2009.0048>

- Cavallo, A., Heyes, C., Becchio, C., Bird, G., & Catmur, C. (2014). Timecourse of mirror and counter-mirror effects measured with transcranial magnetic stimulation. *Social Cognitive and Affective Neuroscience*, 9(8), 1082–1088. <http://doi.org/10.1093/scan/nst085>
- Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40(2), 249–268. <http://doi.org/10.1016/j.wocn.2011.10.007>
- Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, 41(6), 520–533. <http://doi.org/10.1016/j.wocn.2013.09.006>
- Chang, C. B. (2019). Phonetic Drift. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 190–203). Oxford University Press. <http://doi.org/10.1093/oxfordhb/9780198793595.013.16>
- Chang, E. F., Niziolek, C. A., Knight, R., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, 110(7), 2653–2658. <http://doi.org/10.1073/pnas.1216827110>
- Chartier, J., Anumanchipalli, G. K., Johnson, K., & Chang, E. F. (2018). Encoding of Articulatory Kinematic Trajectories in Human Speech Sensorimotor Cortex. *Neuron*, 98(5), 1042–1054.e4. <http://doi.org/10.1016/j.neuron.2018.04.031>
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893–910. <http://doi.org/10.1037/0022-3514.76.6.893>
- Chartrand, T. L., & Lakin, J. L. (2013). The Antecedents and Consequences of Human Behavioral Mimicry. *Annual Review of Psychology*, 64(1), 285–308. <http://doi.org/10.1146/annurev-psych-113011-143754>
- Chartrand, T. L., & van Baaren, R. (2009). Chapter 5 Human Mimicry. In *Advances in Experimental Social Psychology* (pp. 219–274). [http://doi.org/10.1016/S0065-2601\(08\)00405-X](http://doi.org/10.1016/S0065-2601(08)00405-X)
- Chong, T. T. J., Cunnington, R., Williams, M. A., Kanwisher, N., & Mattingley, J. B. (2008). fMRI Adaptation Reveals Mirror Neurons in Human Inferior Parietal Cortex. *Current Biology*, 18(20), 1576–1580. <http://doi.org/10.1016/j.cub.2008.08.068>
- Colenbrander, A. (2002). *Visual standards: aspects and ranges of vision loss with emphasis on population surveys*. Australia.
- Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: From origin to function. *Behavioral and Brain Sciences*, 37(2), 177–192. <http://doi.org/10.1017/S0140525X13000903>
- Cook, R., Press, C., Dickinson, A., & Heyes, C. (2010). Acquisition of automatic imitation is sensitive to sensorimotor contingency. *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 840–852. <http://doi.org/10.1037/a0019256>
- Cook, V. (2003). *Effects of the Second Language on the First*. (V. Cook, Ed.). Multilingual Matters.
- Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., ... Brass, M. (2018). Automatic imitation: A meta-analysis. *Psychological Bulletin*, 144(5), 453–500. <http://doi.org/10.1037/bul0000143>
- Cross, E. S., Hamilton, A. F. d. de C., & Grafton, S. T. (2006). Building a motor simulation de novo:

- Observation of dance by dancers. *NeuroImage*, 31(3), 1257–1267. <http://doi.org/10.1016/j.neuroimage.2006.01.033>
- Cross, E. S., Hamilton, A. F. de C., Kraemer, D. J. M., Kelley, W. M., & Grafton, S. T. (2009). Dissociable substrates for body motion and physical experience in the human action observation network. *European Journal of Neuroscience*, 30(7), 1383–1392. <http://doi.org/10.1111/j.1460-9568.2009.06941.x>
- Cross, E. S., Kraemer, D. J. M., Hamilton, A. F. D. d. C., Kelley, W. M., & Grafton, S. T. (2009). Sensitivity of the Action Observation Network to Physical and Observational Learning. *Cerebral Cortex*, 19(2), 315–326. <http://doi.org/10.1093/cercor/bhn083>
- Darwin, C. (1872). *The expression of the emotions in man and animals*. London: John Murray. <http://doi.org/10.1037/10001-000>
- De Jong, R., Liang, C.-C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology: Human Perception and Performance*, 20(4), 731–750. <http://doi.org/10.1037/0096-1523.20.4.731>
- Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, 64(2–3), 145–173. <http://doi.org/10.1159/000107914>
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91(1), 176–180. <http://doi.org/10.1007/BF00230027>
- Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to ?Do as I say, not as I do? *Developmental Psychobiology*, 29(4), 315–334. [http://doi.org/10.1002/\(SICI\)1098-2302\(199605\)29:4<315::AID-DEV2>3.0.CO;2-T](http://doi.org/10.1002/(SICI)1098-2302(199605)29:4<315::AID-DEV2>3.0.CO;2-T)
- Diehl, R. L., & Kluender, K. R. (1989). On the Objects of Speech Perception. *Ecological Psychology*, 1(2), 121–144. [http://doi.org/10.1207/s15326969eco0102\\_2](http://doi.org/10.1207/s15326969eco0102_2)
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Annual Review of Psychology*, 55(1), 149–179. <http://doi.org/10.1146/annurev.psych.55.090902.142028>
- Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In *Advances in Experimental Social Psychology* (pp. 1–40). Academic Press. [http://doi.org/10.1016/S0065-2601\(01\)80003-4](http://doi.org/10.1016/S0065-2601(01)80003-4)
- Donohue, S. E., Appelbaum, L. G., Park, C. J., Roberts, K. C., & Woldorff, M. G. (2013). Cross-Modal Stimulus Conflict: The Behavioral Effects of Stimulus Input Timing in a Visual-Auditory Stroop Task. *PLoS ONE*, 8(4). <http://doi.org/10.1371/journal.pone.0062802>
- Dushanova, J., & Donoghue, J. (2010). Neurons in primary motor cortex engaged during action observation. *European Journal of Neuroscience*, 31(2), 386–398. <http://doi.org/10.1111/j.1460-9568.2009.07067.x>
- Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 229–240. <http://doi.org/10.1037/0096-1523.27.1.229>

- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. <http://doi.org/10.3758/BF03203267>
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, *15*(2), 399–402. <http://doi.org/10.1046/j.0953-816x.2001.01874.x>
- Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611. <http://doi.org/10.1.1.299.4524>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. <http://doi.org/10.3758/BRM.41.4.1149>
- Ferrari, P. F., Bonini, L., & Fogassi, L. (2009). From monkey mirror neurons to primate behaviours: possible ‘direct’ and ‘indirect’ pathways. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2311–2323. <http://doi.org/10.1098/rstb.2009.0062>
- Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, *17*(8), 1703–1714. <http://doi.org/10.1046/j.1460-9568.2003.02601.x>
- Ferrari, P. F., Gerbella, M., Coudé, G., & Rozzi, S. (2017). Two different mirror neuron networks: The sensorimotor (hand) and limbic (face) pathways. *Neuroscience*, *358*, 300–315. <http://doi.org/10.1016/j.neuroscience.2017.06.052>
- Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, *46*(3), 199–210. <http://doi.org/10.1037/h0062827>
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. [http://doi.org/10.1016/S0095-4470\(19\)30537-6](http://doi.org/10.1016/S0095-4470(19)30537-6)
- Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers’ perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973–2987. <http://doi.org/10.1121/1.428116>
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal Lobe: From Action Organization to Intention Understanding. *Science*, *308*(5722), 662–667. <http://doi.org/10.1126/science.1106138>
- Frith, C. D., & Frith, U. (2012). Mechanisms of Social Cognition. *Annual Review of Psychology*, *63*(1), 287–313. <http://doi.org/10.1146/annurev-psych-120710-100449>
- Galantucci, B., Fowler, C. A., & Goldstein, L. (2009). Perceptuomotor compatibility effects in speech. *Attention, Perception, & Psychophysics*, *71*(5), 1138–1149. <http://doi.org/10.3758/APP.71.5.1138>
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377. <http://doi.org/10.3758/BF03193857>

- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593–609. <http://doi.org/10.1093/brain/119.2.593>
- Gallese, V., Gernsbacher, M. A., Heyes, C., Hickok, G., Iacoboni, M., & Glenberg, A. M. (2011). Mirror neuron forum. *Perspectives on Psychological Science*, *6*(4), 369–407. <http://doi.org/10.1177/1745691611413392>
- Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology*, *4*(June), 340. <http://doi.org/10.3389/fpsyg.2013.00340>
- Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage*, *35*(4), 1674–1684. <http://doi.org/10.1016/j.neuroimage.2007.02.003>
- Gentilucci, M., & Bernardis, P. (2007). Imitation during phoneme production. *Neuropsychologia*, *45*(3), 608–615. <http://doi.org/10.1016/j.neuropsychologia.2006.04.004>
- Gentilucci, M., & Corballis, M. (2006). From manual gesture to speech: A gradual transition. *Neuroscience & Biobehavioral Reviews*, *30*(7), 949–960. <http://doi.org/10.1016/j.neubiorev.2006.02.004>
- Ghaffarvand Mokari, P., Gafos, A., & Williams, D. (2020). Perceptuomotor compatibility effects in vowels: Beyond phonemic identity. *Attention, Perception, & Psychophysics*, *82*(5), 2751–2764. <http://doi.org/10.3758/s13414-020-02014-1>
- Gillmeister, H., Catmur, C., Liepelt, R., Brass, M., & Heyes, C. (2008). Experience-based priming of body parts: A study of action imitation. *Brain Research*, *1217*, 157–170. <http://doi.org/10.1016/j.brainres.2007.12.076>
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R.” *Neuropsychologia*, *9*(3), 317–323. [http://doi.org/10.1016/0028-3932\(71\)90027-3](http://doi.org/10.1016/0028-3932(71)90027-3)
- Gottlieb, G. (1976). The Roles of Experience in the Development of Behavior and the Nervous System (pp. 25–54). <http://doi.org/10.1016/B978-0-12-609303-2.50008-X>
- Graziano, M. S. A. (2016). A New View of the Motor Cortex and Its Relation to Social Behavior. In S. S. Obhi & E. S. Cross (Eds.), *Shared Representations* (pp. 38–58). Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9781107279353.004>
- Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review*, *77*(2), 73–99. <http://doi.org/10.1037/h0028689>
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. <http://doi.org/10.1016/j.tics.2005.11.006>
- Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *The Journal of the Acoustical Society of America*, *105*(5), 2854–2865. <http://doi.org/10.1121/1.426900>
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280–301.

