Auditory-motor interactions during speech perception and production

DING-LAN TANG





St Anne's College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy
November 2020

Acknowledgements

I am very grateful for the funding from the China Scholar Council.

I am extremely grateful to my supervisor, Prof. Kate Watkins for her continual support, academically and personally. Her immense knowledge and plentiful experience helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor for my DPhil. I would also like to thank my second supervisor, Prof. Charlotte Stagg for her insightful comments and encouragement. Many thanks also to my master supervisor, Dr Yu-Xuan Zhang (张语轩) who encouraged me to pursue my doctoral degree and looked after me in all the many ways.

I would like to thank all the members in the Speech and Brain Lab. In particular to Salome and Gabe, for all the support they gave me during the pandemic. It truly has been very, very good time in this lab.

To Miriam, I am so glad I found you during the departmental induction-thanks for always being there for me. To TaoYe (陶烨) and Xu YueHua(徐悦华), for your friendship and reassurance across the seas. To Bea, for all the calls that have encouraged and inspired me. I feel so lucky to have had the support of great friends throughout my DPhil.

Finally, I would like to express my gratitude to my great family (my parents and my sister-刘兰), without their tremendous understanding and encouragement in the past few years, it would be impossible for me to leave my home country and complete my study in UK.

A version of Chapter 3 is available as a preprint on bioRxiv DOI: https://doi.org/10.1101/733121. Tang, D-L., Lametti, D. R; Watkins, K. E., (2019). Altered auditory feedback induces coupled changes in formant frequencies during speech production.

A version of Chapter 4 has been submitted to a journal where it is currently under review (under review). Tang, D-L., McDaniel, A; Watkins, K. E., (2020). Disruption of Speech Motor Adaptation with Repetitive Transcranial Magnetic Stimulation of the Articulatory Representation in Primary Motor Cortex.

A version of Chapter 5 has been published on JNeurosci. Tang, D-L., Möttönen, R; Asaridou, S. S, Watkins, K. E., (2020). Asymmetry of auditory-motor speech processing is determined by language experience. DOI: https://doi.org/10.1523/JNEUROSCI.1977-20.2020

Abstract

Speech perception and production are two central aspects of human language. It remains unclear, however, how sensory and motor systems interact with each other during these two processes. The overall aim of this thesis was to investigate auditory-motor interactions during both speech perception and production.

There were two lines of research. The first explored sensorimotor adaptation in speech production, which plays an important role in the development and maintenance of fluent speech production. In study 1 (chapter 3), I examined the compensatory responses of speakers when they experienced perturbations to the auditory feedback of their own speech. The results showed feedback perturbation induced changes in subsequent speech production that were evident not only in the shifted formant, but also in the unaltered one. Moreover, such combined changes resulted in moving the produced speech token closer to a known pattern of speech production for an existing vowel category. This study confirmed the utility of this specific experimental paradigm, and that the apparatus used to apply real-time feedback perturbation was effective. It allowed the development of additional research questions to be tackled using the speech feedback perturbation paradigm.

Following the successful implementation of the paradigm in study 1, I carried out a brain stimulation study (chapter 4) exploring the neural basis of sensorimotor adaptation in speech using transcranial magnetic stimulation (TMS), which temporarily inhibits function in stimulated brain areas. This allowed testing of causal hypotheses concerning the role of different parts of the motor system in

speech motor adaptation. Here, I aimed to determine whether temporary disruption of the articulatory representation in left primary motor cortex (M1) by repetitive transcranial magnetic stimulation (rTMS) impairs sensorimotor adaptation in speech. Results showed that disruption of tongue, but not hand representation in left M1 almost abolished the expected compensatory response for the upwards F1 shift in auditory feedback, namely a decrease in F1 frequency alongside an increase in F2 frequency. This is the first study to address the key question as to whether cortical motor areas that control movements of articulators causally contribute to speech motor adaptation.

The second line of research explored in this thesis was focused on the asymmetry of auditory-motor speech processing in tonal and non-tonal language speakers (study 3, chapter 5). To achieve this, I temporarily disrupted activity in either the left or right motor cortex using the same form of rTMS used in Study 2 and measured the impact of these disruptions on auditory processing of phoneme and tone changes in sequences of syllables using electroencephalography (EEG). I found that the effect of motor disruptions on processing of tone changes differed in tonal and nontonal language speakers: disruption of the right speech motor cortex affected processing of tone changes in non-tonal but not tonal language speakers, whereas disruption of the left speech motor cortex significantly impaired processing of tone changes in tonal language speakers. In contrast, disruption of the speech motor cortex in the left, but not the right hemisphere, impaired processing of phoneme changes in both language groups equally. For the first time, I provide causal evidence that

auditory-motor speech processing asymmetries are shaped by the functional role of the acoustic cues in the listener's native language.

Overall, the work presented in this thesis emphasizes the importance of understanding the interactions between auditory and motor systems when investigating speech perception and production. It also calls for more comparative studies between speakers of different languages, since ignoring the diversity of spoken languages can to lead to an incomplete understanding of the neurobiology of speech and language.

Contents

ABSTRACT ..... **IV**

LIST OF FIGURES ..... **X**

LIST OF TABLES ..... **XI**

CHAPTER 1 GENERAL INTRODUCTION **1**

1.1 OVERVIEW OF SPEECH MOTOR CONTROL 1
1.1.1 Feedforward control vs. feedback control 4
1.1.2 The DIVA (directions into velocities of articulators) model 5
1.2 AUDITORY FEEDBACK IN SPEECH MOTOR CONTROL 7
1.2.1 Pitch perturbation 8
1.2.2 Formant perturbation 9
1.2.2.1 Unpredictable formant perturbation *12*
1.2.2.2 Predictable formant perturbation *12*
1.3 THE TARGETS OF SPEECH MOTOR CONTROL 17
1.4 NEURAL BASIS OF SPEECH MOTOR CONTROL 19
1.4.1 Cerebellum 19
1.4.2 Motor areas 21
1.4.2.1 primary motor cortex *21*
1.4.2.2 premotor cortex and supplementary motor areas *24*
1.4.3 Auditory areas 27
1.5 THE DUAL PATHWAYS OF SPEECH PROCESSING 28
1.5.1 The Ventral "What" Stream, from Sound to Meaning 30
1.5.2 The dorsal "How" Stream, from Sound to Action 32
1.6 THE ROLE OF MOTOR SYSTEM IN SPEECH PERCEPTION 34
1.6.1 Motor theory of speech perception 35
1.6.2 Neurobiology evidence of motor contribution 36
1.7 HEMISPHERIC LATERALIZATION OF AUDITORY PROCESSING 41
1.8 PROCESSING TONAL LANGUAGES 43
1.8.1 Hemispheric lateralization of tone in tonal languages 43
1.8.2 Tonal versus segmental units 46
1.8.3 The asymmetry of motor contributions during auditory speech processing 48
1.9 AIMS AND SCOPE OF THE THESIS 49
1.9.1 Study 1: Speech adaption: behavioural testing 49
1.9.2 Study 2: Speech adaption: TMS interference 50
1.9.3 Study 3: Asymmetry of auditory-motor speech processing: EEG and TMS 50

CHAPTER 2 METHODOLOGY **53**

2.1 SUMMARY 53
2.2 TRANSCRANIAL MAGNETIC STIMULATION 53
2.2.1 Monophasic versus biphasic stimulation 54
2.2.2 Single-pulse TMS over the primary motor cortex 56
2.2.2.1 The size and latency of TMS-induced MEPs. *57*
2.2.2.2 Defining the threshold for corticomotor excitation: active motor threshold and resting motor threshold *61*
2.2.2 Repetitive transcranial magnetic stimulation (rTMS) 62
2.2.3 Effect of low-frequency rTMS on cortical excitability of tongue representation 64
2.2.3.1 Methods *65*
2.2.3.2 Results *67*
2.2.3.3 Conclusion *68*
2.2.4 Effect of low-frequency rTMS on cortical excitability in Mandarin speakers 68
2.2.4.1 Methods *69*
2.2.4.2 Results *70*
2.2.4.3 Conclusions *71*
2.3 ELECTROENCEPHALOGRAPHY (EEG) 71
2.3.1 Mismatch negativity (MMN) 72

CHAPTER 3 ALTERED AUDITORY FEEDBACK INDUCES COUPLED CHANGES IN FORMANT FREQUENCIES DURING SPEECH PRODUCTION **77**

3.1 ABSTRACT 77
3.2 INTRODUCTION 78
3.3 METHODS 82
3.3.1 Participants 82
3.3.2. Apparatus 83
3.3.3 Procedure 83
3.3.4. Real-Time Feedback Alteration 85
3.3.5 Auditory analysis 86
3.3.6 Statistical analysis 86
3.3.7 Unplanned Analyses 88
3.4 RESULTS 89
3.4.1 Experiment 1: F1 shift of +110 mel 89
3.4.2 Experiment 2: F2 shift of -110 mels 91
3.4.3 Comparison of Experiment 1 and 2 94
3.5 DISCUSSION 97

CHAPTER 4 DISRUPTION OF SPEECH MOTOR ADAPTATION WITH REPETITIVE TRANSCRANIAL MAGNETIC STIMULATION OF THE ARTICULATORY REPRESENTATION IN PRIMARY MOTOR CORTEX 105

4.1 ABSTRACT 105
4.2 INTRODUCTION 106
4.3 METHOD 109
4.3.1 Participants 109
4.3.2 Procedure 109
4.3.3 Real-Time Feedback Alteration 111
4.3.4 Transcranial Magnetic Stimulation 112
4.3.5 Electromyography recordings 114
4.3.6 MEPs Analysis 115
4.3.7 Analysis of speech data 115
4.3.8 Statistical Analysis 116
4.4 RESULTS 117
4.4.1 Effect of rTMS on cortical excitability 118
4.4.2 Effect of rTMS on speech adaptation 119
4.5 DISCUSSION 125

CHAPTER 5 ASYMMETRY OF AUDITORY-MOTOR SPEECH PROCESSING IS DETERMINED BY LANGUAGE EXPERIENCE 135

5.1 ABSTRACT 135
5.2 INTRODUCTION 136
5.3 MATERIALS AND METHODS 139
5.3.1 Participants 139
5.3.2 Procedure 140
5.3.3 Stimuli 141
5.3.4 Transcranial Magnetic Stimulation 143
5.3.5 Electromyography (EMG) recordings 144
5.3.5.1 Electroencephalographic (EEG) recording parameters 144
5.3.5.2 EEG data Analyses 145
5.3.6 Statistical analyses 147
5.4 RESULTS 148
5.4.1 Effects of TMS-induced disruption in the right speech motor cortex 148
5.4.2 Effects of TMS-induced disruption in the left speech motor cortex 154
5.6 DISCUSSION 161

CHAPTER 6 GENERAL DISCUSSION 168

6.1 REAL-TIME FORMANT PERTURBATION 168
6.1.1 Main findings of study 1 and study 2 168
6.1.2 Develop standards to test and report speech adaptation induced by real-time formant perturbation 169
6.1.3 Clinical applications of real-time perturbation paradigm 172

6.2 THE EFFECTS OF TMS OVER M1 ON THE NEURAL ACTIVITIES IN FUNCTIONALLY CONNECTED REMOTE AREAS 174
6.3 HEMISPHERIC ASYMMETRY DURING AUDITORY SPEECH PROCESSING 175
6.3.1 Main findings of study 3 175
6.4 TOP-DOWN INFLUENCES DURING AUDITORY SPEECH PROCESSING 176
6.5 IMPLICATIONS FOR CURRENT SPEECH PRODUCTION AND PERCEPTION MODELS 177
6.5.1 speech production model 177
6.5.2 speech perception models 179
6.6 CONCLUSION 181
APPENDICES 183
APPENDIX 1: STUDY 1 RESULTS 184
APPENDIX 2: STUDY 2 RESULTS 188
REFERENCES 192