- <http://doi.org/10.1016/j.bandl.2005.06.001>
- Hallett, M. (2007). Transcranial Magnetic Stimulation: A Primer. *Neuron*, 55(2), 187–199. <http://doi.org/10.1016/j.neuron.2007.06.026>
- Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., & Ceballos-Baumann, A. O. (2005). Transmodal Sensorimotor Networks during Action Observation in Professional Pianists. *Journal of Cognitive Neuroscience*, 17(2), 282–293. <http://doi.org/10.1162/0898929053124893>
- Hattori, K. (2010). *Perception and production of English /r-/l/ by adult Japanese speakers*. University College London.
- Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, 5(6), 253–261. [http://doi.org/10.1016/S1364-6613\(00\)01661-2](http://doi.org/10.1016/S1364-6613(00)01661-2)
- Heyes, C. (2005). Imitation by Association. In S. Hurley & N. Chater (Eds.), *Perspectives on Imitation: From Neuroscience to Social Science: Volume 1 Mechanisms of Imitation and Imitation in Animals* (pp. 157–176). MIT Press.
- Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, 34(4), 575–583. <http://doi.org/10.1016/j.neubiorev.2009.11.007>
- Heyes, C. (2011). Automatic imitation. *Psychological Bulletin*, 137(3), 463–483. <http://doi.org/10.1037/a0022288>
- Heyes, C. (2013). What can imitation do for cooperation? In K. Sterelny, R. Joyce, B. Calcott, & B. Fraser (Eds.), *Life and mind: Philosophical issues in biology and psychology. Cooperation and its evolution* (pp. 313–331). MIT Press. <http://doi.org/10.1093/icb/icu011>
- Heyes, C. (2016). Homo imitans? Seven reasons why imitation couldn't possibly be associative. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1686), 20150069. <http://doi.org/10.1098/rstb.2015.0069>
- Heyes, C. (2018). *Cognitive gadgets: the cultural evolution of thinking*. Harvard University Press.
- Heyes, C., Bird, G., Johnson, H., & Haggard, P. (2005). Experience modulates automatic imitation. *Cognitive Brain Research*, 22(2), 233–240. <http://doi.org/10.1016/j.cogbrainres.2004.09.009>
- Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, 29(1), 2–20. <http://doi.org/10.1080/01690965.2013.834370>
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–Motor Interaction Revealed by fMRI: Speech, Music, and Working Memory in Area Spt. *Journal of Cognitive Neuroscience*, 15(5), 673–682. <http://doi.org/10.1162/jocn.2003.15.5.673>
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*, 69(3), 407–422. <http://doi.org/10.1016/j.neuron.2011.01.019>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, 8(5), 393–402. <http://doi.org/10.1038/nrn2113>
- Hisanaga, S., Sekiyama, K., Igasaki, T., & Murayama, N. (2016). Language/Culture Modulates Brain and Gaze Processes in Audiovisual Speech Perception. *Scientific Reports*, 6(1), 35265. <http://doi.org/10.1038/srep35265>

- Hogeveen, J., Obhi, S. S., Banissy, M. J., Santiesteban, I., Press, C., Catmur, C., & Bird, G. (2015). Task-dependent and distinct roles of the temporoparietal junction and inferior frontal cortex in the control of imitation. *Social Cognitive and Affective Neuroscience*, *10*(7), 1003–1009. <http://doi.org/10.1093/scan/nsu148>
- Holt, L. L., & Lotto, A. J. (2008). Speech Perception Within an Auditory Cognitive Science Framework. *Current Directions in Psychological Science*, *17*(1), 42–46. <http://doi.org/10.1111/j.1467-8721.2008.00545.x>
- Hommel, B. (1993). The relationship between stimulus processing and response selection in the Simon task: Evidence for a temporal overlap. *Psychological Research*, *55*(4), 280–290. <http://doi.org/10.1007/BF00419688>
- Hommel, B. (1997). Interactions between stimulus-stimulus congruence and stimulus-response compatibility. *Psychological Research*, *59*(4), 248–260. <http://doi.org/10.1007/BF00439302>
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878. <http://doi.org/10.1017/S0140525X01000103>
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213–1216. <http://doi.org/10.1126/science.279.5354.1213>
- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor Adaptation of Speech I. *Journal of Speech, Language, and Hearing Research*, *45*(2), 295–310. [http://doi.org/10.1044/1092-4388\(2002\)023](http://doi.org/10.1044/1092-4388(2002)023)
- Houde, J. F., & Nagarajan, S. S. (2011). Speech Production as State Feedback Control. *Frontiers in Human Neuroscience*, *5*. <http://doi.org/10.3389/fnhum.2011.00082>
- Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta Burst Stimulation of the Human Motor Cortex. *Neuron*, *45*(2), 201–206. <http://doi.org/10.1016/j.neuron.2004.12.033>
- Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin*, *139*(1), 133–151. <http://doi.org/10.1037/a0028566>
- Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical Mechanisms of Human Imitation. *Science*, *286*(5449), 2526–2528. <http://doi.org/10.1126/science.286.5449.2526>
- Jansson, E., Wilson, A. D., Williams, J. H. G., & Mon-Williams, M. (2007). Methodological problems undermine tests of the ideo-motor conjecture. *Experimental Brain Research*, *182*(4), 549–558. <http://doi.org/10.1007/s00221-007-1013-1>
- Jarick, M., & Jones, J. A. (2009). Effects of seeing and hearing speech on speech production: A response time study. *Experimental Brain Research*, *195*(2), 175–182. <http://doi.org/10.1007/s00221-009-1765-x>
- Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, *108*(3), 1246. <http://doi.org/10.1121/1.1288414>
- Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: studies of

- Mandarin tone production. *Journal of Phonetics*, 30(3), 303–320.  
<http://doi.org/10.1006/jpho.2001.0160>
- Jones, J. A., & Munhall, K. G. (2005). Remapping Auditory-Motor Representations in Voice Production. *Current Biology*, 15(19), 1768–1772. <http://doi.org/10.1016/j.cub.2005.08.063>
- Kartushina, N., Frauenfelder, U. H., & Golestani, N. (2016). How and When Does the Second Language Influence the Production of Native Speech Sounds: A Literature Review. *Language Learning*, 66(S2), 155–186. <http://doi.org/10.1111/lang.12187>
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, 138(2), 817–832. <http://doi.org/10.1121/1.4926561>
- Kerzel, D. (2002). Evidence for effects of phonological correspondence between visible speech and written syllables. *Psychological Research*, 66(3), 195–200. <http://doi.org/10.1007/s00426-002-0088-3>
- Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus response compatibility. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2), 634–647. <http://doi.org/10.1037/0096-1523.26.2.634>
- Keysers, C., Kohler, E., Umiltà, M. A., Nanetti, L., Fogassi, L., & Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research*, 153(4), 628–636. <http://doi.org/10.1007/s00221-003-1603-5>
- Kilner, J. M., Neal, A., Weiskopf, N. W., Friston, K. J., & Frith, C. D. (2009). Evidence of Mirror Neurons in Human Inferior Frontal Gyrus. *Journal of Neuroscience*, 29(32), 10153–10159. <http://doi.org/10.1523/JNEUROSCI.2668-09.2009>
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science*, 297(5582), 846–848. <http://doi.org/10.1126/science.1070311>
- Kokkinaki, T., & Kugiumutzakis, G. (2000). Basic aspects of vocal imitation in infant-parent interaction during the first 6 months. *Journal of Reproductive and Infant Psychology*, 18(3), 173–187. <http://doi.org/10.1080/713683042>
- Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of Stroop- and Simon-like stimuli. *Psychological Research*, 56(3), 130–135. <http://doi.org/10.1007/BF00419699>
- Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility--a model and taxonomy. *Psychological Review*, 97(2), 253–270. <http://doi.org/10.1037/0033-295X.97.2.253>
- Kornblum, S., Stevens, G. T., Whipple, A., & Requin, J. (1999). The effects of irrelevant stimuli: 1. The time course of stimulus–stimulus and stimulus–response consistency effects with Stroop-like stimuli, Simon-like tasks, and their factorial combinations. *Journal of Experimental Psychology: Human Perception and Performance*, 25(3), 688–714. <http://doi.org/10.1037/0096-1523.25.3.688>

- Krekelberg, B., Boynton, G. M., & van Wezel, R. J. A. (2006). Adaptation: from single cells to BOLD signals. *Trends in Neurosciences*, *29*(5), 250–256. <http://doi.org/10.1016/j.tins.2006.02.008>
- Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory Preference in Speech Production Revealed by Simultaneous Alteration of Auditory and Somatosensory Feedback. *Journal of Neuroscience*, *32*(27), 9351–9358. <http://doi.org/10.1523/JNEUROSCI.0404-12.2012>
- Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., & Ostry, D. J. (2014). Plasticity in the Human Speech Motor System Drives Changes in Speech Perception. *Journal of Neuroscience*, *34*(31), 10339–10346. <http://doi.org/10.1523/JNEUROSCI.0108-14.2014>
- Leighton, J., & Heyes, C. (2010). Hand to mouth: automatic imitation across effector systems. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1174–1183. <http://doi.org/10.1037/a0019953>
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. <http://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. <http://doi.org/10.1037/h0044417>
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. [http://doi.org/10.1016/0010-0277\(85\)90021-6](http://doi.org/10.1016/0010-0277(85)90021-6)
- Lingnau, A., Gesierich, B., & Caramazza, A. (2009). Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proceedings of the National Academy of Sciences*, *106*(24), 9925–9930. <http://doi.org/10.1073/pnas.0902262106>
- Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD*, *20*(3), 384–422. <http://doi.org/10.1080/00437956.1964.11659830>
- Longo, M. R., Kosobud, A., & Bertenthal, B. I. (2008). Automatic imitation of biomechanically possible and impossible actions: Effects of priming movements versus goals. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 489–501. <http://doi.org/10.1037/0096-1523.34.2.489>
- Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *The Journal of the Acoustical Society of America*, *102*(2), 1134–1140. <http://doi.org/10.1121/1.419865>
- Massaro, D. W. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. MIT Press.
- Massen, C., & Prinz, W. (2009). Movements, actions and tool-use actions: an ideomotor approach to imitation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2349–2358. <http://doi.org/10.1098/rstb.2009.0059>
- Masur, E. F., & Rodemaker, J. E. (1999). Mothers' and infants' spontaneous vocal, verbal, and action imitation during the second year. *Merrill-Palmer Quarterly*, *45*(3), 392–412. Retrieved from <http://www.jstor.org/stable/23092579>
- Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect, and fMRI. *Journal of*

- Cognitive Neuroscience*, 26(3), 606–620. [http://doi.org/10.1162/jocn\\_a\\_00515](http://doi.org/10.1162/jocn_a_00515)
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748. <http://doi.org/10.1038/264746a0>
- Mclaren, I. P. L., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior*, 30(3), 177–200. <http://doi.org/10.3758/BF03192828>
- Meltzoff, A. N. (2002). Elements of a developmental theory of imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The Imitative Mind Development, Evolution and Brain Bases* (pp. 19–41). Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9780511489969.002>
- Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1431), 491–500. <http://doi.org/10.1098/rstb.2002.1261>
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science*, 198(4312), 75–78. <http://doi.org/10.1126/science.198.4312.75>
- Meltzoff, A. N., & Moore, M. K. (1983). Newborn Infants Imitate Adult Facial Gestures. *Child Development*, 54(3), 702. <http://doi.org/10.2307/1130058>
- Meltzoff, A. N., & Moore, M. K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, 25(6), 954–962. <http://doi.org/10.1037/0012-1649.25.6.954>
- Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: a theoretical model. *Early Development and Parenting*, 6(3–4), 179–192. [http://doi.org/10.1002/\(SICI\)1099-0917\(199709/12\)6:3/4<179::AID-EDP157>3.0.CO;2-R](http://doi.org/10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R)
- Ménard, L. (2015). Multimodal Speech Production. In *The Handbook of Speech Production* (pp. 200–221). Hoboken, NJ, NJ: John Wiley & Sons, Inc. <http://doi.org/10.1002/9781118584156.ch10>
- Mengotti, P., Ticini, L. F., Waszak, F., Schütz-Bosbach, S., & Rumiati, R. I. (2013). Imitating others' actions: transcranial magnetic stimulation of the parietal opercula reveals the processes underlying automatic imitation. *European Journal of Neuroscience*, 37(2), 316–322. <http://doi.org/10.1111/ejn.12019>
- Mitsuya, T., MacDonald, E. N., & Munhall, K. G. (2014). Temporal control and compensation for perturbed voicing feedback. *The Journal of the Acoustical Society of America*, 135(5), 2986–2994. <http://doi.org/10.1121/1.4871359>
- Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 18(5), 331–340. <http://doi.org/10.3758/BF03211209>
- Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. *Neuroscience and Biobehavioral Reviews*, 36(1), 341–349. <http://doi.org/10.1016/j.neubiorev.2011.07.004>
- Möttönen, R., Farmer, H., & Watkins, K. E. (2010). Lateralization of motor excitability during observation of bimanual signs. *Neuropsychologia*, 48(10), 3173–3177.

<http://doi.org/10.1016/j.neuropsychologia.2010.06.033>

- Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-Neuron Responses in Humans during Execution and Observation of Actions. *Current Biology*, *20*(8), 750–756. <http://doi.org/10.1016/j.cub.2010.02.045>
- Murakami, T., Abe, M., Wiratman, W., Fujiwara, J., Okamoto, M., Mizuochi-Endo, T., ... Ugawa, Y. (2018). The Motor Network Reduces Multisensory Illusory Perception. *The Journal of Neuroscience*, *38*(45), 9679–9688. <http://doi.org/10.1523/JNEUROSCI.3650-17.2018>
- Murakami, T., Restle, J., & Ziemann, U. (2011). Observation–execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. *Neuropsychologia*, *49*(7), 2045–2054. <http://doi.org/10.1016/j.neuropsychologia.2011.03.034>
- Nasir, S. M., & Ostry, D. J. (2006). Somatosensory Precision in Speech Production. *Current Biology*, *16*(19), 1918–1923. <http://doi.org/10.1016/j.cub.2006.07.069>
- Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences*, *106*(48), 20470–20475. <http://doi.org/10.1073/pnas.0907032106>
- Newman-Norlund, R. D., Ondobaka, S., Van Schie, H. T., van Elswijk, G., & Bekkering, H. (2010). Virtual lesions of the IFG abolish response facilitation for biological and non-biological cues. *Frontiers in Behavioral Neuroscience*, *4*(5). <http://doi.org/10.3389/neuro.08.005.2010>
- Nuttall, H. E., Kennedy-Higgins, D., Hogan, J., Devlin, J. T., & Adank, P. (2016). The effect of speech distortion on the excitability of articulatory motor cortex. *NeuroImage*, *128*, 218–226. <http://doi.org/10.1016/j.neuroimage.2015.12.038>
- O’Toole, R., & Dubin, R. (1968). Baby feeding and body sway: An experiment in George Herbert Mead’s “taking the role of the other.” *Journal of Personality and Social Psychology*, *10*(1), 59–65. <http://doi.org/10.1037/h0026387>
- Okada, K., & Hickok, G. (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language*, *98*(1), 112–117. <http://doi.org/10.1016/j.bandl.2006.04.006>
- Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience Letters*, *452*(3), 219–223. <http://doi.org/10.1016/j.neulet.2009.01.060>
- Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., ... Slaughter, V. (2016). Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology*, *26*(10), 1334–1338. <http://doi.org/10.1016/j.cub.2016.03.047>
- Pa, J., & Hickok, G. (2008). A parietal–temporal sensory–motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. *Neuropsychologia*, *46*(1), 362–368. <http://doi.org/10.1016/j.neuropsychologia.2007.06.024>
- Pani, P., Theys, T., Romero, M. C., & Janssen, P. (2014). Grasping Execution and Grasping Observation Activity of Single Neurons in the Macaque Anterior Intraparietal Area. *Journal of Cognitive Neuroscience*, *26*(10), 2342–2355. [http://doi.org/10.1162/jocn\\_a\\_00647](http://doi.org/10.1162/jocn_a_00647)

- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. <http://doi.org/10.1121/1.2178720>
- Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, *94*(1), 61–73. <http://doi.org/10.1037/0033-295X.94.1.61>
- Peperkamp, S., & Bouchon, C. (2011). The Relation Between Perception and Production in L2 Phonological Processing. In *INTERSPEECH-2011* (pp. 161–164). Florence, Italy, Italy.
- Perry, A., Stiso, J., Chang, E. F., Lin, J. J., Parvizi, J., & Knight, R. T. (2017). Mirroring in the Human Brain: Deciphering the Spatial-Temporal Patterns of the Human Mirror Neuron System. *Cerebral Cortex*, 1–10. <http://doi.org/10.1093/cercor/bhx013>
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347. <http://doi.org/10.1017/S0140525X12001495>
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*(2), 191–215. <http://doi.org/10.1006/jpho.2001.0134>
- Press, C., Bird, G., Flach, R., & Heyes, C. (2005). Robotic movement elicits automatic imitation. *Cognitive Brain Research*, *25*(3), 632–640. <http://doi.org/10.1016/j.cogbrainres.2005.08.020>
- Press, C., Gillmeister, H., & Heyes, C. (2007). Sensorimotor experience enhances automatic imitation of robotic action. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1625), 2509–2514. <http://doi.org/10.1098/rspb.2007.0774>
- Press, C., Kok, P., & Yon, D. (2020). The Perceptual Prediction Paradox. *Trends in Cognitive Sciences*, *24*(1), 13–24. <http://doi.org/10.1016/j.tics.2019.11.003>
- Press, C., Weiskopf, N., & Kilner, J. M. (2012). Dissociable roles of human inferior frontal gyrus during action execution and observation. *NeuroImage*, *60*(3), 1671–1677. <http://doi.org/10.1016/j.neuroimage.2012.01.118>
- Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, M. F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(20), 7865–7870. <http://doi.org/10.1073/pnas.0509989103>
- Purcell, D. W., & Munhall, K. G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966–977. <http://doi.org/10.1121/1.2217714>
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. <http://doi.org/10.1038/nn.2331>
- Ray, E., & Heyes, C. (2011). Imitation in infancy: The wealth of the stimulus. *Developmental Science*, *14*(1), 92–105. <http://doi.org/10.1111/j.1467-7687.2010.00961.x>
- Redding, G. M., & Gerjets, D. A. (1977). Stroop Effect: Interference and Facilitation with Verbal and Manual Responses. *Perceptual and Motor Skills*, *45*(1), 11–17. <http://doi.org/10.2466/pms.1977.45.1.11>
- Restle, J., Murakami, T., & Ziemann, U. (2012). Facilitation of speech repetition accuracy by theta burst stimulation of the left posterior inferior frontal gyrus. *Neuropsychologia*, *50*(8), 2026–2031. <http://doi.org/10.1016/j.neuropsychologia.2012.05.001>

- Riecker, A., Ackermann, H., Wildgruber, D., Meyer, J., Dogil, G., Haider, H., & Grodd, W. (2000). Articulatory/Phonetic Sequencing at the Level of the Anterior Perisylvian Cortex: A Functional Magnetic Resonance Imaging (fMRI) Study. *Brain and Language*, 75(2), 259–276. <http://doi.org/10.1006/brln.2000.2356>
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5), 188–194. [http://doi.org/10.1016/S0166-2236\(98\)01260-0](http://doi.org/10.1016/S0166-2236(98)01260-0)
- Rizzolatti, G., & Craighero, L. (2004). THE MIRROR-NEURON SYSTEM. *Annual Review of Neuroscience*, 27(1), 169–192. <http://doi.org/10.1146/annurev.neuro.27.070203.144230>
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2), 131–141. [http://doi.org/10.1016/0926-6410\(95\)00038-0](http://doi.org/10.1016/0926-6410(95)00038-0)
- Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, 11(4), 264–274. <http://doi.org/10.1038/nrn2805>
- Roon, K. D., & Gafos, A. I. (2015). Perceptuo-motor effects of response-distractor compatibility in speech: beyond phonemic identity. *Psychonomic Bulletin & Review*, 22(1), 242–250. <http://doi.org/10.3758/s13423-014-0666-6>
- Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, 24(2), 113–121. [http://doi.org/10.1016/S0166-2236\(00\)01685-4](http://doi.org/10.1016/S0166-2236(00)01685-4)
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, 23(2–3), 429–435. <http://doi.org/10.1016/j.cogbrainres.2004.11.006>
- Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4), 421–436. <http://doi.org/10.1006/jpho.1997.0051>
- Santiesteban, I., White, S., Cook, J., Gilbert, S. J., Heyes, C., & Bird, G. (2012). Training social cognition: From imitation to Theory of Mind. *Cognition*, 122(2), 228–235. <http://doi.org/10.1016/j.cognition.2011.11.004>
- Schmitz, J., Bartoli, E., Maffongelli, L., Fadiga, L., Sebastian-Galles, N., & D’Ausilio, A. (2018). Motor cortex compensates for lack of sensory and motor experience during auditory speech perception. *Neuropsychologia*, (July 2017), 1–7. <http://doi.org/10.1016/j.neuropsychologia.2018.01.006>
- Schuerman, W. L., Nagarajan, S., McQueen, J. M., & Houde, J. (2017). Sensorimotor adaptation affects perceptual compensation for coarticulation. *The Journal of the Acoustical Society of America*, 141(4), 2693–2704. <http://doi.org/10.1121/1.4979791>
- Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America*, 90(4), 1797–1805. <http://doi.org/10.1121/1.401660>
- Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, 21(4), 427–444. [http://doi.org/10.1016/S0095-4470\(19\)30229-3](http://doi.org/10.1016/S0095-4470(19)30229-3)

- Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, *125*(2), 1103–1113. <http://doi.org/10.1121/1.3058638>
- Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*(3), 422–429. <http://doi.org/10.3758/BF03194890>
- Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300–304. <http://doi.org/10.1037/h0020586>
- Simon, J. R., & Small, A. M. (1969). Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology*, *53*(5), 433–435. <http://doi.org/10.1037/h0028034>
- Simpson, A., & Riggs, K. J. (2011). Under what conditions do children have difficulty in inhibiting imitation? Evidence for the importance of planning specific responses. *Journal of Experimental Child Psychology*, *109*(4), 512–524. <http://doi.org/10.1016/j.jecp.2011.02.015>
- Singer, T., Seymour, B., O’Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for Pain Involves the Affective but not Sensory Components of Pain. *Science*, *303*(5661), 1157–1162. <http://doi.org/10.1126/science.1093535>
- Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing Lips and Seeing Voices: How Cortical Areas Supporting Speech Production Mediate Audiovisual Speech Perception. *Cerebral Cortex*, *17*(10), 2387–2399. <http://doi.org/10.1093/cercor/bhl147>
- Sowden, S., & Catmur, C. (2015). The role of the right temporoparietal junction in the control of imitation. *Cerebral Cortex*, *25*(4), 1107–1113. <http://doi.org/10.1093/cercor/bht306>
- Strafella, A. P., & Paus, T. (2000). Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. *Neuroreport*, *11*(10), 2289–2292. <http://doi.org/10.1097/00001756-200007140-00044>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. <http://doi.org/10.1037/h0054651>
- Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: a study of imitation. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1746–1759. <http://doi.org/10.1037/0096-1523.26.6.1746>
- Sumbly, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. <http://doi.org/10.1121/1.1907309>
- Swaminathan, S., MacSweeney, M., Boyles, R., Waters, D., Watkins, K. E., & Möttönen, R. (2013). Motor excitability during visual perception of known and unknown spoken languages. *Brain and Language*, *126*(1), 1–7. <http://doi.org/10.1016/j.bandl.2013.03.002>
- Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, *1*(OCT), 1–23. <http://doi.org/10.3389/fpsyg.2010.00166>
- Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology*, *5*. <http://doi.org/10.3389/fpsyg.2014.00725>
- Tkach, D., Reimer, J., & Hatsopoulos, N. G. (2007). Congruent Activity during Action and Action

- Observation in Motor Cortex. *Journal of Neuroscience*, 27(48), 13241–13250.  
<http://doi.org/10.1523/JNEUROSCI.2895-07.2007>
- Tomasello, M. (1996). Do Apes Ape? In *Social Learning in Animals* (pp. 319–346). Elsevier.  
<http://doi.org/10.1016/B978-012273965-1/50016-9>
- Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981.  
<http://doi.org/10.1080/01690960903498424>
- Tremblay, C., Champoux, F., Bacon, B. A., & Theoret, H. (2007). Evidence for a Generic Process Underlying Multisensory Integration. *The Open Behavioral Science Journal*, 1(1), 1–4.  
<http://doi.org/10.2174/1874230000701010001>
- Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, 423(6942), 866–869. <http://doi.org/10.1038/nature01710>
- van Schie, H. T., van Waterschoot, B. M., & Bekkering, H. (2008). Understanding action beyond imitation: Reversed compatibility effects of action observation in imitation and joint action. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6), 1493–1500.  
<http://doi.org/10.1037/a0011750>
- Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics*, 60(6), 926–940.  
<http://doi.org/10.3758/BF03211929>
- Venezia, J. H., Fillmore, P., Matchin, W., Lisette Isenberg, A., Hickok, G., & Fridriksson, J. (2016). Perception drives production across sensory modalities: A network for sensorimotor integration of visual speech. *NeuroImage*, 126, 196–207. <http://doi.org/10.1016/j.neuroimage.2015.11.038>
- Vigneswaran, G., Philipp, R., Lemon, R. N., & Kraskov, A. (2013). M1 Corticospinal Mirror Neurons and Their Role in Movement Suppression during Action Observation. *Current Biology*, 23(3), 236–243. <http://doi.org/10.1016/j.cub.2012.12.006>
- Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, 122(4), 2306–2319. <http://doi.org/10.1121/1.2773966>
- Vogt, S., Buccino, G., Wohlschläger, A. M., Canessa, N., Shah, N. J., Zilles, K., ... Fink, G. R. (2007). Prefrontal involvement in imitation learning of hand actions: Effects of practice and expertise. *NeuroImage*, 37(4), 1371–1383. <http://doi.org/10.1016/j.neuroimage.2007.07.005>
- Wang, Y., & Hamilton, A. F. de C. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. *Frontiers in Human Neuroscience*, 6(June), 1–10.  
<http://doi.org/10.3389/fnhum.2012.00153>
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8), 989–994.  
[http://doi.org/10.1016/S0028-3932\(02\)00316-0](http://doi.org/10.1016/S0028-3932(02)00316-0)
- Wiggett, A. J., Hudson, M., Clifford, A., Tipper, S. P., & Downing, P. E. (2012). Doing, seeing, or both: Effects of learning condition on subsequent action perception. *Social Neuroscience*, 7(6), 606–621. <http://doi.org/10.1080/17470919.2012.686926>