List of Figures

Figure 1-1 The DIVA model of speech acquisition and production (Tourville & Guenther, 2011) 6
Figure 1-2 Experimental set-up (adapted from Rochet-Capellan & Ostry, 2011). 11
Figure 1-3 Homunculus of primary motor cortex 23
Figure 1-4 The dual-stream model of speech processing (from Hickok & Poeppel, 2007). 30
Figure 2-1 Time-course of monophasic (left) and biphasic (right) pulses (modified from Sommer et al., 2002). 56
Figure 2-2 Neurophysiological basis of the motor evoked potential (Groppa et al., 2012). 57
Figure 2-3 Using TMS to elicit MEPs in the lip muscle (adapted from Möttönen et al., 2014). 59
Figure 2-4 TMS coil orientation adapted from Brasil-Neto et al., 1992). 60
Figure 2-5 Effect of rTMS on cortical excitability of tongue motor cortex. 67
Figure 2-6 Effect of rTMS on cortical excitability in Mandarin speakers. 71
Figure 3-1 Vowel production. 82
Figure 3-2 Experimental procedure. 85
Figure 3-3 Experiment 1: F1 shifted upwards by 110 mel. Change in normalized formant frequencies during speech production. 91
Figure 3-4 Experiment 2: F2 shifted downwards by 110 mel. 94
Figure 3-5 Comparison of changes in vowel production in Experiment 1 and Experiment 2. 96
Figure 4-1 Vowel spaces and experimental procedure. 111
Figure 4-2 Effect of rTMS on cortical excitability. 118
Figure 4-3 Changes in normalized F1 production (z-score) related to sensorimotor learning and rTMS location. 120
Figure 4-4 Changes in normalized F2 production (z-score) related to sensorimotor learning and rTMS location. 122
Figure 4-5 Coupled F1-F2 changes in vowel space for production during F1 feedback perturbation. 124
Figure 5-1 The trajectories of fundamental frequency (F0) of the three syllables used in current study. 142
Figure 5-2 Effect of rTMS over right speech motor cortex on MMN responses elicited by the phoneme (A) and tone (B) changes in tonal and non-tonal language speakers. 151
Figure 5-3 Comparison of amplitudes of MMN responses elicited by the phoneme (A) and tone (B) changes in tonal and non-tonal language speakers. 153
Figure 5-4 Effect of rTMS over left speech motor cortex on MMN responses elicited by the phoneme (A) and tone (B) changes in tonal and non-tonal language speakers. 157
Figure 5-5 Comparison of amplitudes of MMN responses elicited by the phoneme (A) and tone (B) changes in tonal and non-tonal language speakers. 158

List of Tables

Table 4-1 117
Table 5-1 159
Table 5-2 160

Chapter 1 General Introduction

Speech is one of the main modes of communication used by humans to convey language information.

In speaking, speakers perform a variety of speech acts to achieve different goals, such as asking, warning, persuading and comforting. Speakers can also change articulation, pitch, loudness, rhythm, tempo and other paralinguistic aspects of vocalization to express different meaning.

Overall, speech is a complex vocal communication system that allows human beings to make connections with each other. Speech also enables us to build knowledge and develop cognitive and mental capacities through time.

Speech is typically divided into two processes: production and perception, respectively the acts of movement to articulate and listening to recognize and understand. Understanding how the motor and auditory systems interact during both speech production and perception is the overarching goal of this thesis.

1.1 Overview of Speech motor control

How are spoken words produced? This seemingly straightforward question has a fascinating but complicated answer. In normal fluent conversation, people produce roughly ten or twelve phonemes, four syllables, and two or three words each second. These words are selected from between 50 and 100 thousand

words contained in the huge repository of a normal adult known as the mental lexicon (Levelt, 1999). Despite this, speech production is very accurate with errors occurring less than once or twice in every 1000 words (Garnham, Shillcock, Brown, Mill, & Cutler, 2011).

Speech production is a multi-step process by which thoughts, concepts, feelings and so forth are converted into articulation of words. According to the model of Levelt and colleagues (Levelt, 1999), the first step of speech production is “conceptual preparation” which involves the activation of a lexical concept. This is followed by “lexical selection” where appropriate lemmas (the canonical form, dictionary form, or citation form of a set of words) are selected for the intended concept. The next step is “morphological encoding”, which is the process of transforming lemmas into the morphemes. For example, the lemma “play” may be transformed into the morphemes “play” and “ing” to show an ongoing status. The selected morphemes are then broken down into syllables to be produced in overt speech. The following step is “phonetic encoding”. This involves the activation of articulatory gestures dependent on the syllables selected, creating a phonetic gestural score. The score then specifies which motor acts are to be executed in order to generate the syllable (“articulation”). This thesis focuses on speech motor control, the last two stages of this model.

Articulatory control is a highly complicated motor behavior, which requires pulmonary pressure provided by the lungs (respiration), vibration of the vocal folds in the larynx (phonation) and selection of resonant frequencies by articulators (articulation), including the jaw, lips, tongue. At least 100 different

muscles are located in chest wall, larynx, velopharynx, jaw, tongue, lips, and other speech-related organs, and each muscle includes around 100 different motor units that comprise motor neurons and related muscle fibers. In addition, fluent speech production requires integration with simultaneous auditory and somatosensory feedback, which adds additional layers of complexity when investigating speech motor control. How this coordination of movement is achieved given the considerable number of muscles and degrees of freedom of movement, and how the neural controller of articulation integrates such complicated auditory and somatosensory feedback when creating commands to send to the musculature are largely unresolved problems in the field of speech motor control.

One model of speech motor control borrows from models of visuomotor control of eye and limb movements. Accordingly, the term speech motor control refers to the control of movements of an articulator system (plant), such as a human vocal tract to achieve the desired output (acoustic or motor goal). The plant receives input from the controller, a system to generate motor commands to achieve desired movement in the plant. For speech production, the controller consists of a large number of cortical and subcortical regions of the brain involved in speech motor control. Here, I will mainly focus on the controller, the system that that receives the speech target (plan) and sends inputs to the plant in order to achieve the desired output.

1.1.1 Feedforward control vs. feedback control

Broadly speaking, the problem of speech motor control has been addressed by two types of control systems.

Feedforward control, a control system that does not rely on any outputs from the plant for maintaining control. Instead, it generates a pre-defined motor command for the desired output. This system is also called open-loop control system, as the signals generated in the system are totally feedforward. Feedforward control is often used to control rapid movements that end before any sensory information can be processed. In the case of speech production, a feedforward control system is one that passes a control signal (motor commands) for a syllable (or phoneme) completely from memory, without taking account of the complex auditory and somatosensory feedback that results from these commands.

Feedback control refers to a system that makes use of outputs from the plant when determining the motor command. The sensory feedback, which conveys the sensory state information (actual output) is compared with the desired output to generate an error signal that is then sent out as a corrective motor command. This type of controller is also referred to as a closed-loop controller, since the output signal is fed back to the input and forms a loop from the motor command to the error signal. Feedback control systems have many advantages over feedforward control. One apparent advantage is the fact that the use of sensory feedback means that the system can naturally deal with unstable or unpredictable environments, including external perturbations or internal

variations. However, poorly designed closed-loop systems can also cause the system to lose control or oscillate indefinitely.

Overall, both feedback and feedforward control systems, have their own strengths and weaknesses. Feedforward control can be completed quickly but is unstable as it cannot compensate for external or internal perturbations to the system, while feedback control is relatively stable when facing perturbations, but is slow and becomes less accurate when sensory feedback is delayed and noisy, as in most neural systems. Therefore, a fine motor control system, such as that needed for the complexity of speech production, may make use of both feedforward and feedback control.

1.1.2 The DIVA (directions into velocities of articulators) model

Among the speech motor control models described in the literature, the DIVA (directions into velocities of articulators) model (Tourville & Guenther, 2011) is one of the most influential. According to the DIVA model (see Figure 1-1), production of a phoneme or syllable starts with activation of a cell/node representing a “speech sound” in a speech sound map, which is assumed to be located in the left ventral premotor cortex (vPMC). Once a speech sound map cell has been activated, signals from the speech sound map are conveyed to the articulator map, which is putatively located in the motor cortex, via two control systems: a feedforward control system and a feedback control system. The feedback control system is further divided into two components: an auditory feedback control subsystem and a somatosensory feedback control subsystem.

The feedback control subsystems in the DIVA model play an important role in maintaining stable speech by detecting and correcting differences between the desired sensory signal output and the actual sensory feedback. First, activation of the speech sound map also sends signals (efference copy) to the auditory target map and the somatosensory target map, which provide readouts of learned sensory targets for that sound (corollary discharge). The sensory targets for the speech sound are then compared to the current sensory states, available in the auditory state map or somatosensory state map. When a mismatch between the sensory targets and current sensory states is detected, error signals would be generated in the auditory/ somatosensory error map, which are hypothesized to lie in the higher-order sensory cortex. These error signals are then sent to the feedback control map in right ventral premotor cortex (vPMC) to create a corrective motor command.

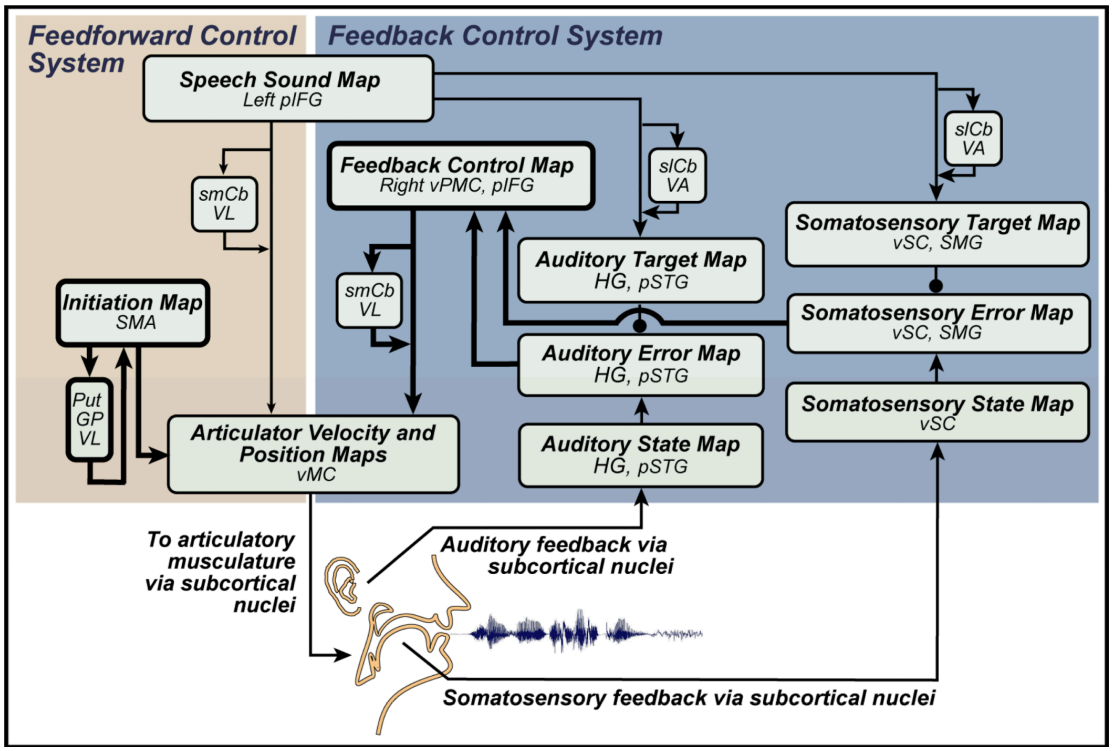


Figure 1-1 The DIVA model of speech acquisition and production (Tourville & Guenther, 2011). **Reprinted with permission from Taylor & Francis, copyright (2011).**

In the model, both feedforward (brown portion) and feedback control (blue portion) subsystems are involved and integrated. Abbreviations: GP = globus pallidus; HG =Heschl's gyrus; pIFg = posterior inferior frontal gyrus; pSTg = posterior superior temporal gyrus; Put = putamen; slCB = superior lateral cerebellum; smCB = superior medial cerebellum; SMA = supplementary motor area; SMG = supramarginal gyrus; VA = ventral anterior nucleus of the cerebellum; VL = ventral lateral nucleus of the thalamus; vMC = ventral motor cortex; vPMC = ventral premotor cortex; vSC = ventral somatosensory cortex.