- Wiggett, A. J., Hudson, M., Tipper, S. P., & Downing, P. E. (2011). Learning associations between action and perception: Effects of incompatible training on body part and spatial priming. *Brain and Cognition*, 76(1), 87–96. <http://doi.org/10.1016/j.bandc.2011.02.014>
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701–702. <http://doi.org/10.1038/nn1263>
- Wu, Y., Evans, B. G., & Adank, P. (2019). Sensorimotor training modulates automatic imitation of visual speech. *Psychonomic Bulletin & Review*. <http://doi.org/10.3758/s13423-019-01623-8>
- Yuval-Greenberg, S., & Deouell, L. Y. (2009). The dog's meow: asymmetrical interaction in cross-modal object recognition. *Experimental Brain Research*, 193(4), 603–614. <http://doi.org/10.1007/s00221-008-1664-6>

## Appendices

### Appendix A Experiment 3

The aim of Experiment 3 was to replicate findings from Experiment 2 reported in Chapter 3, but with participants instructed to look at the speaker's mouth, instead of face, while completing each task. If different compatibility effect patterns across tasks were due to the experimental manipulation of S-R and/or S-S overlap, rather than gaze behaviours, similar results were expected between Experiment 2 and 3.

#### *Method*

##### *Participants*

Twenty-eight participants were recruited following the same criteria used in Experiment 2, but four were excluded for having no eye-tracking data collected, which could be because they were not looking at the screen at all. Hence, the final analysis included 24 participants (17 female;  $M_{\text{age}} = 21.00$  years,  $SD_{\text{age}} = 3.93$ ,  $\text{range}_{\text{age}} = 18 - 31$ ). Six participants were further excluded from eye-tracking analysis for their eye-tracking data failed to meet the selection requirements. Hence, the final eye-tracking analysis included 18 participants (12 female;  $M_{\text{age}} = 20.94$  years,  $SD_{\text{age}} = 3.72$ ,  $\text{range}_{\text{age}} = 18 - 31$ ). Participants received £15 or course credit. The University Research Ethics Committee approved the procedures, and all participants gave written informed consent.

##### *Material and stimuli*

Material and stimuli were identical to those in Experiment 2.

##### *Design and procedure*

The design and procedure were identical to those in Experiment 2, except that participants were instructed to look at the speaker's mouth but to ignore what she was saying in distractor videos.

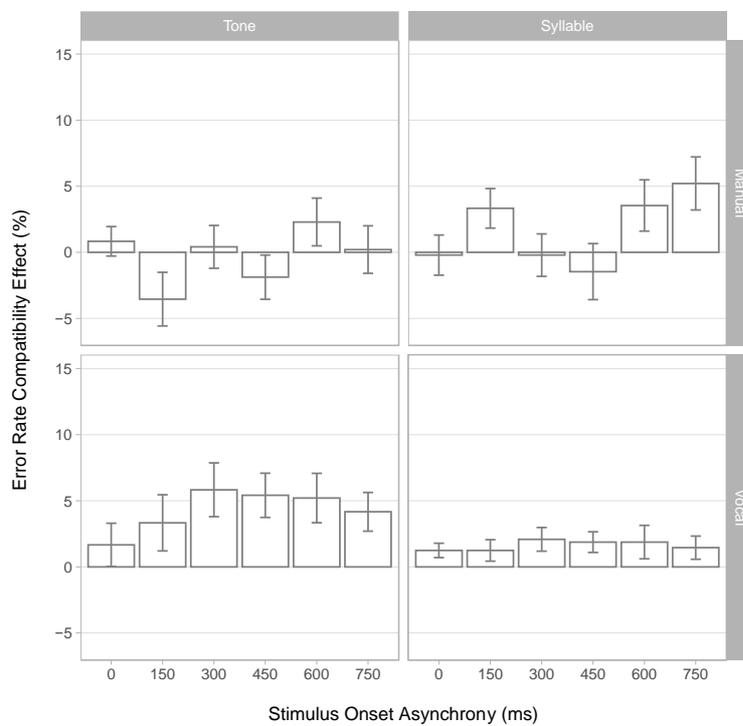
##### *Data recording and analysis*

Technical equipment used to record responses, and data processing methods were identical to those in Experiment 2. The eye-tracking equipment and data processing methods were also identical to those in Experiment 2. Trials with more than 33.33% (i.e., 72 samples) data loss

were first excluded (percentage of trials included per participant per task:  $M = 64.5$ ,  $SD = 30.7$ ) from eye-tracking data. Then, six participants with less than 30 trials in any task condition were excluded from eye-tracking analysis. However, all participants were included in behavioural analysis for eye-tracking results were not of primary interest in the current study but were used to examine whether participants had followed the instructions. Therefore, behavioural and eye-tracking data were analysed separately. Planned statistical analyses were identical to those used in Experiment 2.

## Results

### Error rate analysis



**Figure A.1** ER compatibility effects  $\pm$  SEs in each experimental condition. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where SSC+SRC effects were expected.

**Table A.1** Three-way ANOVA summary for ER compatibility effects, as a function of prompt type, response modality, and SOA.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	$Epsilon$	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Prompt	1	23	1.00	15.67	897.87	0.40	.533	.017
Response	1	23	1.00	722.27	2903.78	5.72	.025	.199
SOA	5	115	0.79	480.43	5583.12	1.98	.087	.079
Prompt x Response	1	23	1.00	767.75	2341.62	7.54	.012	.247
Prompt x SOA	3.57	82.17	0.72	365.84	6164.37	1.37	.256	.056
Response x SOA	3.68	84.58	0.74	490.50	6202.21	1.82	.138	.073
Prompt x Response x SOA	5	115	0.75	351.26	6758.12	1.20	.316	.049

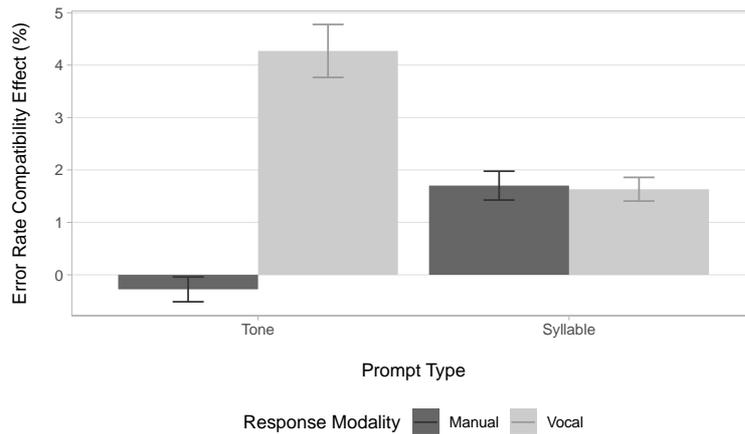
*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator.  $Epsilon$  indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

On average, participants made 5.5% errors (incorrect responses: 4.6%; missed trials: 0.6%; anticipatory responses: 0.3%; outliers: 0.0%). Analyses for ER compatibility effects (see Figure A.1, Table A.1, and Table A.2) revealed a main effect of response modality, with a smaller ER compatibility effect for manual responses than for vocal responses.

**Table A.2** ER compatibility effects, SEs, and 95% CIs for each level within response modality.

	Level	$Mean$	$SE$	95% CI
Response	Manual	0.71	0.48	[-0.27, 1.69]
	Vocal	2.95	0.81	[1.28, 4.62]

The two-way interaction between prompt type and response modality (see Figure A.2 and Table A.3) was significant, and follow-up  $t$  tests revealed a smaller ER compatibility effect for manual responses than for vocal responses when responding to tone prompts ( $t(23) = 2.94, p = .007$ ) but no difference between manual and vocal responses when responding to syllable prompts ( $t(23) = 0.08, p = .94$ ).

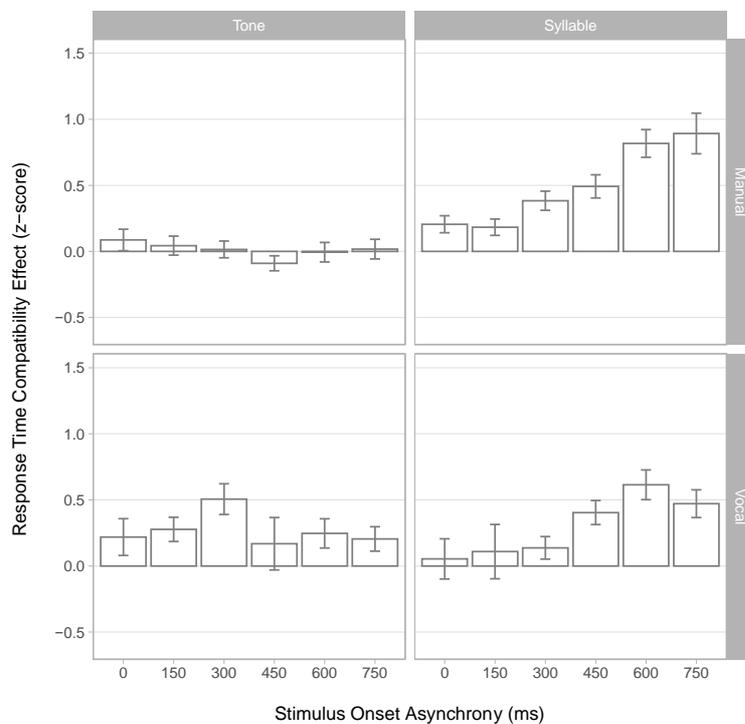


**Figure A.2** ER compatibility effects  $\pm$  SEs for each prompt type for each response modality.

**Table A.3** ER compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and response modality.

Prompt	Response	Mean	SE	95% CI
Tone	Manual	-0.28	0.59	[-1.50, 0.94]
	Vocal	4.27	1.26	[1.66, 6.88]
Syllable	Manual	1.70	0.69	[0.28, 3.13]
	Vocal	1.63	0.56	[0.47, 2.80]

### Response time analysis



**Figure A.3** Standardised RT compatibility effects  $\pm$  SEs in each experimental condition. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-

manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where SSC+SRC effects were expected.

**Table A.4** Three-way ANOVA summary for standardised RT compatibility effects, as a function of prompt type, response modality, and SOA.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	$Epsilon$	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Prompt	1	23	1.00	9.46	8.74	24.90	< .001	.520
Response	1	23	1.00	0.14	9.86	0.32	.577	.014
SOA	3.15	72.41	0.63	6.61	25.68	5.92	.001	.205
Prompt x Response	1	23	1.00	7.49	6.41	26.86	< .001	.539
Prompt x SOA	3.32	76.38	0.66	10.90	39.91	6.28	< .001	.215
Response x SOA	3.36	77.34	0.67	0.89	30.09	0.68	.582	.029
Prompt x Response x SOA	2.80	64.44	0.56	1.00	31.22	0.73	.527	.031

*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

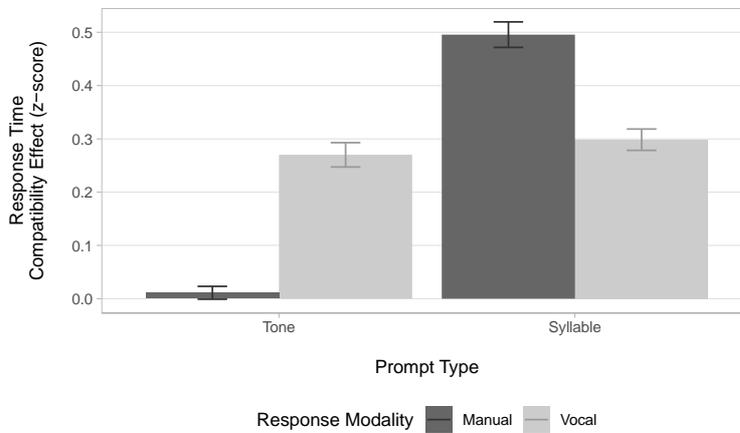
After errors were removed, analyses for standardised RT compatibility effects (see Figure A.3, Table A.4, and Table A.5) revealed a main effect of prompt type, with a smaller RT compatibility effect for tone prompts than for syllable prompts. Follow-up  $t$  tests for the main effect of SOA revealed a smaller RT compatibility effect at SOA1 than at SOA5-6 and a smaller RT compatibility effect at SOA2 than at SOA5 (all  $p < .03$ ).

**Table A.5** Standardised RT compatibility effects, SEs, and 95% CIs for each level within prompt type and SOA.

	Level	$Mean$	$SE$	95% CI
Prompt	Tone	0.14	0.03	[0.08, 0.21]
	Syllable	0.40	0.04	[0.31, 0.48]
SOA	1	0.14	0.04	[6.56, 6.69]
	2	0.15	0.07	[6.53, 6.66]
	3	0.26	0.03	[6.52, 6.65]
	4	0.24	0.06	[6.48, 6.61]
	5	0.42	0.05	[6.45, 6.58]
	6	0.40	0.06	[6.41, 6.54]

The two-way interaction between prompt type and response modality (see Figure A.4 and Table A.6) was significant, follow-up  $t$  tests revealed a smaller RT compatibility effect for tone prompts than for syllable prompts when performing manual responses ( $t(23) = 7.62$ ,

$p < .001$ ) but no difference between tone and syllable prompts when performing vocal responses ( $t(23) = 0.40, p = .70$ ).

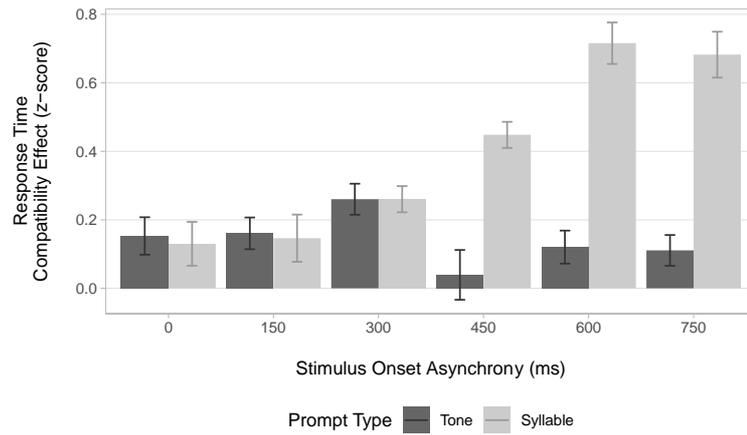


**Figure A.4** Standardised RT compatibility effects  $\pm$  SEs for each prompt type for each response modality.

**Table A.6** Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and response modality.

Prompt	Response	Mean	SE	95% CI
Tone	Manual	0.01	0.03	[-0.05, 0.07]
	Vocal	0.27	0.06	[0.15, 0.39]
Syllable	Manual	0.50	0.06	[0.37, 0.62]
	Vocal	0.30	0.05	[0.20, 0.40]

The two-way interaction between prompt type and SOA (see Figure A.5 and Table A.7) was significant, and follow-up  $t$  tests revealed smaller RT compatibility effects for tone prompts than for syllable prompts at SOA4-6 (all  $p < .04$ ). Further comparisons of RT compatibility effects across SOAs for each prompt type revealed no difference for RT compatibility effects across SOAs for tone prompts but smaller RT compatibility effects at SOA1-3 than at SOA5-6 for syllable prompts (all  $p < .03$ ).

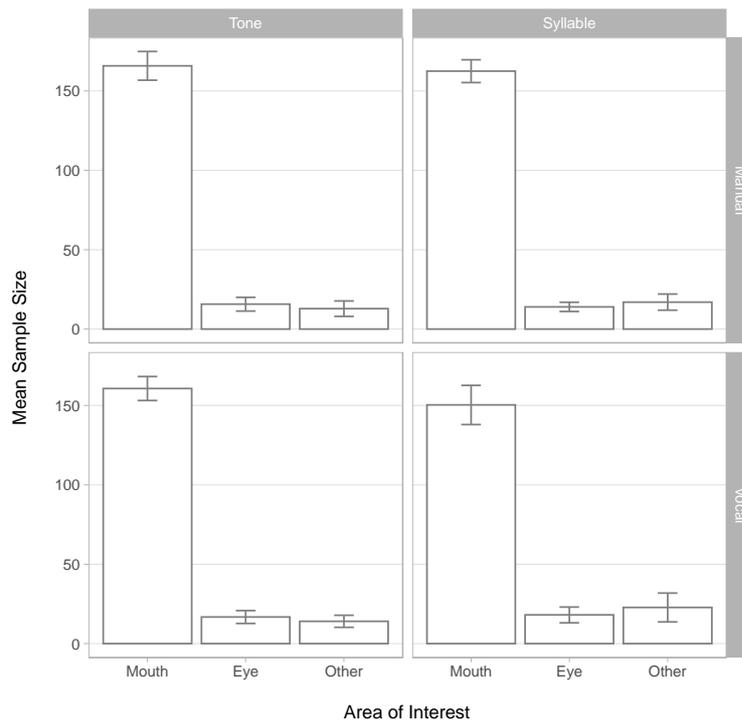


**Figure A.5** Standardised RT compatibility effects  $\pm$  SEs at each SOA for each prompt type.

**Table A.7** Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and SOA.

Prompt	SOA	Mean	SE	95% CI
Tone	1	0.15	0.08	[-0.01, 0.32]
	2	0.16	0.07	[0.02, 0.30]
	3	0.26	0.07	[0.13, 0.39]
	4	0.04	0.10	[-0.18, 0.25]
	5	0.12	0.07	[-0.02, 0.26]
	6	0.11	0.06	[-0.02, 0.24]
Syllable	1	0.13	0.09	[-0.06, 0.32]
	2	0.15	0.10	[-0.06, 0.35]
	3	0.26	0.06	[0.15, 0.37]
	4	0.45	0.05	[0.34, 0.56]
	5	0.72	0.09	[0.54, 0.90]
	6	0.68	0.10	[0.48, 0.88]

## Eye-tracking analysis



**Figure A.6** Mean sample sizes  $\pm$  SEs for each AOI in each task. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where SSC+SRC effects were expected.

**Table A.8** Three-way ANOVA summary for same sizes, as a function of prompt type, response modality, and AOI.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	$Epsilon$	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Prompt	1	17	1.00	2.34	402.40	0.10	.757	.006
Response	1	17	1.00	38.05	365.09	1.77	.201	.094
AOI	1.30	22.09	0.65	987015.46	127269.12	131.84	< .001	.886
Prompt x Response	1	17	1.00	0.59	343.65	0.03	.866	.002
Prompt x AOI	2	34	0.85	1572.71	12723.80	2.10	.138	.110
Response x AOI	1.41	24.04	0.71	1636.80	5718.75	4.87	.026	.223
Prompt x Response x AOI	1.40	23.80	0.70	353.58	20601.23	0.29	.671	.017

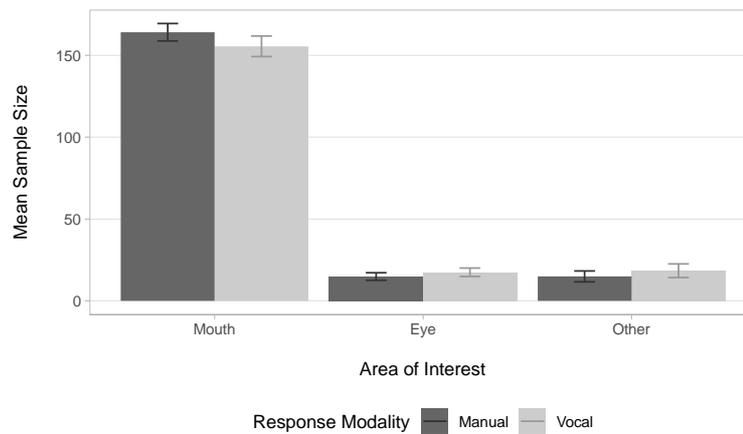
*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

Eye-tracking analyses (see Figure A.6, Table A.8, and Table A.9) revealed a main effect of AOI, with more samples collected from the mouth than from the eye and the other areas (both  $p < .001$ ).

**Table A.9** Mean sample sizes, SEs, and 95% CIs for each level within AOI.

	Level	Mean	SE	95% CI
AOI	Mouth	165.15	8.71	[146.47, 183.83]
	Eye	13.06	2.74	[7.18, 18.93]
	Other	16.38	6.37	[2.73, 30.03]

The two-way interaction between response modality and AOI (see Figure A.7 and Table A.10) was significant, and follow-up  $t$  tests revealed that, for the mouth area, more samples were collected for manual responses than for vocal responses ( $t(17) = 2.46, p = .025$ ).

**Figure A.7** Mean sample sizes  $\pm$  SEs for each AOI for each response modality.**Table A.10** Mean sample sizes, SEs, and 95% CIs for each level within the interaction between response modality and AOI.

Response	AOI	Mean	SE	95% CI
Manual	Mouth	164.11	7.64	[147.99, 180.22]
	Eye	14.84	3.33	[7.81, 21.86]
	Other	14.93	4.79	[4.83, 25.02]
Vocal	Mouth	155.50	9.02	[136.46, 174.53]
	Eye	17.43	3.68	[9.67, 25.18]
	Other	18.43	5.95	[5.88, 30.97]

### Discussion

The aim of Experiment 3 was to replicate behavioural findings from Experiment 2 reported in Chapter 3, but with participants instructed to look at the speaker's mouth, instead of face, while completing each task. Firstly, the eye-tracking analysis in the current experiment revealed that more eye-tracking samples were collected from the mouth area than from the eye or the other areas, suggesting that participants were following the instructions to look at the speaker's mouth while responding to auditory prompts. Interestingly, the current eye-

tracking analysis also revealed that more eye-tracking samples were collected from the mouth area when manual responses were required than when vocal responses were required. This finding was consistent with the observation in Experiment 2 that people tended to look at the mouth area more in the syllable-manual task. It could be that, when required to say something, participants unavoidably tended to look away from the speaker's mouth because the speaker's articulation might interfere with their own articulatory performance.