1.2 Auditory feedback in speech motor control

It is uncontroversial that auditory feedback is particularly important in speech production; without it, fluent speech production cannot be maintained, or even learned. Imagine you are having a bad telephone connection. A slight delay or tiny noise will significantly affect your speaking fluency. The disruptive effect of altered auditory feedback is just one source of evidence for the important role of auditory feedback in fluent speech production. Pre-lingually deaf children, who have difficulties with speech articulation is another (Svirsky, Teoh, & Neuburger, 2004). Auditory feedback is also crucial for the maintenance of speech intelligibility, which can deteriorate in post-lingually deaf adults (Waldstein, 1990).

In recent decades, using real-time perturbation techniques, more subtle manipulations have been applied to investigate the contributions of auditory feedback to speech motor control. The majority of perturbation studies involve modifications of fundamental frequency (pitch perturbation experiments) or formant frequencies (formant perturbation experiments).

1.2.1 Pitch perturbation

The first study to explore the compensatory responses induced by real-time pitch perturbation was conducted by Elman (1981), where he electronically increased all frequencies of the speech signal by around 10% and fed this back to the speakers, using a device called the Lexicon Varispeech. In response to this pitch perturbation, speakers decreased the pitch of their speech productions. Such compensatory responses were unrelated to a speaker's awareness of the pitch shift.

More subtle manipulations were applied in later studies which largely improved our understanding of the pitch-shift response. Burnett, Freedland, Larson, & Hain (1998) conducted three experiments, in which the effects of different stimulus variables on pitch shift responses were carefully examined. In experiment 1, it was found that the magnitude or latency of pitch shift response was insensitive to either the absolute level of feedback intensity or the presence of pink masking noise. In experiment 2, the magnitude of perturbation level (pitch shift) was systematically manipulated, increasing from 25 to 300 cents. The results showed that increasing the perturbation magnitude reduced the proportion of following responses (i.e. responses that were in the same direction as the shift). However, the average magnitude (approximately 30 cents) of compensatory responses (i.e. responses that were in the opposite direction to the shift) was not significantly affected by perturbation magnitudes. In the final experiment, pitch-shift stimulus durations were increased from 20 to 500 ms and corresponding responses were examined. They found that durations from 20 to 100 ms led to no differences in

the pitch-shift response, while durations between 100 and 500 ms led to longer durations and larger magnitude of responses.

The pitch-shift response has been investigated in native Mandarin speakers. Mandarin is a tonal language in which the pitch or pitch contour can change the meaning of a word. In the study of Jones and Munhall (2002), native Mandarin speakers received abrupt pitch feedback perturbation when producing a Mandarin syllable /ma/ with a high, flat tone. Rapid compensations in response to the perturbation were found in almost every participant. Responses matched those found in English speakers, regardless of conscious awareness. A later study further examined the pitch-shift response in Mandarin speakers during dynamic tone productions by introducing pitch feedback shift when they produced disyllabic sequences /ma ma/ with different tonal combinations at a natural speaking rate. Results indicated dynamic changes in the fundamental frequency (F0) led to increased response magnitudes and decreased response latencies (Xu, Larson, Bauer, & Hain, 2004).

1.2.2 Formant perturbation

Pitch perturbation experiments have improved our understanding of auditory contributions during speech motor control. In English or most other European languages, however, pitch is a suprasegmental component that is mainly used to express a speaker's emotions or intentions and is not an essential feature of the meaning of a word. Formant frequencies, in contrast, are closely associated with phoneme identity in all languages. Formants are distinctive frequency components evident in the acoustic signal reflecting the size and shape of different chambers in the vocal tract i.e. their resonances. The first two formants

(F1 and F2) in a vowel sound contain most of the acoustical energy and are sufficient to disambiguate vowel sounds. In addition, the frequencies of F1 and F2 are closely related to tongue positions. For example, F1 frequency mainly reflects the tongue height in the oral cavity, while F2 frequency reflects the front-back position of the tongue body (Ladefoged, 2001).

Twenty years ago, to question feedforward control in speech, Houde & Jordan (1998) first introduced the paradigm in which the first two formant frequencies of whispered monosyllabic words containing the vowel /ɛ/ were perturbed in a consistent way and fed back to speakers in real time. They found that speakers learned to adjust their speech production to compensate for feedback perturbations that change the vowel's perceived phonetic identity. Such compensatory responses were maintained when participants spoke with auditory feedback masked by noise, suggesting modification of stored feedforward motor commands for the words.

Here, I will give a relatively detailed review about the formant perturbation, as it is the experimental paradigm used in study 1 (Chapter 3) and study 2 (Chapter 4).

In study 1 (chapter 3), real-time formant perturbation was used to examine the compensatory responses of native English speakers when they received feedback involving single formant perturbation. A typical experimental set up is illustrated in Figure 1-2. As shown in the figure, when the participant says “head”, if the first formant (F1) is shifted up, moving the bottom green line to the red line

in the right spectrogram, the participant would hear a sound more similar to “had”. After receiving such shifted feedback, the participant can compensate by changing their own productions (e.g. lowering their F1 frequencies). In most of studies, auditory feedback of the speech produced is provided for participants through headphones in almost real time (i.e. with delays of 20 ms or less). Therefore, in most of cases, participants do not notice the perturbation. By checking the adaptive movements induced by such real-time perturbations, we can gain insight into how speech movements are controlled.



Figure 1-2 Experimental set-up (adapted from Rochet-Capellan & Ostry, 2011).

The participant reads words aloud when the words appear on the screen and hears their altered voice through headphones. The first formant of the word “head” was shifted up (moved green line to the red line in the right spectrogram), which made the vowel sound more similar to that in “had”.

Experiments employing the formant feedback perturbation can be divided into those that use either unpredictable or predictable alterations.

1.2.2.1 Unpredictable formant perturbation

Unpredictable formant perturbation is mainly used to investigate feedback control. In a previous study, feedback of F1 production was unexpectedly shifted up or down when participants produced the steady-state isolated vowel /ɛ/ in consonant-vowel-consonant (CVC) contexts (as in “head”), which resulted in the percept of speech sounding more like /æ/ (as in “had”) or /ɪ/ (as in “hid”) respectively (Purcell & Munhall, 2006b). Results showed that participants responded to the unexpected formant shifts by changing their formant productions in the opposite direction to the shift applied. Compensation responses started approximately 460 ms after stimulus onset. Compensation for the perturbation was incomplete (approximately 16% and 11% of the applied shift, upward or downward, respectively). These results are similar to those seen for pitch perturbation (see above) and suggest that auditory feedback control is similar for both pitch and formant production.

1.2.2.2 Predictable formant perturbation

Formant perturbation can be also be applied predictably, as was the case in the original study by Houde & Jordan (1998). Applying a consistent formant perturbation to auditory feedback over a number of utterances will result in a sustained adjustment of the motor commands originating from the feedforward control system. According to the DIVA model, the feedforward control system continuously monitors the corrective commands produced by the feedback control system, and gradually integrates such corrective commands into later feedforward commands (Tourville & Guenther, 2011). Such a process is referred to as sensorimotor adaptation. The learned changes in production reduce the

perceived auditory error and persist for a period even after the perturbation is turned off, which is often called the aftereffect. Aftereffects are usually smaller than adaptation measured under noise-masked conditions (Purcell & Munhall, 2006a). Therefore, it is suggested that compensatory changes measured in the course of adaptation when the perturbation is applied, are the result of both feedback and feedforward control (Guenther, 2016).

By using this paradigm, Purcell & Munhall (2006a) characterized sensorimotor adaptation and aftereffects induced by predictable formant perturbation. Results indicated that when F1 feedback was shifted up or down (±200 Hz), participants compensated by changing F1 to the opposite frequency direction from baseline. The average magnitude of maximal compensations in F1 was around 60 Hz (i.e. about 30% of the applied shift), which did not fully compensate for the perturbation. Similar magnitudes of compensation were reported in later work, regardless of the delay in the auditory feedback (Mitsuya, Munhall, & Purcell, 2017) or the types of headphone (circumaural headphones vs. insert earphones) used (Mitsuya & Purcell, 2016).

When feedback perturbation is removed (i.e. feedback is returned to normal), F1 returns gradually to baseline, showing an exponential de-adaptation pattern (also called an aftereffect; Purcell & Munhall, 2006a). This de-adaptation pattern is observed for both upwards and downwards shifts in feedback. The de-adaptation pattern is unaffected by the duration of the hold phase in which feedback shift is maintained at the maximum level.

The effect of conscious awareness and strategy on compensatory responses to feedback perturbations has also been investigated (Munhall, MacDonald, Byrne, & Johnsrude, 2009). Three groups of participants were given different instructions when the formants in their speech were perturbed in real time: (1) control, in which no information about feedback shift was given to participants; (2) ignore headphones, in which participants were told that some changes might be applied to their voice but try to ignore the auditory feedback from the headphones; (3) avoid compensation, in which participants were fully-informed about the shift manipulation and were instructed to avoid any compensation. Results showed that participants produced robust and consistent compensations to the feedback perturbation in all conditions. There were no significant differences in the magnitudes of compensatory changes among groups. These results indicated that sensorimotor adaption induced by real-time feedback perturbation is relatively automatic and implicit and is not easily modulated by awareness or application of any explicit strategy.

Using real-time formant perturbation, sensorimotor adaptation has been investigated cross-linguistically. Compensatory responses of three language groups were examined (Mitsuya, MacDonald, Purcell, & Munhall, 2011): (1) native English speakers producing an English vowel /ɛ/; (2) native Japanese speakers producing a Japanese vowel /e/; and (3) native Japanese speakers learning English, producing /ɛ/. All three groups showed similar compensatory patterns when F1 feedback was decreased; however, when F1 feedback was increased, the average adaption of both Japanese groups was not as large as that of native English speakers. This asymmetry is taken as evidence for the

influence of phonological processing in adaptation, that is, how auditory feedback is perceptually processed and applied to formant control. For English speakers, both increased and decreased F1 feedback may have resulted in a comparable perceptual change from the intended vowel (i.e. sounding more like /I/ or /æ/), which in turn, induced a comparable amount of compensation. For Japanese speakers, however, the production of vowel /æ/ is unstable and further away from /ɛ/ in vowel space. Therefore, when the perturbation (F1 downward shift) aimed to move the vowel closer to /æ/, the sound might have been heard as an acceptable instance of the intended vowel /ɛ/. Similar results also were found between native English and French speakers, when they received decreased F2 feedback (Mitsuya, Samson, Ménard, & Munhall, 2013). Compared with English speakers, French speakers made larger compensations in F2 and modified the third formant (F3) in their vowels even though F3 was unperturbed. In short, the authors suggest that results from these cross-linguistic studies indicate that the amount speakers compensate is affected by the perceptual consequence of the perturbation, reflecting language-specific phonemic processes.

The influence of higher-level linguistic knowledge, such as lexical status, on speech motor control has been investigated as well. In the study of Bourguignon, Baum, & Shiller (2014), participants were instructed to produce the vowel /ɛ/ in either monosyllabic real words (e.g. less) or pseudo-words (e.g. kess), while they received a downward shift in F1 feedback of their speech. Such F1 alteration made the vowel sound more similar to /ɪ/ (as in “hid”), which could result in different changes to different words dependent on their lexical status. Results indicated that the magnitude of compensation varied when the feedback

alteration affected the lexical status of the words; compensatory responses were significantly larger when a downshift in F1 changed perception of an utterance from a pseudo-words to a real word (e.g. kess-kiss), than when the same F1 alteration resulted in the reverse change (i.e. from real to pseudo-words; e.g. less-liss). This study successfully demonstrated the interaction between sensory and lexical information in guiding speech motor control.