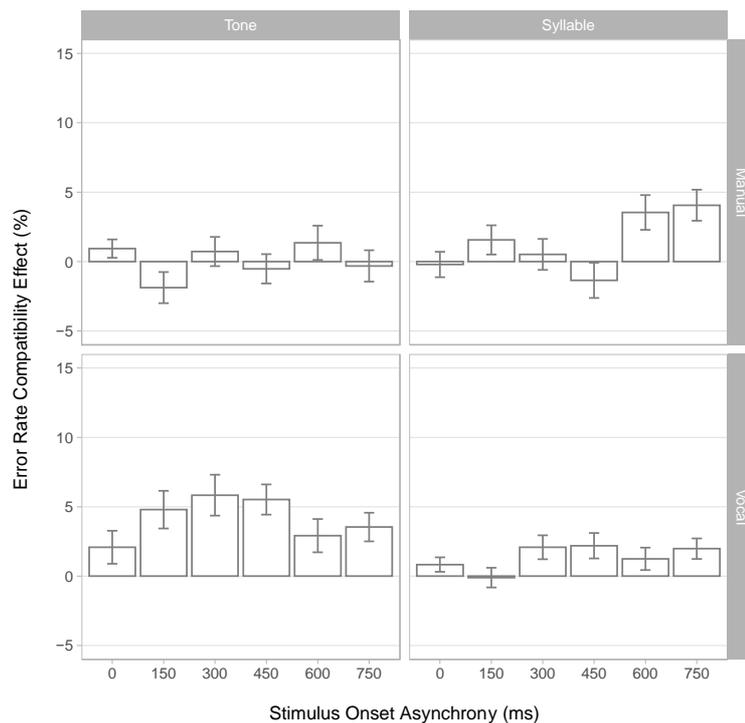
Behavioural results from Experiment 2 and 3 were similar, except that a few interactions involving SOA did not reach significance in Experiment 3. Different from Experiment 2, the interaction between prompt type, response modality, and SOA in the current experiment was not significant for RT compatibility effects. Additionally, different from Experiment 2, the current experiment did not find a significant interaction between prompt type and SOA for ER compatibility effects or between response modality and SOA for RT compatibility effects. It seemed that the time courses for compatibility effects in the current experiment were less clear than those in Experiment 2. In Appendix B, further analyses for both behavioural and eye-tracking results were conducted by combining data from Experiment 2 and 3. These analyses were conducted to examine whether participants' gaze behaviours were statistically different between two experiments and if so, whether this difference in gaze behaviours affected compatibility effects elicited by the perception of speech distractors.

## Appendix B Further analysis of Experiment 2 and 3

In this section, ER and RT compatibility effects from both Experiment 2 (Chapter 3) and Experiment 3 (Appendix A) were subjected to repeated-measures ANOVA with prompt type (tone versus syllable), response modality (manual versus vocal), and SOA (0, 150, 300, 450, 600, or 750 ms) as within-subjects variables and experiment (Experiment 2 versus 3) as a between-subjects variable. Eye-tracking sample sizes from both experiments were also subjected to repeated-measures ANOVA with prompt type, response modality, and AOI (mouth, eye, or other) as within-subjects variables and experiment as a between-subjects variable. The significance level was set to  $p < .05$ . Greenhouse-Geisser correction for non-sphericity and Bonferroni-correction for multiple comparisons were applied whenever appropriate.

### Results

#### Error rate analysis



**Figure B.1** ER compatibility effects  $\pm$  SEs in each experimental condition from both Experiment 2 and 3. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where SSC+SRC effects were expected.

**Table B.1** Four-way ANOVA summary for ER compatibility effects, as a function of prompt type, response modality, SOA, and experiment.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	$Epsilon$	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Experiment	1	46	1.00	13.56	4206.73	0.15	.702	.003
Prompt	1	46	1.00	149.50	2851.35	2.41	.127	.050
Prompt x Experiment	1	46	1.00	43.95	2851.35	0.71	.404	.015
Response	1	46	1.00	1198.46	4809.85	11.46	.001	.199
Response x Experiment	1	46	1.00	11.48	4809.85	0.11	.742	.002
SOA	5	230	0.85	402.45	10230.51	1.81	.112	.038
SOA x Experiment	5	230	0.85	257.66	10230.51	1.16	.330	.025
Prompt x Response	1	46	1.00	1178.15	4063.85	13.34	.001	.225
Prompt x Response x Experiment	1	46	1.00	23.63	4063.85	0.27	.607	.006
Prompt x SOA	4.12	189.48	0.82	439.43	11086.94	1.82	.124	.038
Prompt x SOA x Experiment	4.12	189.48	0.82	535.09	11086.94	2.22	.066	.046
Response x SOA	5	230	0.88	838.39	11581.55	3.33	.006	.068
Response x SOA x Experiment	5	230	0.88	91.52	11581.55	0.36	.873	.008
Prompt x Response x SOA	5	230	0.86	482.14	11099.44	2.00	.080	.042
Prompt x Response x SOA x Experiment	5	230	0.86	59.05	11099.44	0.25	.942	.005

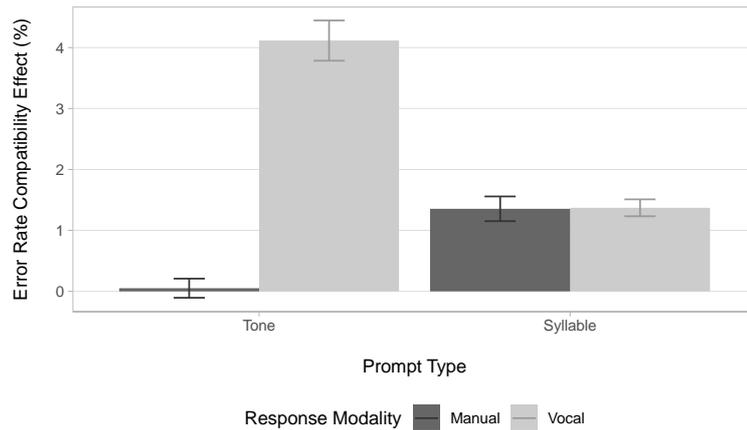
*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

Analyses for ER compatibility effects (see Figure B.1, Table B.1, and Table B.2) revealed a main effect of response modality, with a smaller ER compatibility effect for manual responses than for vocal responses.

**Table B.2** ER compatibility effects, SEs, and 95% CIs for each level within response modality.

	Level	Mean	SE	95% CI
Response	Manual	0.70	0.30	[0.104, 1.303]
	Vocal	2.74	0.50	[1.733, 3.753]

The two-way interaction between prompt type and response modality (see Figure B.2 and Table B.3) was significant, and follow-up  $t$  tests revealed a smaller ER compatibility effect for manual responses than for vocal responses when responding to tone prompts ( $t(47) = 4.13$ ,  $p < .001$ ) but no difference between manual and vocal responses when responding to syllable prompts ( $t(47) = 0.03$ ,  $p = .98$ ).

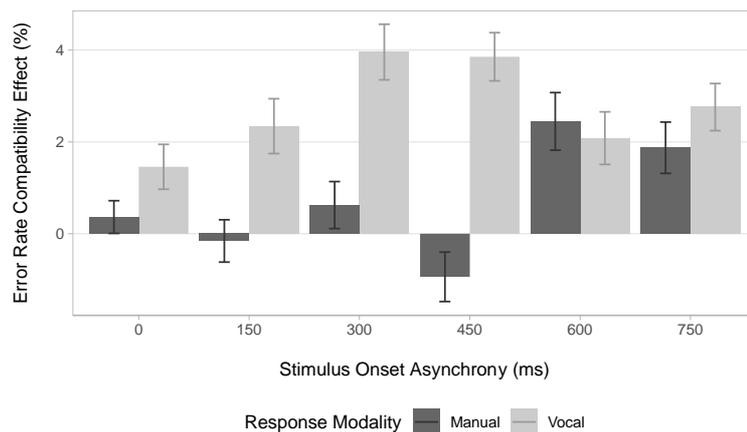


**Figure B.2** ER compatibility effects  $\pm$  SEs for each prompt type for each response modality.

**Table B.3** ER compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and response modality.

Prompt	Response	Mean	SE	95% CI
Tone	Manual	0.05	0.39	[-0.73, 0.84]
	Vocal	4.12	0.82	[2.46, 5.77]
Syllable	Manual	1.35	0.50	[0.34, 2.37]
	Vocal	1.37	0.34	[0.68, 2.06]

The two-way interaction between response modality and SOA (see Figure B.3 and Table B.4) was significant, and follow-up *t* tests revealed smaller ER compatibility effects for manual responses than for vocal responses at SOA2-4 (all  $p < .04$ ).



**Figure B.3** ER compatibility effects  $\pm$  SEs at each SOA for each response modality.

**Table B.4** ER compatibility effects, SEs, and 95% CIs for each level within the interaction between response modality and SOA.

Response	SOA	Mean	SE	95% CI
Manual	1	0.37	0.51	[-0.66, 1.39]
	2	-0.16	0.66	[-1.49, 1.18]
	3	0.63	0.73	[-0.85, 2.10]
	4	-0.94	0.77	[-2.48, 0.60]
	5	2.45	0.90	[0.64, 4.26]
	6	1.88	0.79	[0.28, 3.47]
Vocal	1	1.46	0.70	[0.05, 2.87]
	2	2.34	0.86	[0.61, 4.07]
	3	3.96	0.87	[2.21, 5.71]
	4	3.85	0.75	[2.34, 5.37]
	5	2.08	0.79	[0.49, 3.68]
	6	2.76	0.74	[1.28, 4.25]

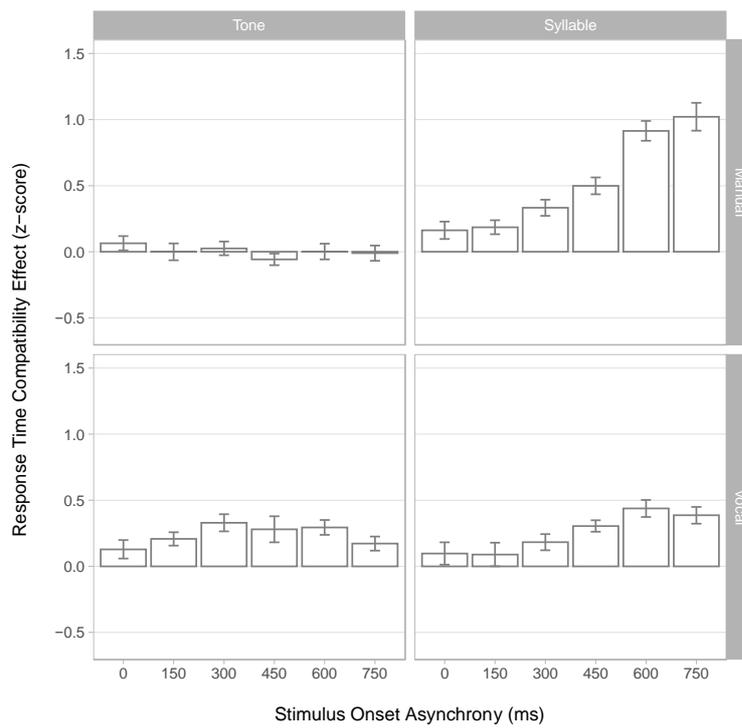
Interestingly, the three-way interaction between prompt type, response modality, and SOA (see Figure B.1 and Table B.5) was marginally significant ( $p = .08$ ). Follow-up analyses were conducted. Again, NO compatibility effects were expected from the tone-manual task, SSC effects were expected from the syllable-manual task, SRC effects were expected from the tone-vocal task, and SSC+SRC effects were expected from the syllable-vocal task. Comparisons of ER compatibility effects across four tasks at each SOA were conducted and revealed that (all  $p < .05$ ):

1. SRC effects were larger than NO effects at SOA2 and SOA4, and the larger difference was at SOA2.
2. SRC effects were larger than SSC effects at SOA3-4, and the larger difference was at SOA4.
3. SRC effects were larger than SSC+SRC effects at SOA2.

**Table B.5** ER compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type, response modality, and SOA.

Prompt	Response	SOA	Mean	SE	95% CI
Tone	Manual	1	0.94	0.67	[-0.41, 2.28]
		2	-1.88	1.11	[-4.10, 0.35]
		3	0.73	1.06	[-1.41, 2.87]
		4	-0.52	1.05	[-2.63, 1.59]
		5	1.35	1.24	[-1.14, 3.85]
		6	-0.31	1.14	[-2.60, 1.98]
	Vocal	1	2.08	1.20	[-0.33, 4.50]
		2	4.79	1.35	[2.07, 7.52]
		3	5.83	1.49	[2.83, 8.84]
		4	5.52	1.10	[3.31, 7.73]
		5	2.92	1.16	[0.57, 5.26]
		6	3.54	1.04	[1.45, 5.63]
Syllable	Manual	1	-0.21	0.93	[-2.08, 1.66]
		2	1.56	1.03	[-0.52, 3.64]
		3	0.52	1.12	[-1.74, 2.78]
		4	-1.35	1.28	[-3.94, 1.23]
		5	3.54	1.27	[0.99, 6.09]
		6	4.06	1.12	[1.81, 6.32]
	Vocal	1	0.83	0.53	[-0.22, 1.89]
		2	-0.10	0.69	[-1.48, 1.28]
		3	2.08	0.88	[0.32, 3.85]
		4	2.19	0.93	[0.33, 4.05]
		5	1.25	0.81	[-0.39, 2.89]
		6	1.98	0.75	[0.48, 3.48]

## Response time analysis



**Figure B.4** Standardised RT compatibility effects  $\pm$  SEs in each experimental condition from both Experiment 2 and 3. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where SSC+SRC effects were expected.

**Table B.6** Four-way ANOVA summary for standardised ER compatibility effects, as a function of prompt type, response modality, SOA, and experiment.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	$Epsilon$	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Experiment	1	46	1.00	0.053	17.264	0.142	.708	.003
Prompt	1	46	1.00	20.21	15.66	59.34	< .001	.563
Prompt x Experiment	1	46	1.00	0.00	15.66	0.00	.95	.000
Response	1	46	1.00	0.10	14.60	0.33	.57	.007
Response x Experiment	1	46	1.00	0.08	14.60	0.25	.623	.005
SOA	3.99	183.29	0.80	15.96	39.92	18.39	< .001	.286
SOA x Experiment	3.99	183.29	0.80	0.61	39.92	0.71	.587	.015
Prompt x Response	1	46	1.00	18.06	10.30	80.63	< .001	.637
Prompt x Response x Experiment	1	46	1.00	0.03	10.30	0.14	.712	.003
Prompt x SOA	3.87	178.19	0.78	16.27	55.17	13.57	< .001	.228
Prompt x SOA x Experiment	3.87	178.19	0.78	0.99	55.17	0.83	.505	.018
Response x SOA	4.09	188.25	0.82	3.44	40.69	3.89	.004	.078
Response x SOA x Experiment	4.09	188.25	0.82	0.23	40.69	0.26	.908	.006
Prompt x Response x SOA	3.59	165.03	0.72	4.23	42.75	4.55	.002	.090
Prompt x Response x SOA x Experiment	3.59	165.03	0.72	1.64	42.75	1.76	.146	.037

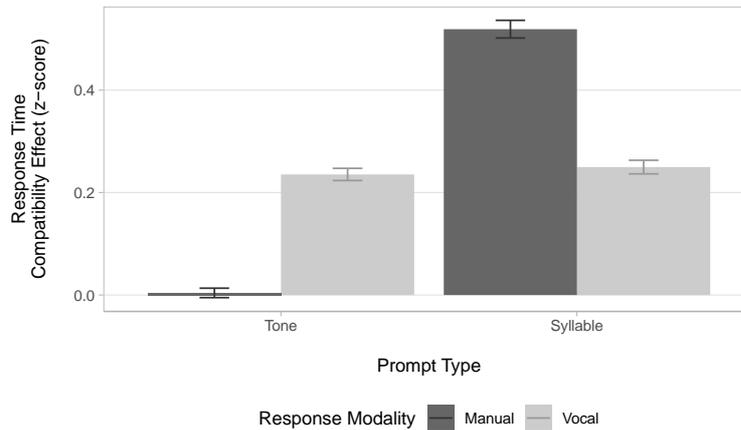
*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

Analyses for standardised RT compatibility effects (see Figure B.4, Table B.6, and Table B.7) revealed a main effect of prompt type, with a smaller RT compatibility effect for tone prompts than for syllable prompts. Follow-up  $t$  tests for the main effect of SOA revealed smaller RT compatibility effects at SOA1-4 than at SOA5-6 (all  $p < .05$ ).

**Table B.7** Standardised RT compatibility effects, SEs, and 95% CIs for each level within prompt type and SOA.

	Level	Mean	SE	95% CI
Prompt	Tone	0.12	0.02	[0.08, 0.16]
	Syllable	0.39	0.03	[0.32, 0.45]
SOA	1	0.11	0.03	[0.06, 0.17]
	2	0.12	0.04	[0.04, 0.20]
	3	0.22	0.02	[0.17, 0.26]
	4	0.26	0.03	[0.19, 0.32]
	5	0.41	0.03	[0.34, 0.48]
	6	0.39	0.04	[0.32, 0.47]

The two-way interaction between prompt type and response modality (see Figure B.5 and Table B.8) was significant, follow-up  $t$  tests revealed a smaller RT compatibility effect for tone prompts than for syllable prompts when performing manual responses ( $t(47) = 10.51$ ,  $p < .001$ ) but no difference between tone and syllable prompts when performing vocal responses ( $t(47) = 0.36$ ,  $p = .72$ ).

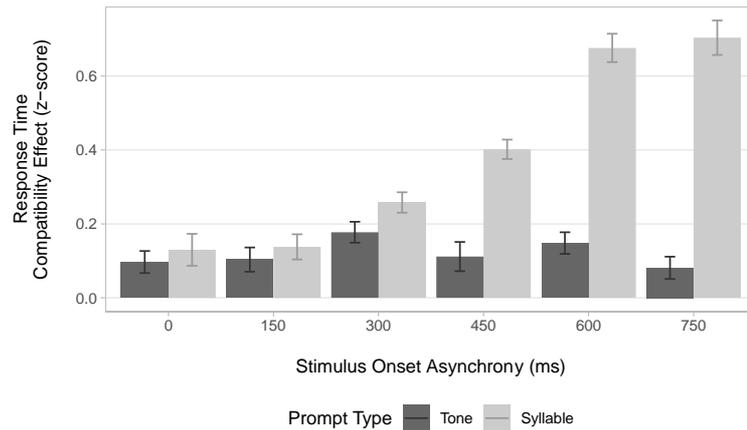


**Figure B.5** Standardised RT compatibility effects  $\pm$  SEs for each prompt type for each response modality.

**Table B.8** Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and response modality.

Prompt	Response	Mean	SE	95% CI
Tone	Manual	0.00	0.02	[-0.04, 0.05]
	Vocal	0.24	0.03	[0.18, 0.30]
Syllable	Manual	0.52	0.04	[0.43, 0.61]
	Vocal	0.25	0.03	[0.18, 0.32]

The two-way interaction between prompt type and SOA (see Figure B.6 and Table B.9) was significant, and follow-up  $t$  tests revealed smaller RT compatibility effects for tone prompts than for syllable prompts at SOA4-6 (all  $p < .001$ ). Further comparisons of RT compatibility effects across SOAs for each prompt type revealed no difference across SOAs for tone prompts but smaller RT compatibility effects at SOA1-2 than at SOA4-6 and smaller effects at SOA3-4 than at SOA5-6 for syllable prompts (all  $p < .004$ ).

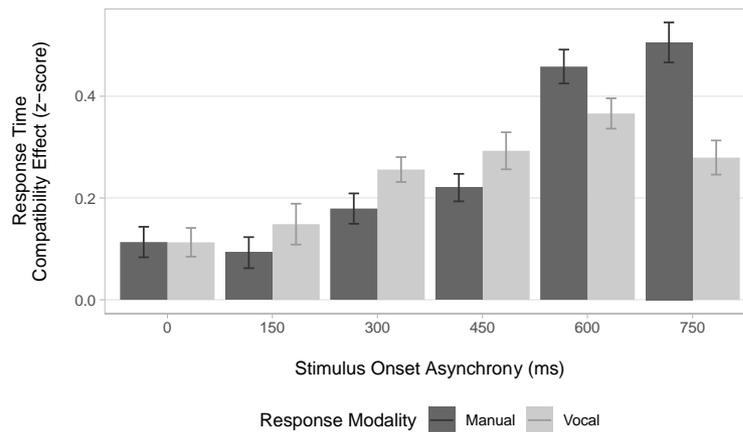


**Figure B.6** Standardised RT compatibility effects  $\pm$  SEs at each SOA for each prompt type.

**Table B.9** Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type and SOA.

Prompt	SOA	Mean	SE	95% CI
Tone	1	0.10	0.04	[0.01, 0.18]
	2	0.10	0.05	[0.01, 0.20]
	3	0.18	0.04	[0.10, 0.26]
	4	0.11	0.06	[0.00, 0.22]
	5	0.15	0.04	[0.06, 0.23]
	6	0.08	0.04	[-0.01, 0.17]
Syllable	1	0.13	0.06	[0.01, 0.26]
	2	0.14	0.05	[0.04, 0.24]
	3	0.26	0.04	[0.18, 0.34]
	4	0.40	0.04	[0.33, 0.48]
	5	0.68	0.06	[0.57, 0.79]
	6	0.70	0.07	[0.57, 0.84]

The two-way interaction between response modality and SOA (see Figure B.7 and Table B.10) was significant, and follow-up  $t$  tests revealed a larger RT compatibility effect for manual responses than for vocal responses at SOA6 (both  $p = .003$ ). Further comparisons of RT compatibility effects across SOAs for each response modality revealed smaller RT compatibility effects at SOA1-4 than at SOA5-6 for manual responses and smaller RT compatibility effects at SOA1-2 than at SOA5 for vocal responses (all  $p < .02$ ).



**Figure B.7** Standardised RT compatibility effects  $\pm$  SEs at each SOA for each response modality.

**Table B.10** Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between response modality and SOA.

Response	SOA	Mean	SE	95% CI
Manual	1	0.11	0.04	[0.03, 0.20]
	2	0.09	0.04	[0.01, 0.18]
	3	0.18	0.04	[0.09, 0.27]
	4	0.22	0.04	[0.14, 0.30]
	5	0.46	0.05	[0.36, 0.55]
	6	0.51	0.06	[0.39, 0.62]
Vocal	1	0.11	0.04	[0.03, 0.19]
	2	0.15	0.06	[0.03, 0.27]
	3	0.26	0.04	[0.19, 0.33]
	4	0.29	0.05	[0.19, 0.40]
	5	0.37	0.04	[0.28, 0.45]
	6	0.28	0.05	[0.18, 0.38]

Of primary interest, the three-way interaction between prompt type, response modality, and SOA (see Figure B.4 and Table B.11) was significant. Again, NO compatibility effects were expected from the tone-manual task, SSC effects were expected from the syllable-manual task, SRC effects were expected from the tone-vocal task, and SSC+SRC effects were expected from the syllable-vocal task. Comparisons of RT compatibility effects across four tasks at each SOA were further conducted and revealed that (all  $p < .05$ ):

1. SSC effects were larger than NO effects at SOA3-6, and the largest difference was at SOA6.
2. SRC effects were larger than NO effects at SOA2-5, and the largest difference was at SOA4.
3. SSC+SRC effects were larger than NO effects at SOA4-6, and the largest difference was at SOA5.

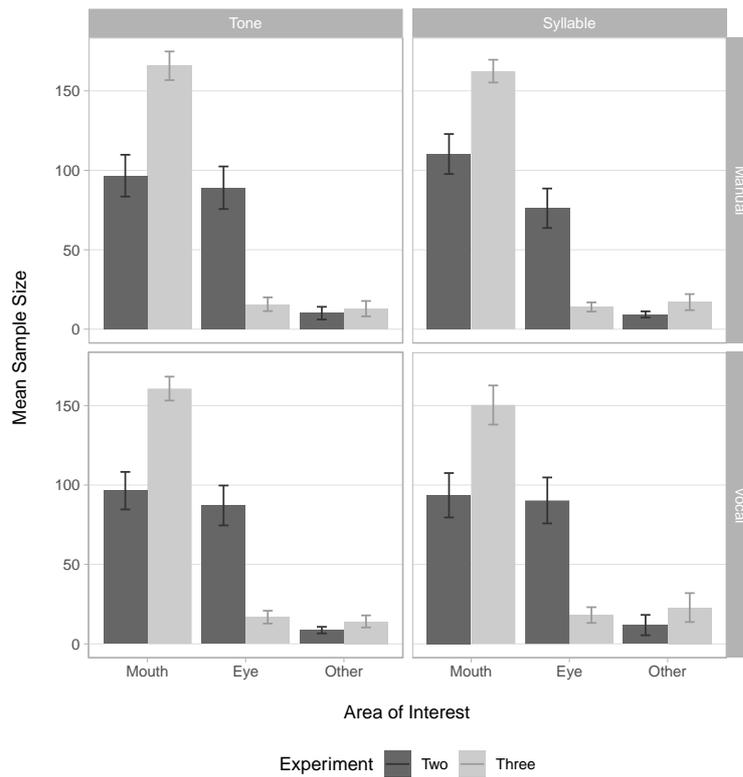
4. SSC effects were larger than SRC effects at SOA5-6, and the larger difference was at SOA6.
5. SSC effects were larger than SSC+SRC effects at SOA5-6, and the larger difference was at SOA6.
6. The SSC+SRC effect was larger than the SRC effect at SOA6.