Many studies have explored the generalizability of sensorimotor adaptation. In the original study of Houde & Jordan (1998), participants were instructed to produce words which were randomly selected from two different sets: a set of training words and a set of test words. The training words consisted of four consonant-vowel-consonants (CVC) words, “pep”, “peb”, “bep”, “beb”, containing the same vowel /ɛ/. These words were presented to participants with (perturbed) auditory feedback or with whispering or masking noise that blocked the auditory feedback. There were seven test words: (1) “gep”, “peg”, “teg”, which contained the same vowel /ɛ/ as the training words and were used to test whether the adaptation of vowel /ɛ/ can transfer between different phonetic contexts; and (2) “pip”, “peep”, “pap,” and “pop”, which were included to test if the adaptation of the vowel /ɛ/ can transfer to other vowels. Results showed that the effect of adaption successfully generalized across phonetic contexts and even to different vowels. Consistent results were found in a later study, in which similar experimental procedures were used and showed that adaptation on the vowel /ɛ/ significantly transferred to the same vowel with different phonetic contexts and to different vowels in the same context (Villacorta, Perkell, & Guenther, 2007). However, it is noteworthy that generalization was incomplete. The compensation

in F1 and F2 of the trained vowel /ɛ/ was approximately 54% of the applied shift, while the compensation of the transfer word with the same vowel was around 35% (Houde & Jordan, 1998). By examining compensatory responses of native Mandarin speakers in response to F1 perturbation during their production of the triphthong /iau/, Cai, Ghosh, Guenther, & Perkell (2010) found the strength of the generalization depended on the similarity between trained and tested utterances. For example, increased distance in the vowel space led to reduced generalization, i.e. there was very weak transfer to /iou/.

1.3 The targets of speech motor control

In order to better understand speech motor control, it is necessary to understand the speech target or goal. In other words, what exactly is it that the speech motor system aims to achieve? In the speech motor control literature, the term “target” or goal is typically used to refer to the input to the controller.

It seems to be obvious that the acoustic signal is the only or primary target of speech production, as speech is mainly used by people to exchange information with others. In other words, the controller aims to generate a desired acoustic trajectory in production of a phoneme, syllable, or word. However, previous studies have demonstrated that speakers also compensate for somatosensory feedback perturbation caused by mechanical load applied to the jaw, when auditory feedback was unaltered (Tremblay, Shiller, & Ostry, 2003), which suggests the existence of a somatosensory target. Such a notion aligns with theories proposing that instead of the acoustic signal, the primary target of speech production is actually the articulatory gesture (e.g., Browman & Goldstein,

1990a, 1990b; Saltzman & Munhall, 1989; Fowler, Rubin, Remez, & Turvey, 1980). These gestures involve applying some particular constrictions on the locations and degrees of parts of the vocal tract, such as the height and front-back position of the tongue body in vowel production. However, these two competing views regarding the speech target (auditory vs. gestural) are often difficult to distinguish from each other experimentally, as the locations and degrees of articulators are closely related to the acoustic signal. For example, F1 frequency is mainly dependent on tongue height, while F2 frequency is mainly dependent on the front-back position of the tongue body (Ladefoged, 2001)

More recent understanding is that both auditory and somatosensory targets play crucial roles in speech control in the mature system. For example, the DIVA model (Tourville & Guenther, 2011) integrates these two hypotheses into a unified model of speech production in which the contributions of both auditory and somatosensory input are fully recognized (see Figure 1-1). However, the DIVA model gives different weight to the auditory target and somatosensory target in the developing speech system. In the model, the auditory target for an intended word or phoneme typically plays a major role in generating the feedforward motor commands, and also plays a role in creating the somatosensory target for the sound. For example, when a developing speaker practises production of a newly learned word, the speaker's brain establishes a somatosensory target for the word to augment the auditory target.

1.4 Neural basis of speech motor control

Our understanding of the neural basis of speech production owes much to the pioneering work of neurologists like Dax, Broca, Lichtheim, and Wernicke, who studied brain lesions in patients who lost the ability to produce fluent speech. With the advent of neuroimaging, our picture of the neural underpinnings of speech production has become more detailed and complex. Undoubtedly, fluent speech production is supported by a large number of cortical and subcortical regions of the brain. Here, I will focus on areas involved in articulatory motor control relevant for this thesis, these include the cerebellum, motor cortex and auditory cortex.

1.4.1 Cerebellum

The cerebellum is a softball-sized structure attached to the bottom of the brain, which is involved in a wide range of tasks, including sensorimotor control, language, emotional processing, and spatial and higher-level cognitive functions (see Ackermann, Mathiak, & Riecker, 2007; Stoodley & Schmahmann, 2009 for reviews).

Experimental research in animals (e.g. Snider & Eldred, 1951), brain imaging studies in humans (e.g. Bushara et al., 2001; Grodd, Hülsmann, & Ackermann, 2005) and clinical reports (e.g. Schoch, Dimitrova, Gizewski, & Timmann, 2006) have suggested that there is topography of motor function within the cerebellum (see Manni & Petrosini, 2004 for a review). For example, a recent study investigated topographical organization in posterior cerebellum by electrically stimulating different parts of the posterior cerebellum of 20 human patients

undergoing surgery (Mottolese et al., 2013). They found that Vermal lobule VI was associated with head movement while hemispheric lobule VI was associated with face/mouth movement. Upper-limb representations were more widely distributed. In the speech domain, increased activity in both inferior and superior posterior lobe was reported during speech production in several neuroimaging studies (Bohland & Guenther, 2006; Golfinopoulos et al., 2011; Tourville, Reilly, & Guenther, 2008). Overt speech production resulted in significantly increased responses in the superior cerebellar cortex bilaterally and in the right inferior cerebellum as well. Increased activity was also found in the superior and inferior cerebellum when participants produced utterances with altered auditory (Tourville et al., 2008) and somatosensory (Golfinopoulos et al., 2011) feedback compared with unaltered feedback.

In terms of feedforward and feedback control, research in other motor domains indicates a crucial role for the cerebellum in feedforward control, but relatively little involvement in feedback control (Morton & Bastian, 2006; Smith & Shadmehr, 2005; Tseng, Diedrichsen, Krakauer, Shadmehr, & Bastian, 2007). The role of the cerebellum in feedforward control and feedback control during speech production was recently tested by comparing speech adaptation between normal controls and patients with cerebellar degeneration (Parrell, Agnew, Nagarajan, Houde, & Ivry, 2017). Compared with normal controls, patients were impaired in adjusting their feedforward control system, showing decreased compensatory responses to a predictable perturbation of auditory feedback. In contrast, patients showed even greater compensatory responses than controls in response to unpredictable perturbations of auditory feedback, which suggested that the

patients with cerebellar degeneration relied more on sensory feedback. These results confirmed that, as for other forms of motor control, the cerebellum is critical for maintaining accurate feedforward control during speech production, but relatively less important in feedback control. Recently, non-invasive brain stimulation was used to examine the role of the cerebellum in speech adaptive learning and compare this with the role of the motor cortex (Lametti, Smith, Freidin, & Watkins, 2018). Compared with sham stimulation, anodal (excitatory) tDCS over both the cerebellum and the motor cortex significantly improved speech motor learning in terms of increased compensatory responses in F1 to an F1 perturbation. Interestingly, the group who received anodal tDCS over the cerebellum showed no change in the second formant (F2), which also showed adaptive responses in both the sham group and the group who received motor cortex stimulation. This pattern of findings suggested that motor cortex might be involved in a more global control of formants to achieve production of a familiar vowel, while stimulation of the cerebellum simply promoted compensatory changes that offset the acoustic error.

1.4.2 Motor areas

Three areas of the cortex are often referred to as motor: primary motor cortex (M1), the premotor cortex (PMC) and the supplementary motor areas (SMA).

1.4.2.1 primary motor cortex

The primary motor cortex (M1) lies immediately anterior to the central sulcus, extending from the posterior sulcal bank onto the gyral surface of the precentral gyrus. Motor representations are somatotopically organized along the primary

motor cortex, from the toe (at its most dorsal extent) to the tongue (ventrally). Furthermore, these cortical representations are distorted relative to the relative sizes of the effects and this distortion is often represented graphically as the motor homunculus (see Figure 1-3). The motor representations of the speech articulators lie primarily in the ventral portion of the primary motor cortex and according to Penfield & Roberts (1937), are organized as follows moving from the dorsal representation towards the Sylvian fissure: lip, vocalization, jaw, tongue and swallowing. Findings from more recent neuroimaging studies provide further detail of the motor homunculus involved in speech production. For example, some studies suggest two laryngeal representations involved in vocalisation instead of one; one located dorsal to the lip representation and the other at the most ventral extent of the central sulcus (see Belyk & Brown, 2017 for review). Recent work using quantitative mapping of these regions indicates that while the dorsal larynx representation is in primary motor cortex the ventral larynx representation is not (Eichert et al., 2020). Furthermore, the motor representations show some overlap (Grabski et al., 2012; Eichert et al., 2020), which is considered necessary for the precise inter-articulator coordination observed in fluent speech (see Figure 1-3).

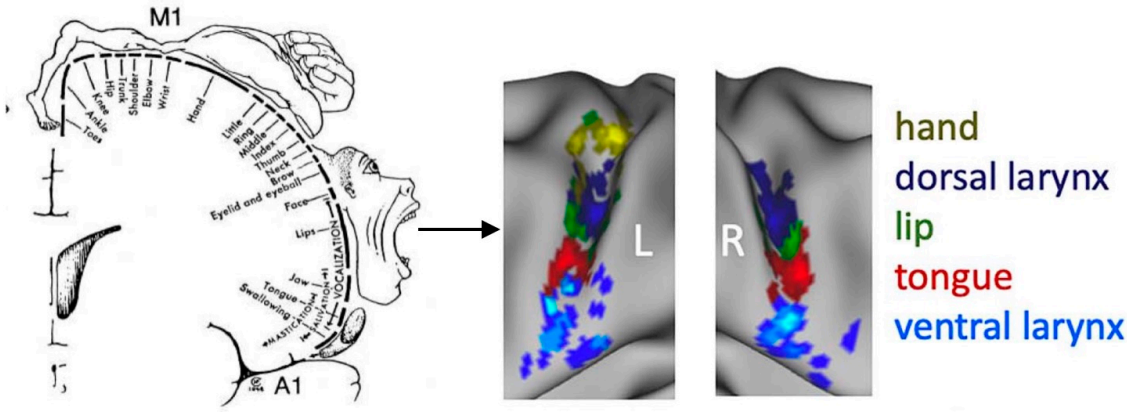


Figure 1-3 Homunculus of primary motor cortex

(Left) Motor homunculus by Wilder Graves Penfield (1950). (Right) Individual activation maxima in a group of 20 participants derived for simple hand (yellow), lip (green), and tongue (red) movements, and vocalization (i.e. larynx activity; dorsal larynx shown in dark blue) (Eichert et al., 2020).

The important role of M1 in generating neural impulses that control the execution of movement might obscure the fact that M1 not only sends outputs to the muscles to generate the movement, but also receives proprioceptive input from the periphery and can influence other cortical areas, directly via corticocortical pathways and indirectly via the corticothalamic pathways.

In contrast to the speech motor adaptation literature, the role of M1 in visuomotor adaptation has been studied extensively. The results from a large number of non-invasive brain stimulation studies suggest that M1 is not involved in initial motor adaptation but is crucial for the retention of learned movements patterns when the feedback return to normal (Galea, Vazquez, Pasricha, Orban de Xivry, & Celnik, 2011; Hadipour-Niktarash, Lee, Desmond, & Shadmehr, 2007; Richardson et al., 2006). For example, disruptive transcranial magnetic

stimulation (TMS) over M1 before adaptive learning did not affect adaptation to a force field, but significantly impaired participants' performances in the same force field 24h later (Richardson et al., 2006). In contrast, a tDCS study found that anodal tDCS applied over M1 during force-field adaptation significantly increased the initial movement trajectory error (in the first 150 ms of movement), suggesting the involvement of M1 in initial adaptation of reaching movement of human participants (Hunter, Sacco, Nitsche, & Turner, 2009). More recently, researchers have started to explore the relationship between neurochemical changes in M1 and sensorimotor learning. Functional magnetic resonance spectroscopy (fMRS) was used to measure γ-aminobutyric acid (GABA) concentrations in left M1 before and after learning. Results showed that there were significant increases in GABA during training, which extended into subsequent rest periods and were correlated with the amount of learning on a subject-level basis (van Vugt, Near, Hennessy, Doyon, & Ostry, 2020).

1.4.2.2 premotor cortex and supplementary motor areas

Aside from the primary motor cortex, two other cortical areas are known to play important roles in motor control and learning: the premotor cortex (PMC) and the supplementary motor areas (SMA). Both send signals to the primary motor cortex as well as to brainstem motor regions.

The SMA is located on the medial surface of the frontal lobe in Brodmann area (BA) 6, occupying the posterior one third of the superior frontal gyrus. The traditionally defined SMA is now regarded as consisting of at least two sub-regions: The supplementary motor area (SMA) proper, which lies caudally,

projecting directly to the primary motor cortex and the spinal cord, and the pre-supplementary motor area (preSMA), which is located in the rostral part, and is densely interconnected with regions of prefrontal cortex and other non-primary motor areas (Bates & Goldman-Rakic, 1993; Lu, Preston, & Strick, 1994; Wang, Isoda, Matsuzaka, Shima, & Tanji, 2005). During the past few decades, there has been increased focus on the important roles of the SMA and pre-SMA in language and speech processing (see Hertrich, Dietrich, & Ackermann, 2016 for a review). Given considerable differences in their connectivity, SMA and pre-SMA display some important differences in their functions during speech production. SMA is thought to serve predominantly simple motor execution, while pre-SMA is found to be involved in the more cognitive aspects of motor behaviour, such as motor plan updating and learning of new motor sequences (see Picard & Strick, 1996 for a review). Previous studies have found that simple speech tasks, such as simple repetition of words (Petersen, Fox, Posner, Mintun, & Raichle, 1988), and the production of over-practised verbal association (Paus, Petrides, Evans, & Meyer, 1993) result in activity in SMA. In contrast, activity in the pre-SMA was observed in association with more complex speech tasks, such as silent verb generation (Wise et al., 1991), oral word production in new conditional associations (Paus et al., 1993), and self-ordered number generation (Petrides, Alivisatos, Meyer, & Evans, 1993).

The premotor cortex is also a motor area lying within the frontal lobe, occupying the lateral surface rostral to the primary motor cortex on the precentral gyrus (part of Brodmann's area 6). PMC receives considerable afference from sensory areas and subcortical regions, and projects to M1 and the brainstem. On the basis of

these connections, PMC, like preSMA, is primarily involved in the motor planning and sensorimotor learning. In the domain of speech, the PMC has been recognized as a key intermediary between sensory and motor cortex, and between higher-level speech areas (Bohland & Guenther, 2006; Houde & Nagarajan, 2011). PMC is bidirectionally and heavily connected to other motor areas (Kandel et al., 2000). Modulation of M1 excitability by ventral PMC stimulation before and during movements has been reported in several studies (e.g. Davare, Lemon, & Olivier, 2008; de Beukelaar, Alaerts, Swinnen, & Wenderoth, 2016). On the basis of these connections, the DIVA model proposed that speech production begins with the activation of a speech sound map, which is located in the left PMC and adjacent inferior frontal cortex (IFC) (see Figure 1-1). In addition to the massive connectivity with M1, PMC is also bidirectionally connected to the auditory and somatosensory areas. Using real-time formant perturbation, a recent tDCS study showed that the rate of speech adaptation induced by predictable formant perturbation was increased by anodal tDCS applied over the left ventral motor cortex (Scott et al., 2020). In the DIVA model, a feedback control map is suggested, which is assumed to transform the auditory and somatosensory error signals into corrective motor commands (Tourville & Guenther, 2011). Based on results from previous neuroimaging studies, the feedback control map has been suggested to lie in the right ventral premotor/frontal cortex (see Figure 1-1). When participants received unpredictable auditory feedback perturbation, speech production related activity increased compared with speech production with unaltered feedback in auditory cortex bilaterally, right motor and premotor cortex, and the right posterior

cerebellum. Similar results were reported in Golfinopoulos et al., (2011), in which brain activity was monitored under normal or altered somatosensory feedback.

Taken together, the PMC is a key intermediary between sensory and motor cortex, which plays a critical role in both feedforward and feedback control. However, it is also worthy to notice that the PMC is also functionally connected with many other cortical and subcortical structures, such as insula and basal ganglia. Therefore, as suggested in Houde & Nagarajan (2011), it is quite possible that the crucial role hypothesized for PMC alone is supported by a large brain network which involves many cortical and subcortical areas.

1.4.3 Auditory areas

As we mentioned before, the critical role of auditory prediction has been well recognized in current models of speech production, such as the DIVA model (Tourville & Guenther, 2011). In DIVA model, they posit that during speech production, a copy of the motor command (efference copy) is sent to the auditory and somatosensory cortex, in which a predicted sensory consequence will be generated (corollary discharge) and compared with the actual speech sound. The DIVA model posits that higher-order auditory areas play a critical role during these processes (see Figure 1-1; Tourville & Guenther, 2011). Evidence from functional Magnetic Resonance Imaging (fMRI) studies using the real-time perturbation paradigm provides support for this hypothesis, in which they found increased activations in auditory areas when participants spoke with delayed (Hashimoto & Sakai, 2003) or altered auditory feedback compared with those who received unaltered feedback (Tourville et al., 2008). A phenomenon, called

speaking-induced suppression, which is commonly observed in studies using electroencephalography (EEG) (Behroozmand & Larson, 2011), magnetoencephalography (MEG) (Houde, Nagarajan, Sekihara, & Merzenich, 2002; Niziolek, Nagarajan, & Houde, 2013), and direct cortical recordings (Chang, Niziolek, Knight, Nagarajan, & Houde, 2013; Chen et al., 2011), provides further support for the existence of corollary discharge in auditory areas. In these studies, the amplitude of N1/M100 responses was found to be decreased during self-produced speech compared with during hearing the speech of others or the same recorded speech sounds played back. N1/M100 is a large, negative-going evoked potential that is thought to mainly reflect neural activation in the thalamo-cortical network, which receives lots of inputs from the auditory cortex, in response to an auditory stimulation (Näätänen & Picton, 1987).

1.5 The dual pathways of speech processing

Now, we shift our attention from speech production to perception. Over the past two decades, it was proposed that, similar to visual processing, auditory processing also involves two pathways: the “Where” pathway, projecting from auditory areas into posterior parietal cortex, which is hypothesized to deal with the spatial information in acoustical signals; and the “What” pathway, projecting from the posterior temporal cortex to anterior temporal cortex, which is assumed to subserve the recognition of sound patterns (Kaas, Hackett, & Tramo, 1999; Rauschecker, 1998; Romanski et al., 1999; see Rauschecker & Scott, 2009 for a review).

Around the same time, a dual-stream model of auditory speech processing was proposed by Hickok and Poeppel (see Figure 1-4; Hickok & Poeppel, 2000, 2004, 2007). The key observations that motivated them to develop the Dual Stream model are from the classic aphasia literature. Patients with focal brain damage to different brain areas showed selective impairment in either speech comprehension (knowing "what" is being said), or repetition (knowing "how" to reproduce heard sounds). For example, patients with transcortical sensory aphasia, which is a kind of aphasia that involves damage to specific regions of the temporal lobe, show severe impairment in understanding the meanings of words and sentences, but they have relatively intact repetition ability (i.e. they can repeat the same words). In contrast, patients with conduction aphasia, have intact speech comprehension, but they are severely deficient at repetition.

Hickok and Poeppel thus claimed that speech processing includes at least two functions. One is the ability to translate the acoustical patterns of speech into a representation of the meaning of spoken sound. The other is to map the acoustical speech signals into articulatory representations, to verbally reproduce the heard sound sounds. Similar to other two pathway models, these two key functions during speech processing are separately supported by two broad processing streams. According to Hickok & Poeppel (2007), the earliest stage of speech processing in the brain involves spectrotemporal analysis, which is achieved in the dorsal superior temporal gyrus (STG) bilaterally. The early stage also involves phonological-level processing, which is hypothesized to be carried out in the middle to posterior portions of the superior temporal sulcus (STS). A weak left-hemisphere bias is suggested during the phonological-level processing.

From here on, the system segregates into two streams: a ventral “what” stream and a dorsal “how” stream.

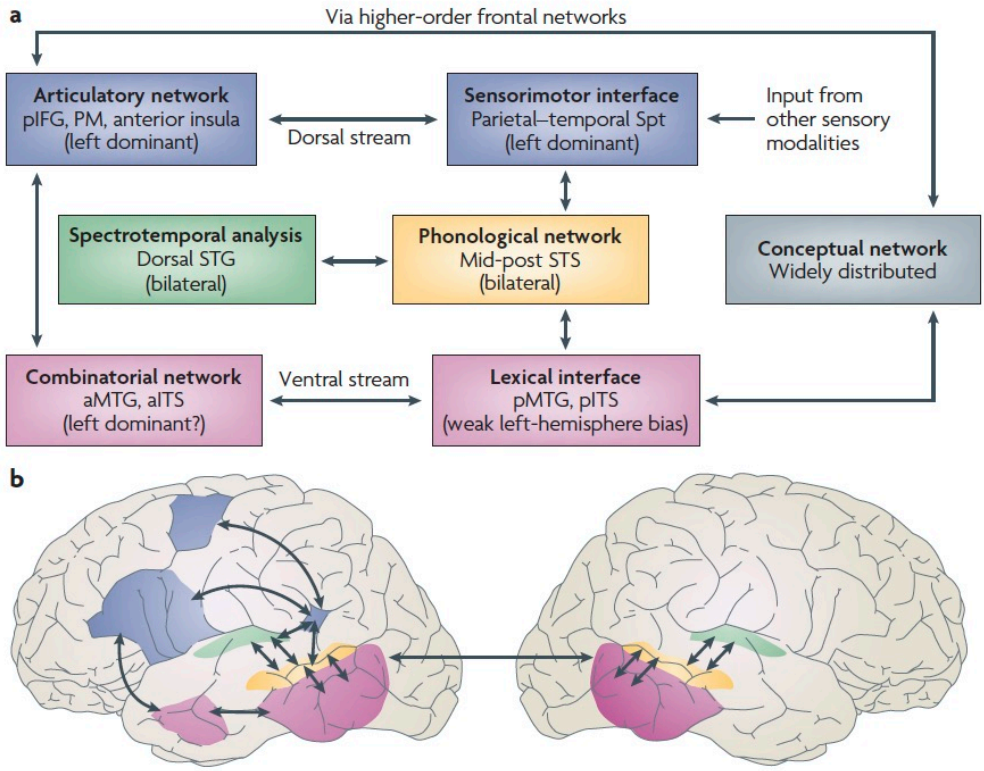


Figure 1-4 The dual-stream model of speech processing (from Hickok & Poeppel, 2007). Reprinted with permission from Nature Publishing Group, copyright (2007).

(a) Schematic diagram of the dual-stream model; (b) Approximate anatomical locations of the dual-stream model components.

1.5.1 The Ventral “What” Stream, from Sound to Meaning

The ventral stream consists of two main functional-anatomical components: the lexical interface and the combinatorial network (Figure 1-4). The lexical interface is considered as a relay station, which transfers the phonological structures of spoken words to their semantic structures. However, how exactly this transfer is achieved remains unclear. In the literature, there is a debate around how many

stages (one or two) are required during this phonological-semantic mapping processing (Caramazza, 1997; Levelt, Roelofs, & Meyer, 1999). The one-stage view claims that a word-level phonological representation is directly projected to the meaning of a word, while the two-stage view suggest there is another level, called the lemma, intervening between the meaning representation and the word-level phonological representation. According to the dual-stream model, the lexical interface also sends information to the combinatorial network, a higher-order combinatorial network that plays a crucial role in understanding the meanings of phrases and sentences. However, again, detail is lacking about how this integration process works.

In terms of the anatomical location of this lexical interface, the dual-stream model suggests that the lexical interface is supported by the posterior middle temporal gyrus (pMTG) and the posterior inferior temporal gyrus (pITG). Such a hypothesis has gained support from a considerable number of brain imaging studies that investigated semantic processing (e.g. Saur et al., 2008; see the meta-analysis by Binder, Desai, Graves, & Conant, 2009). Supporting evidence also comes from patient data. Damage to the MTG and IFG is frequently associated with profound semantic deficits (Lambon-Ralph, Lowe, & Rogers, 2007; Noppeney et al., 2007). For the combinatorial network, it is suggested that it has reciprocal connections with the lexical interface and lies in the anterior middle and inferior portions of the temporal lobes, predominantly in the left hemisphere. Such a hypothesis is consistent with evidence from positron emission tomography (PET) and fMRI studies, showing that when participants listened to semantically coherent sentences compared with listening to word lists and other types of

simple auditory stimuli, greater activations were found in the left anterior temporal lobe (ATL) (Humphries, Willard, Buchsbaum, & Hickok, 2001; Humphries, Binder, Medler, & Liebenthal, 2006; Humphries, Love, Swinney, & Hickok, 2005; Mazoyer et al., 1993; Vandenberghe, Nobre, & Price, 2002).