Consistent with Prediction 5, these results suggested that SSC effects reached maximum at a longer SOA and SRC effects reached maximum at a shorter SOA than SSC+SRC effects.

**Table B.11** Standardised RT compatibility effects, SEs, and 95% CIs for each level within the interaction between prompt type, response modality, and SOA.

Prompt	Response	SOA	Mean	SE	95% CI
Tone	Manual	1	0.07	0.05	[-0.05, 0.17]
		2	0.00	0.06	[-0.13, 0.13]
		3	0.03	0.05	[-0.08, 0.131]
		4	-0.06	0.04	[-0.15, 0.03]
		5	0.00	0.06	[-0.12, 0.12]
		6	-0.01	0.06	[-0.13, 0.11]
	Vocal	1	0.13	0.07	[-0.01, 0.27]
		2	0.21	0.05	[0.10, 0.31]
		3	0.33	0.06	[0.20, 0.46]
		4	0.28	0.10	[0.09, 0.48]
		5	0.29	0.06	[0.18, 0.41]
		6	0.17	0.05	[0.06, 0.28]
Syllable	Manual	1	0.16	0.07	[0.03, 0.30]
		2	0.19	0.05	[0.08, 0.29]
		3	0.33	0.06	[0.21, 0.46]
		4	0.50	0.07	[0.37, 0.63]
		5	0.92	0.08	[0.76, 1.07]
		6	1.02	0.11	[0.81, 1.23]
	Vocal	1	0.10	0.09	[-0.07, 0.27]
		2	0.09	0.09	[-0.09, 0.27]
		3	0.18	0.06	[0.06, 0.31]
		4	0.31	0.04	[0.22, 0.39]
		5	0.44	0.06	[0.31, 0.57]
		6	0.39	0.06	[0.26, 0.52]

## Eye-tracking analysis



**Figure B.8** Mean sample sizes  $\pm$  SEs for each AOI for each task from Experiment 2 and 3. Four panels represent the tone-manual task (left-top) where no compatibility effects were expected, the syllable-manual task (right-top) where SSC effects were expected, the tone-vocal task (left-bottom) where SRC effects were expected, and the syllable-vocal task (right-bottom) where SSC+SRC effects were expected.

**Table B.12** Four-way ANOVA summary for same sizes, as a function of prompt type, response modality, AOI, and experiment.

Main Effect/Interaction	$df_{Num}$	$df_{Den}$	$Epsilon$	$SS_{Num}$	$SS_{Den}$	$F$	$p$	$\eta^2_p$
Experiment	1	35	1.00	56.461	4802.743	0.411	.525	.012
Prompt	1	35	1.00	3.37	549.83	0.21	.646	.006
Prompt x Experiment	1	35	1.00	16.23	549.83	1.03	.316	.029
Response	1	35	1.00	59.92	671.31	3.12	.086	.082
Response x Experiment	1	35	1.00	1.21	671.31	0.06	.803	.002
AOI	1.33	46.71	0.67	1039847.81	503089.76	72.34	< .001	.674
AOI x Experiment	1.33	46.71	0.67	315991.68	503089.76	21.98	< .001	.386
Prompt x Response	1	35	1.00	13.84	623.10	0.78	.384	.022
Prompt x Response x Experiment	1	35	1.00	6.84	623.10	0.38	.539	.011
Prompt x AOI	1.31	45.68	0.65	786.54	53703.65	0.51	.525	.014
Prompt x AOI x Experiment	1.31	45.68	0.65	1809.14	53703.65	1.18	.298	.033
Response x AOI	1.26	44.09	0.63	3484.11	24076.17	5.07	.022	.126
Response x AOI x Experiment	1.26	44.09	0.63	195.68	24076.17	0.28	.650	.008
Prompt x Response x AOI	1.68	58.73	0.84	2274.96	72479.35	1.10	.331	.030
Prompt x Response x AOI x Experiment	1.68	58.73	0.84	596.49	72479.35	0.29	.712	.008

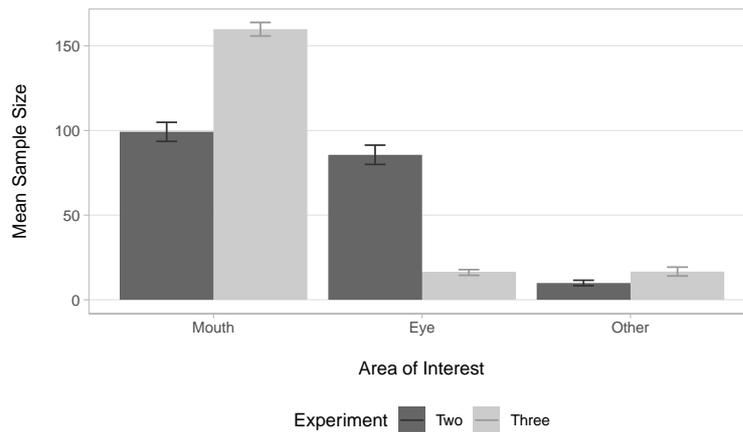
*Note.*  $df_{Num}$  indicates degrees of freedom numerator.  $df_{Den}$  indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom,  $p$ -values and degrees of freedom in the table incorporate this correction.  $SS_{Num}$  indicates Type III sum of squares numerator.  $SS_{Den}$  indicates Type III sum of squares denominator.  $\eta^2_p$  indicates partial eta-squared. Significant main effects and interactions with  $p < .05$  are highlighted.

Eye-tracking analyses (see Figure B.8, Table B.12, and Table B.13) revealed a main effect of AOI, with more samples collected from the mouth area than from the eye and the other areas and more samples collected from the eye area than from the other area (all  $p < .001$ ).

**Table B.13** Mean sample sizes, SEs, and 95% CIs for each level within AOI.

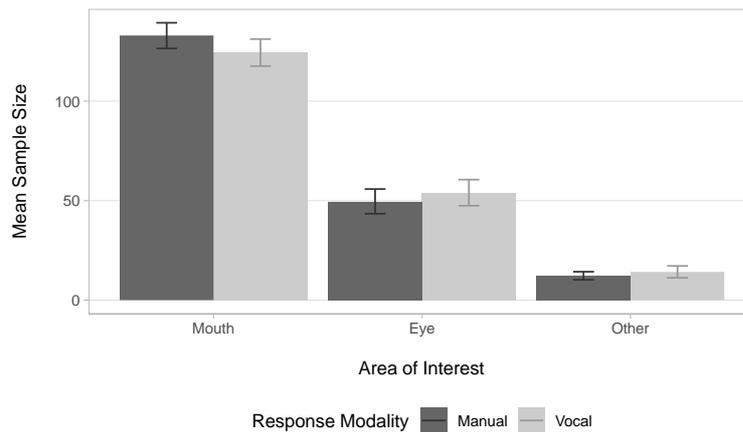
	Level	Mean	SE	95% CI
AOI	Mouth	129.50	7.09	[115.11, 143.88]
	Eye	50.88	6.20	[38.29, 63.47]
	Other	13.30	3.07	[7.07, 19.54]

The two-way interaction between AOI and experiment (see Figure B.9, Table 3.19, and Table A.9) was significant, and it was because more samples were collected from the mouth and the eye areas than from the other area in Experiment 2 and more samples were collected from the mouth area than from the eye and the other areas in Experiment 3.



**Figure B.9** Mean sample sizes  $\pm$  SEs for each AOI for each experiment.

The two-way interaction between response modality and AOI (see Figure B.10 and Table B.14) was significant, and follow-up  $t$  tests revealed that, for the mouth area, more samples were collected for manual responses than for vocal responses ( $t(36) = 2.64, p = .012$ ).



**Figure B.10** Mean sample sizes  $\pm$  SEs for each AOI for each response modality.

**Table B.14** Mean sample sizes, SEs, and 95% CIs for each level within the interaction between response modality and AOI.

Response	AOI	Mean	SE	95% CI
Manual	Mouth	53.05	6.40	[40.06, 66.04]
	Eye	125.22	7.48	[110.04, 140.41]
	Other	14.30	3.49	[7.22, 21.39]
Vocal	Mouth	48.71	6.29	[35.93, 61.48]
	Eye	133.77	7.05	[119.45, 148.08]
	Other	12.31	2.72	[6.78, 17.83]

## *Discussion*

Further analysis of behavioural data from both Experiment 2 and 3 revealed similar results to those observed in Experiment 2 and 3 when analysed separately. Most importantly, neither the main effect of experiment nor any interaction involving experiment was significant. The interaction between prompt type, response modality, and SOA was marginally significant for ER compatibility effects and significant for RT compatibility effects. Consistent with Prediction 5 given in Section 3.3.1, follow-up analyses revealed that SSC effects reached maximum at a longer SOA and SRC effects reached maximum at a shorter SOA than SSC+SRC effects. These findings suggested that automatic imitation effects are confounded with, but cannot be reduced to, SSC effects in the syllable-vocal condition where both S-S and S-R overlap is present. Eye-tracking analysis suggested that participants in Experiment 2 and 3 were following the instructions to look at the speaker's face or to focus on her mouth respectively, but this difference in gaze behaviours did not affect compatibility effects elicited by the perception of speech distractors.

**Appendix C Table C.1**

**Table C.1** Mean log-transformed RTs, back-transformed RTs, SEs, and 95% CIs for each level in the four-way interaction between training, test, compatibility, and SOA for the training study reported in Chapter 4.

Training	Test	Compatibility	SOA (ms)	Mean	RT(ms)	SE	95% CI
Mirror	Pre-training	Incompatible	552	6.43	618	0.04	[6.36, 6.50]
			736	6.39	596	0.03	[6.33, 6.45]
			920	6.35	572	0.03	[6.29, 6.41]
			1104	6.32	556	0.03	[6.27, 6.37]
		Compatible	552	6.38	591	0.04	[6.31, 6.45]
			736	6.33	562	0.04	[6.26, 6.40]
			920	6.28	533	0.03	[6.21, 6.35]
			1104	6.24	512	0.03	[6.18, 6.30]
	Post-training	Incompatible	552	6.35	571	0.02	[6.30, 6.39]
			736	6.30	542	0.02	[6.25, 6.34]
			920	6.26	525	0.02	[6.22, 6.31]
			1104	6.25	519	0.02	[6.21, 6.30]
		Compatible	552	6.29	536	0.03	[6.24, 6.34]
			736	6.22	501	0.02	[6.17, 6.26]
			920	6.15	468	0.02	[6.10, 6.20]
			1104	6.13	459	0.02	[6.08, 6.17]
Counter-mirror	Pre-training	Incompatible	552	6.37	586	0.04	[6.30, 6.44]
			736	6.34	568	0.03	[6.28, 6.41]
			920	6.32	555	0.03	[6.26, 6.38]
			1104	6.30	543	0.03	[6.25, 6.35]
		Compatible	552	6.34	565	0.04	[6.27, 6.41]
			736	6.28	534	0.04	[6.21, 6.35]
			920	6.22	504	0.03	[6.16, 6.29]
			1104	6.20	495	0.03	[6.15, 6.26]
	Post-training	Incompatible	552	6.35	570	0.02	[6.30, 6.39]
			736	6.28	535	0.02	[6.24, 6.33]
			920	6.23	509	0.02	[6.19, 6.28]
			1104	6.22	500	0.02	[6.17, 6.26]
		Compatible	552	6.30	547	0.03	[6.26, 6.35]
			736	6.23	506	0.02	[6.18, 6.28]
			920	6.15	467	0.02	[6.10, 6.19]
			1104	6.15	470	0.02	[6.11, 6.20]