In sum, the dual-stream model proposes that the ventral stream is mainly responsible for mapping auditory speech signals onto conceptual and semantic representations, which is essential for speech comprehension and recognition.

1.5.2 The dorsal "How" Stream, from Sound to Action

Unlike previous hypotheses in which dorsal stream is assumed to process the "where" information, the dual-stream model suggests that the dorsal stream plays an important role in auditory-motor integration. According to the dual-stream model, like the ventral stream, the dorsal stream has two main functional-anatomical components: the sensorimotor interface and the articulatory network (Figure 1-4).

In the dorsal stream, there is a particular region located in the posterior extent of the Sylvian fissure at the junction of the parietal and temporal lobes known as area Spt (Sylvian parieto-temporal), which is proposed by Hickok and Poeppel (2004, 2007) to be the location of a sensorimotor interface. Area Spt is proposed as an intermediary between phonological network in the temporal lobe bilaterally and the motor-based articulatory network. More specifically, area Spt acts an integration device that uses primarily auditory information from the phonological network to help generate/adjust speech motor movements. Speech requires

extensive sensory-motor integration, both during production and perception. As described previously, during speech production, speakers continuously monitor auditory feedback and use such feedback to adjust their next speech movement (Houde & Nagarajan, 2011; Tourville & Guenther, 2011). Several fMRI studies have shown that area Spt is involved in speech production even when the production is covert (Buchsbaum, Hickok, & Humphries, 2001; Buchsbaum, Olsen, Koch, & Berman, 2005; Hickok, 2009; Okada & Hickok, 2006). For instance, brain responses of five participants were examined when they listened to or covertly rehearsed, either meaningless sentences (‘‘jabberwocky’’) or melodic tonal sequences (Hickok, Buchsbaum, Humphries, & Muftuler, 2003). Results showed that activity in area Spt increased not only during listening conditions but also during covert rehearsal conditions. Importantly, articulatory rehearsal was conducted sub-vocally, so activation observed during the rehearsal condition was not due to hearing one's own voice. Interestingly, the study also showed that area Spt is not speech-specific, as it responded equally well to other sorts of sounds (tonal melodies) during both perception and production (covert rehearsing) (Pa & Hickok, 2008). However, a follow-up fMRI study suggested that area Spt is specifically involved in sensory-motor integration of vocal tract gestures specifically. Skilled pianists were scanned using fMRI while they listened to a melody, and then either covertly hummed the melody (vocal tract regulation), or covertly played the melody on a piano (manual regulation). Results showed that area Spt activated more during the covert humming condition compared with the covert playing condition.

The “sensorimotor interface” has not only reciprocal connections with the phonological network, but also reciprocal connections with the “articulatory

network", which comprises several motor-related regions, including Broca's region, premotor cortex, and probably the primary motor cortex for controlling the vocal tract. FMRI was used to test the neuroanatomical basis of the dorsal stream and identify the most probable brain areas for the "articulatory network" (Saur et al., 2008). Results showed that compared to rest, the repetition task during which participants were required to overtly repeat words and pseudowords immediately after presentation, resulted in strong bilateral temporo-frontal activation with peaks in primary auditory and motor areas. When repetition of pseudowords was contrasted with repetition of real words, activity was focused on the left superior temporal gyrus and shifted from primary motor to premotor and prefrontal areas in the left frontal lobe. The authors therefore suggested that this increased activation of prefrontal and premotor areas was due to the higher phonological processing demands necessary for preparing production of pseudowords compared with high-frequency real words (Saur et al., 2008).

In sum, a dorsal stream, which involves brain regions in the posterior frontal lobe and the posterior dorsal-most aspect of the temporal lobe and parietal operculum, is hypothesized to be responsible for translating acoustic speech signals into articulatory representations, which is particularly important for speech development and speech production. Interestingly, Hickok and Poeppel (2007) point out that unlike the ventral stream which is largely bilaterally organized, the dorsal stream is strongly left-hemisphere dominant.

1.6 The role of motor system in speech perception

1.6.1 Motor theory of speech perception

Fifty years ago, Liberman and colleagues first developed the motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985) to explain some unexpected findings from their experiments. In the original experiment, they tried to create a reading machine for the blind, that worked by substituting individual letters with specific sounds. However, the project failed because listeners were not able to perceive alphabetic sequences at practical rates required to detect words. This finding motivated them to ask the question how normal individuals can perceive speech (e.g. syllables and words) at such rapid rates in their daily life. They then conducted a series of experiments to investigate the acoustic structure of speech and how the acoustic signal corresponds to the phonetic segments of syllables and words (Liberman, Delattre, Cooper, & Gerstman, 1954; Liberman, Harris, Hoffman, & Griffith, 1957). They found that successive consonants and vowels sometimes overlap with one another ("coarticulation"). Thus, according to Liberman's words, speech is perceived as an efficient "code" of overlapping speech gestures instead of an acoustic "alphabet". They also found that the speech signal is highly context sensitive: a given speech sound is influenced by a preceding or following speech sound, which is also referred to as "coarticulation". For example, the English phoneme /n/ sounds differently to normal (alveolar consonant as in "ten") if it appears in the word "tenth", as the following sound, /θ/, is dental.

The motor theory of speech perception has gone through several changes during the past few decades. However, there are a few core claims that remain. First, is

the claim that the primary target of speech perception is the articulatory gesture rather than the acoustic signal. More specifically, it is hypothesized that listeners perceive spoken words by recognizing the vocal tract gestures rather than identifying the acoustical structures generated by speech. A second important claim of the motor theory is that “speech is special”. The perception of speech sounds relies on specialized decoders/neural mechanisms that are specific to speech and language, are unique to humans, and cannot be attributed to general perceptual/cognitive mechanisms (Liberman & Mattingly, 1985; Liberman & Whalen, 2000). The motor theory of speech perception and these claims gained support from some experimental studies (see Galantucci, Fowler, & Turvey, 2006 for a review). For example, listeners perceive speech in noise more accurately when they can see the speakers (Sumby & Pollack, 1954). These effects have been replicated when speech is not seen but is perceived haptically (Fowler & Dekle, 1991). On the other hand, the motor theory of speech perception also suffers from several critiques (see Lane, 1965). A main limitation of the gesture perception claim is that it would predict infants and animals who do not produce speech and therefore do not have access to gestures cannot perceive speech in the same way as adult humans who can produce speech. In actuality, however, it is the other way around (Tsao, Liu, & Kuhl, 2004). It would also predict that defects in speech production would impair speech perception, but again they do not (MacNeilage, Rootes, & Chase, 1967).

1.6.2 Neurobiology evidence of motor contribution

As discussed previously, the importance of auditory feedback from auditory system in speech production has been widely accepted. The controversy arises

concerning the role of the motor system in speech perception (D'Ausilio et al., 2009; Pulvermüller & Fadiga, 2010). The motor theory of speech perception proposes that speech perception involves extracting information about articulatory gestures from the acoustic signal, which might suggest a significant role of the speech motor cortex in the initial stages of speech perception (Devlin & Aydelott, 2009; Liberman et al., 1967; Liberman & Mattingly, 1985; Pulvermüller et al., 2006). In contrast, the dual-stream model does not predict a critical and necessary role of motor system in speech perception but suggests that information is mainly sent from perception to production and not in the reverse direction (see Hickok & Poeppel, 2000, 2004, 2007).

During the past two decades, a large number of studies have addressed this disagreement. Early data from neurophysiology, neuroimaging, and non-invasive brain stimulation have demonstrated the involvement of motor system during passive speech perception, providing strong or weak support for the motor theory of speech perception (e.g. Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Watkins, Strafella, & Paus, 2003; Wilson et al., 2004). For example, Fadiga and colleagues (2002) found increased excitability in the speech motor system, which they measured using TMS to elicit motor-evoked potentials (MEPs) recorded with electromyography (EMG) from listeners' tongue muscles when they passively listened to speech sounds that required the tone in their production e.g. "terra". Similar results were observed in lip muscles (Watkins et al., 2003): compared with control conditions (listening to non-verbal sounds or viewing eye and brow movements), both listening to and viewing speech significantly increased the size of the MEP recorded in lip

muscles elicited by stimulation of the lip representation in primary motor cortex. Supporting evidence also came from brain imaging studies (Pulvermüller et al., 2006; Wilson et al., 2004). For example, Pulvermüller et al., (2006) showed that the motor regions in the precentral gyrus were differentially activated in a somatotopic manner when participants listened to syllables produced by different articulators (lips and tongue).

However, all the evidence mentioned above came from correlational studies. A core prediction of the motor theory of speech perception is that disruption (temporary or permanent) of the speech motor cortex should impair speech perception. By applying repetitive TMS, which can create a virtual lesion in the region stimulated, researchers tested the causal role of speech motor cortex in speech perception in healthy human participants (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Mottonen & Watkins, 2009). For example, a previous study found that TMS-induced disruption of speech motor cortex significantly impaired participants' performances in speech task involving discriminating stop consonants embedded in noise (Meister et al., 2007). However, there are also alternative explanations of these results. An important concern is the non-focal effect. TMS might have a spreading effect, which might affect the ongoing activation of elsewhere. Second, motor areas are not necessary for speech perception, but are recruited only under certain conditions that make speech discrimination hard (Stokes, Venezia, & Hickok, 2019; Wu, Chen, Wu, & Li, 2014). For instance, in the two studies mentioned before, stimuli were embedded in noise (Meister et al., 2007) or the task was quite demanding, in which participants were required to detect subtle differences on a continuum of phonemes

(Möttönen & Watkins, 2009). Later studies tested the involvement of speech motor cortex in speech processing under different noise-level or task demand (D'Ausilio, Bufalari, Salmas, & Fadiga, 2012; Sato, Tremblay, & Gracco, 2009; Stokes et al., 2019). In the work of D'Ausilio et al., (2012), participants' performances were compared when they discriminated syllables under normal or noisy environment after they received disruptive rTMS over the speech motor cortex. Results indicated that the motor cortex may play a more significant role in speech perception only when the stimuli were degraded. The involvement of motor areas might also be modulated by task complexity. Compared with sham stimulation, rTMS applied over the left ventral PMC resulted in slower phoneme discrimination requiring phonemic segmentation, while no effect was observed in the same discrimination tasks without need for phonemic segmentation (Sato et al., 2009). Another concern regarding the causal role of speech motor cortex is disruptions of motor areas might have affected response bias, rather than speech perception (Hickok, 2010).

To address these concerns, a series of further experiments have been carried out. First, in terms of spatial focality related to TMS, a previous study showed that low-frequency rTMS over lip but not hand representations selectively affected speech discrimination (Möttönen & Watkins, 2009). Moreover, TMS-induced disruption of the lip representation only affected perception of sounds involving the lips in their articulation but not sounds that did not involve the lips, which suggests a relatively high precision of rTMS (Möttönen & Watkins, 2009). A later study replicated the findings of this previous study and also investigated the effects of TMS-induced disruption of the primary motor cortex on response

bias (Smalle, Rogers, & Möttönen, 2015). They found that disruption of the lip representation by rTMS significantly reduced the sensitivity for between-category pairs as before. In contrast, disruption of the hand representation temporarily reduced response bias but not sensitivity. These results provide strong support for the idea that the speech motor cortex contributes to speech perception rather than to response bias. Furthermore, combining rTMS and EEG, Möttönen and colleagues (2013) successfully demonstrated the automatic involvement of speech motor cortex in speech perception. Instead of using behavioural tasks, which might be under the constraints of task demands or require attention, they recorded mismatch negativity (MMN) responses to changes in syllables while participants passively listened to an oddball sound sequence and watched a silent film (i.e., attention was directed away from the speech sound). Results indicate that disruption of speech motor cortex by rTMS impaired the brain's ability to detect a change in a syllable indexed by automatic EEG responses. In a follow-up study, previous findings were replicated using MEG and natural speech sounds which were easily discriminated. In addition, they investigated the effect of attention on specificity and timing of interactions between the auditory and motor cortex during speech processing and found that when speech sounds were ignored, the motor contributions to auditory speech processing were late and not feature specific (>170 ms), whereas when speech sounds were attended, there was an additional early (<100 ms) motor effect on auditory speech processing, which was feature-specific (Möttönen, van de Ven, & Watkins, 2014).

1.7 Hemispheric lateralization of auditory processing

Since the work of Broca, much evidence exists indicating that speech and language are processed in the brain by mechanisms engaging cortical and subcortical areas within the left hemisphere (see Binder et al., 1997 for review), while brain regions in the right hemisphere have been implicated in processing of non-linguistic sounds (Zatorre, Evans, & Meyer, 1994). Using fMRI, researchers have repeatedly found stronger activations in the left temporal regions when participants processed speech information than non-speech sounds (see Zatorre and Binder, 2000 for review). In contrast, researchers found that patients with right but not left auditory cortex excisions, were specifically impaired at pitch processing, such as perception of missing fundamental pitch (Zatorre, 1988) and direction of pitch change (Johnsrude, Penhune, & Zatorre, 2000).

Despite much research, researchers are still far from a consensus on the mechanisms underlying such hemispheric lateralization. Concepts such as spectro-temporal resolution (Schönwiesner, Rübsamen, & Von Cramon, 2005; Zatorre & Belin, 2001) and time integration windows (Poeppel, 2003) have been proposed trying to explain the hemispheric specialization during auditory processing.

The temporal-spectral trade-off theory which links the hemispheric asymmetries in auditory processing to differential spectro-temporal resolution of auditory cortex in left and right hemispheres, was proposed, in which the left auditory areas are specialised for rapid temporal processing, while the right auditory areas

show preference for spectral processing (Schönwiesner et al., 2005; Zatorre & Belin, 2001; Zatorre, Belin, & Penhune, 2002; Zatorre, Evans, Meyer, & Gjedde, 1992). This hypothesis has received extensive experimental support. By using PET, Zatorre and Berlin. (2001) examined the responses of human auditory cortex to spectral and temporal information. Participants were presented with two kinds of sound sequence, in which either spectral information (tone spacing) or temporal information (speed of alternation of tones with a fixed spacing) was changed. Results indicated that auditory responses to the temporal variations were weighted towards the left, while responses to the spectral variations were weighted towards the right. Obleser, Eisner, & Kotz (2008) further examined differential contributions of left and right auditory areas in word comprehension under different sound degradation levels. They found brain activation in the left anterolateral superior temporal sulcus (STS) was mainly due to increases in the temporal detail of the signal, whereas the right auditory areas showed higher sensitivity to the variations in spectral detail. A more recent study reported similar results showing: (i) processing of speech sounds are most affected by degradation of information in the temporal dimension, while processing of melodic sounds are most affected by degradation in the spectral dimension; (ii) neural decoding of speech and melodic sounds depends on neural activity patterns in the left and right auditory regions, respectively (Albouy, Benjamin, Morillon, & Zatorre, 2020).

Based on the features of speech signal and involvement of both left and right auditory cortex during speech sound processing, Poeppel (2003) proposed the Asymmetric Sampling in Time hypothesis (AST), which aimed to explain

hemispheric asymmetries during speech processing by the neuronal entrainment to speech. As for the temporal-spectral trade-off theory of Zatorre and colleagues, AST hypothesized that the speech signal is processed bilaterally in the left and right primary auditory cortex. However, beyond the initial processing, the signal is elaborated asymmetrically in the non-primary auditory cortex which includes two different temporal integration windows resulting from the intrinsic properties of local neuronal ensembles (Assaneo et al., 2019). The auditory areas in the right hemisphere have a larger population of neurons with longer temporal integration windows (~150–250 ms), while the left auditory areas showed preference to extract information from short (~20–40 ms) temporal integration windows. It is further suggested that these temporal windows are reflected in neuronal oscillatory activity in different frequency bands (e.g. longer window: theta; shorter window: low gamma), that accord with basic units of speech, namely syllables (theta) and segmental information (gamma) (Ding et al., 2017; Giroud et al., 2020).

1.8 Processing tonal languages

1.8.1 Hemispheric lateralization of tone in tonal languages

A considerable number of studies have reported a rightward lateralization of the processing of pitch-related sound, which is consistent with both the temporal-spectral trade-off theory (spectral information is selectively processed in the right hemisphere) and AST (the right hemisphere preferentially extracts slowly-modulated acoustic information). However, most of the studies providing experimental support for these hypotheses or models used non-speech signals

or stimuli derived from European languages (Boemio, Fromm, Braun, & Poeppel, 2005; Obleser et al., 2008; Telkemeyer et al., 2009; Zatorre & Belin, 2001).

Phonology includes segmental and suprasegmental information. In spoken language, segmental information typically refers to consonants and vowels, which are used for distinguishing words, while suprasegmental information is often regarded as the "musical" aspects of speech, which serves to express a speaker's emotions or intentions (e.g. intonation) through stress or prosody. In English, or most other European languages, pitch changes are suprasegmental information and provide little lexical information. In contrast to European languages, pitch changes or tones in tonal languages are lexically distinctive and are an essential feature of the meaning of a word. Many Asian languages are tonal languages, such as Chinese, Vietnamese, Thai, which are estimated to consist of 60–70% of the world's languages (Yip, 2002). Studying tonal language offers the possibility to investigate the neural mechanisms underlying pitch processing at different levels because of their special phonemic status (see Wong, 2002; Gandour and Krishnan, 2015 for reviews).

Here, when referring to tonal languages, we mean languages that use pitch to change the meaning of a word, which are different from pitch accented languages (e.g. Japanese, Hebrew), in which the meaning of a word can be changed if a different syllable is stressed (Yip, 2002). Mandarin Chinese, which has four distinct tones, is perhaps the best-known example of a tonal language. For example, the word "ma" can mean entirely different things depending on the tone used, ma1 (high-level tone) "mother", ma2 (high-rising) "hemp", ma3 (low-falling-

rising) “horse”, and ma4 (high-falling) “scold” (Xu, 1997). Thai (Bangkok) has five different tones: tones 1 to 5 are mid-level (T1), low-falling (T2), high-falling (T3), high-rising (T4), and low-rising (T5) (Tingsabadh & Abramson, 1999).

As mentioned above, it is well-established that pitch-related sounds are preferentially processed in the right hemisphere in non-tonal language speakers (Zatorre & Belin, 2001; Zatorre et al., 2002, 1992). Researchers have explored whether tonal language speakers would exhibit the same rightward lateralization as non-tonal language speakers when processing pitch sounds. On the behavioural level, early research with patients showed impaired lexical tone production and perception of tonal language speakers following damage to the left hemisphere (Eng, Obler, Harris, & Abramson, 1996; Gandour, Ponglorpisit, & Dardarananda, 1992; Yiu & Fok, 1995). By using the dichotic listening paradigm, several experiments showed that tonal language speakers, including native speakers of Mandarin (Yue Wang, Behne, Jongman, & Sereno, 2004; Yue Wang, Jongman, & Sereno, 2001) and Thai (Van Lancker & Fromkin, 1973), could more accurately report lexical tones presented to their right ear (a left-hemisphere advantage). On the neural level, using PET, Gandour and his colleagues (1998) compared the brain responses of native speakers of English and Thai when they listened to the same Thai tones. Results showed that when discriminating pitch patterns in Thai words, only Thai speakers showed activation in the left frontal operculum, which suggested that pitch information when in a linguistic context is preferentially processed in the language-related regions of the left hemisphere. Similar results were reported when comparing Mandarin speakers and English speakers (Klein, Zatorre, Milner, & Zhao, 2001). Cerebral

blood flow was found to increase in some common regions, including the superior temporal gyrus, lateral cerebellum and the thalamus, in both groups. However, only native Mandarin speakers showed additional activity in frontal, parietal, and parieto- occipital regions of the left hemisphere. In contrast, activations in some right-hemisphere regions were observed exclusively in the English group.

The differences in auditory processing of tone changes in speech between the two language groups provide strong support for the influence of a listener's language experience on hemispheric lateralization of speech. Such lateralization of auditory speech processing is determined early during language acquisition: internationally adopted children who had been completely cut off from exposure to Mandarin since adoption (at 12.8 months of age, on average) exhibited left lateralization in the auditory system during lexical tone processing when studied over 10 years later as teenagers (Pierce, Klein, Chen, Delcenserie, & Genesee, 2014).

1.8.2 Tonal versus segmental units

Another interesting question related to tonal languages is whether processing of lexical tone in the brain is similar to processing of segmental components (e.g. vowels or consonants). On the behavioural level, evidence from a variety of tasks suggests that compared with segmental information, lexical tone information is relatively harder to extract or later extracted (Taft and Chen, 1992; Tong, Francis, & Gandour, 2008; Wiener & Turnbull, 2016; Ye & Connine, 1999). For instance, Cutler and Chen (1997) found that when making same-different judgements, native Cantonese speakers were slower and less accurate if the difference

between two syllables was in tone. Similarly, slower or less accurate responses were also found in Cantonese speakers when making real word/non-word judgments if the only difference between the nonword and the word was in tone. In addition, researchers examined the interaction between segmental and lexical tone processing in Mandarin speakers (Tong et al., 2008). Participants were instructed to attend to one dimension while ignoring the other. Results indicated that segmental dimensions (vowels or consonants) exerted larger interference on lexical tone dimensions than vice versa. Such asymmetric interference effects observed between segmental and lexical tone dimensions suggest that vowel/consonant and lexical tone processing might be supported by different neural mechanisms. On the neural level, depending on task demands, some inconsistent results were found regarding the time course of processing of segments and lexical tone. For example, Schirmer, Tang, Penney, Gunter, & Chen (2005) found that in Cantonese speakers, scalp-recorded event-related potentials (N400, a negative component that is traditionally considered to be associated with semantic processing) are similarly modulated by semantic violations at the tone level or at the segmental level, that occurred in the end of a sentence. This suggests that listeners probably access tone and segmental information at a similar point in time when extracting semantic meaning during word processing in Cantonese. However, different results were found in Mandarin speakers using a different task (Zhao, Guo, Zhou, & Shu, 2011). During the experiment, participants were asked to judge whether two pictures (the name in Chinese character/orthography for the items was also shown in the first picture) belonged to the same semantic category while spoken words were also presented between the two pictures. Phonological competition was manipulated

by presenting the spoken words that either matched or mismatched with the first picture either in tone or segment. Results showed that segmental violations (i.e. the spoken word mismatched the item in the first picture in segment) elicited an earlier and stronger N400 than tone.

1.8.3 The asymmetry of motor contributions during auditory speech processing

As discussed above, the mechanisms underlying hemispheric asymmetries during processing of auditory speech signal are extensively studied, vigorously debated, and still not understood in a satisfactory manner. Several sophisticated proposals have been developed, but most of them can be categorized into one of two general theoretical perspectives. The acoustic hypothesis (e.g. the temporal-spectral trade-off theory) suggests the brain asymmetry of speech processing is determined by differential spectral and temporal sensitivities of the left and right auditory cortex to low-level acoustic signals (Boemio et al., 2005; Zatorre & Belin, 2001). Alternatively, the functional hypothesis claims that the asymmetry is mainly cued by more abstract, linguistically relevant properties of a stimulus, such as whether it changes the meaning of a word (Liberman & Whalen, 2000). Both of these hypotheses have received extensive experimental support, but, at the same time suffered from several shortcomings. One important limitation of these hypotheses is that they assume brain asymmetry of speech processing can be explained by a single factor or brain area. Given the complexity of speech processing, it is more likely that the asymmetry is determined by multiple factors, such as the modulatory role of top-down

processing mechanisms on bottom-up processing of acoustic properties (Zatorre & Gandour, 2008).

In study 3 (Chapter 5), I hypothesized that the asymmetry of auditory speech processing is under the influence of the motor system. I investigated the asymmetry of motor contributions to auditory speech processing in tonal (e.g. Mandarin) and non-tonal language speakers capitalizing on the fact that the same low-level acoustic signal (e.g. tone) serves different linguistic functions in tonal and non-tonal language groups.

1.9 Aims and Scope of the thesis

This thesis focuses on "auditory-motor interactions during speech perception and production". One behavioural study, comprising two experiments, used a novel formant perturbation paradigm to explore how participants modified F1 and F2 in response to a perturbation in only one of these. This paradigm was next used in combination with TMS over the motor cortex to evaluate the contribution of speech motor cortex to this form of sensorimotor adaptation. Finally, TMS was used in combination with EEG in a series of experiments with tonal and non-tonal language speakers to evaluate the contribution of the left and right speech motor cortex to asymmetry in speech processing.

1.9.1 Study 1: Speech adaption: behavioural testing

In study 1 (Chapter 3), by using a real-time perturbation paradigm, I examined the compensatory responses of speakers when they experienced perturbations to their auditory feedback. Forty native English speakers were recruited. Half of

them heard feedback of their speech production with a 110-mel increase in the first formant (F1) while the other half experienced a 110-mel decrease in the second formant (F2). Changes in F1 and F2 production were measured in response to this altered feedback. By manipulating the formant shift, I aimed to test the independence of speech control of individual formants.

1.9.2 Study 2: Speech adaption: TMS interference

I next conducted a brain stimulation study exploring the neural basis of speech adaptation using the real-time perturbation paradigm I developed in study 1 coupled with TMS, which temporarily inhibits function in specific brain areas (Chapter 4). Although multiple lines of evidence implicate the role of M1 in motor learning and memory, whether M1 causally contributes to sensorimotor adaptation in the speech domain remains unclear. Here, I aimed to determine whether temporary disruption of the articulatory representation in left M1 by rTMS impairs sensorimotor adaptation in speech.

1.9.3 Study 3: Asymmetry of auditory-motor speech processing: EEG and TMS

In study 3 (chapter 5), I used the same form of low-frequency rTMS used in study 2 but this time coupled with EEG to explore the asymmetry of motor contributions to auditory speech processing in different language groups. More specifically, I temporarily disrupted the left or right speech motor cortex using rTMS in groups of tonal and non-tonal language speakers and measured the impact of these disruptions on automatic auditory processing of phonemic and tonal changes in sequences of syllables using EEG.

Chapter 2 Methodology

2.1 Summary

Three main techniques were used in this thesis: real-time feedback perturbation, transcranial magnetic stimulation (TMS), and electroencephalography (EEG). The details of real-time feedback perturbation and related topics that can be tackled by this method were discussed in chapter 1. In this chapter, I give a brief introduction to TMS methods used in this thesis, namely single pulse TMS and low-frequency repetitive TMS. These methods were used in two small proof-of-principle experiments to confirm: (i) that the TMS protocol we used to stimulate the tongue representation in Study 2 (chapter 4) produced the effects previously described for lip and hand stimulation; and (ii) that the TMS protocol was effective in speakers of tonal languages who were ethnically Han Chinese and typically have different skull shapes and thicknesses to Caucasian participants who made up the non-tonal language groups in Study 3 (chapter 5). Finally, I cover EEG, specifically the mismatch negativity (MMN) response, which was used to measure automatic speech discrimination in Study 3 (chapter 5).

2.2 Transcranial magnetic stimulation

TMS was introduced in 1985 by Barker and his colleagues who demonstrated that a single TMS pulse applied over the primary motor cortex (M1) elicits responses in those muscles that receive corticomotor input from the stimulated motor cortical area (Barker, Jalinous, & Freeston, 1985). Such transcranially

evoked motor responses (MEPs: motor-evoked potentials) can be recorded by attaching electrodes on the target muscles. Since then, TMS has become a widely used non-invasive brain stimulation technique, which allows researchers and clinicians to stimulate different brain areas by applying magnetic pulses over the scalp (see Terao and Ugawa, 2002; Groppa et al., 2012 for reviews).

A stimulation device consists of a transducing coil, which comprises windings of copper wires and is connected to a discharge system. When a current is discharged and passed through the coil, it creates a small and precise magnetic field around the coil. Conductors within this rapidly changing magnetic field can be induced to carry a charge and the nearest conductors will be the neurons in the cortex if the stimulation coil is placed tangentially on the head. The power of the magnetic field decreases rapidly with distance from the coil. Therefore, the effects of TMS are focal and maximal at the stimulated area (1.5–2 cm deep from the scalp: see Deng, Lisanby, & Peterchev, 2013).

2.2.1 Monophasic versus biphasic stimulation

TMS can be delivered using monophasic or biphasic pulses (see figure 2-1; see Groppa et al., 2012 for a review). The monophasic pulse starts with a strong initial current, which is then dampened by a return current, resulting in a monodirectional (i.e. of single polarity) current flow. Only the initial rapidly changing current flow produces current flow in the stimulated brain, while the dampened return current produces no neuronal stimulation in the brain. The biphasic pulse configuration resembles a cosine waveform, in which an initial current rise is followed by a reversed current (i.e. below zero) and by a

subsequent increase of current (i.e. rise back–zero). In contrast to the monophasic pulse, each phase of the biphasic pulse induces physiologically significant tissue current, which flows in the same or opposite direction as the initial current rise of the biphasic pulse. While all phases of the biphasic pulse produce effective neuronal stimulation in the brain, the second (reversal) phase has the largest amplitude and the longest duration, which causes a stimulation of different populations of neurons as opposed to the initial rising phase.

Monophasic pulses are conventionally used for single pulse TMS, while biphasic pulses are most commonly used for repetitive TMS (rTMS) protocols because of the ability to achieve higher frequencies of stimulation. However, previous studies showed that monophasic TMS has a relatively stronger effect on cortical excitability inhibition or facilitation during repetitive stimulation compared to biphasic TMS (Arai et al., 2005; Sommer, Lang, Tergau, & Paulus, 2002), perhaps because monophasic pulses preferentially activate one population of neurons oriented in the same direction so that their effects readily summate. Biphasic pulses in contrast may activate several different populations of neurons (both facilitatory and inhibitory) so that summation of the effects is less clear than with monophasic pulses.

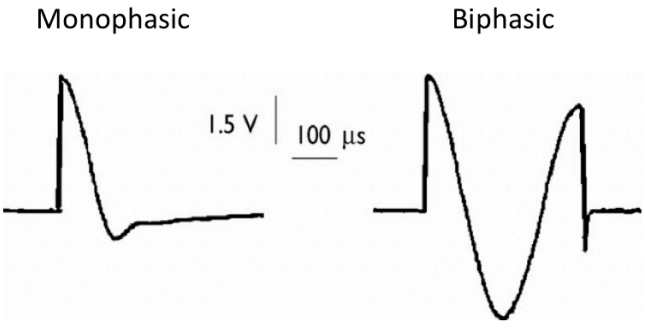


Figure 2-1 Time-course of monophasic (left) and biphasic (right) pulses (modified from Sommer et al., 2002). **Reprinted with permission from Wolters Kluwer Health, Inc. copyright (2002).**

2.2.2 Single-pulse TMS over the primary motor cortex

When a single pulse of TMS is applied over human M1 at sufficient threshold intensities, a number of descending corticospinal volleys (waves) can be recorded at the level of the cervical spinal cord. Such descending volleys have previously been recorded in the pyramidal tract after a single electrical stimulus applied directly to the surface of the motor cortex in animal studies (Patton and Amassian, 1954). The first of these volleys, termed the direct or D wave, is shown to be the result of direct excitation of the corticospinal neuron at its axon hillock. With increasing stimulus intensity, later waves, termed indirect or I waves, follow the D-wave at intervals of 1.2–2.0 ms, which are considered to arise from indirect trans-synaptic corticospinal excitation via different sets of intracortical neurons that synapse on the pyramidal neurons (Figure 2-2 A).

Each descending corticospinal volley then triggers glutamate release in cortico-motoneuronal synapses, which can be summed up temporally and spatially, leading to a depolarization at the postsynaptic cell membrane (Figure 2-2 B). If the volleys are strong enough to exceed the firing threshold, an action potential in the spinal motoneurons can be induced, which travels along the peripheral motor axons, triggering a motor response that can be recorded as the MEP (Figure 2-2 B&C).

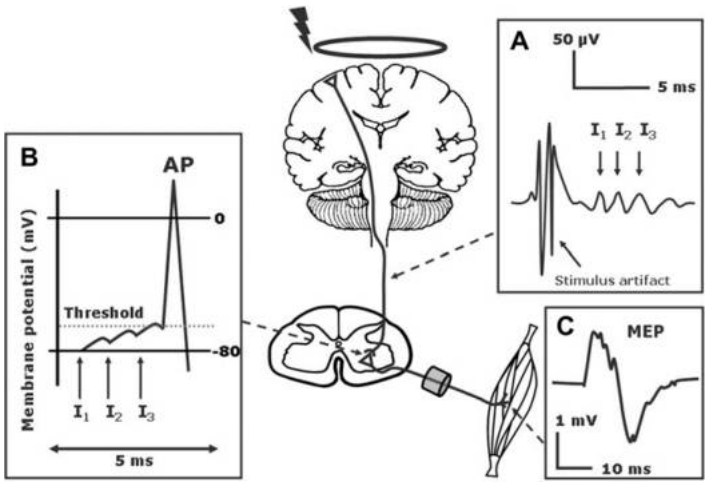


Figure 2-2 Neurophysiological basis of the motor evoked potential (Groppa et al., 2012). Reprinted with permission from Elsevier. copyright (2012).

(A) TMS-induced activation of the corticospinal neurons with a predominant contribution by late Indirect waves (I waves); (B) Temporo-spatial summation at the cortico-motoneuronal synapses (C) Motor evoked potential (MEP).

2.2.2.1 The size and latency of TMS-induced MEPs.

As showed above, a single TMS pulse applied at suprathreshold intensities would produce a series of descending volleys (multiple I-waves and possibly a D-wave) and evoke MEPs in the target muscles (Figure 2.2). The size and latency of MEP response is determined by the strength of the descending cortical input and the local excitability of the spinal motoneurons (Groppa et al., 2012). The descending excitatory drive depends on external (e.g., the intensity and direction of the induced electrical field) and internal (e.g., the maturation and integrity of corticospinal connectivity) factors. Manipulations of those factors can modulate the size and latency of TMS-induced MEPs.

MEP size can be measured from the positive through to the negative peak, referred to as the peak-to-peak amplitude (Figure 2-3). The size of a MEP can

also be measured as the area under the rectified MEP curve. Area under the curve is determined not only by the peak-to-peak amplitude but also by the duration of MEP and captures multiphasic responses. Here, I focus on the peak-to-peak amplitude, the measurement used in this thesis, which can be differently modulated by a number of factors. First, the peak-to-peak amplitude of an MEP depends on the location of the magnetic stimulus and on the direction of the induced electrical field, which is determined by the coil position and direction. To achieve more focal stimulation, figure-8 coils were designed and used (as shown in Figure 2-3; Rösler, Hess, Heckmann, & Ludin, 1989). In this type of coil, the currents in the two wings are in opposite directions, producing a maximum filed strength at the junctions of the two wings (stimulation is maximal underlying this point). Moving the coil over the scalp will influence the size of the MEP. Therefore, it is important to adjust the coil position to find the “hot spot” at which maximal responses can be observed in the contralateral target muscle. The direction of the induced current, which is perpendicular to the long axis of the figure-8 coil can also significantly influence the size of MEP. For the hand and articulatory muscles, inducing a current perpendicular to the central sulcus has been found to be optimal (Brasil-Neto et al., 1992; Dubach, Guggisberg, Rösler, Hess, & Mathis, 2004; Mills, Boniface, & Schubert, 1992). Therefore, it is suggested to orient the handle of the coil 45 degrees from the midline when trying to localize an appropriate "hot spot” for hand and articulatory representations in M1 (Adank, Kennedy-Higgins, Maegherman, Hannah, & Nuttall, 2018; Möttönen, Rogers, & Watkins, 2014) (see Figure 2-4).

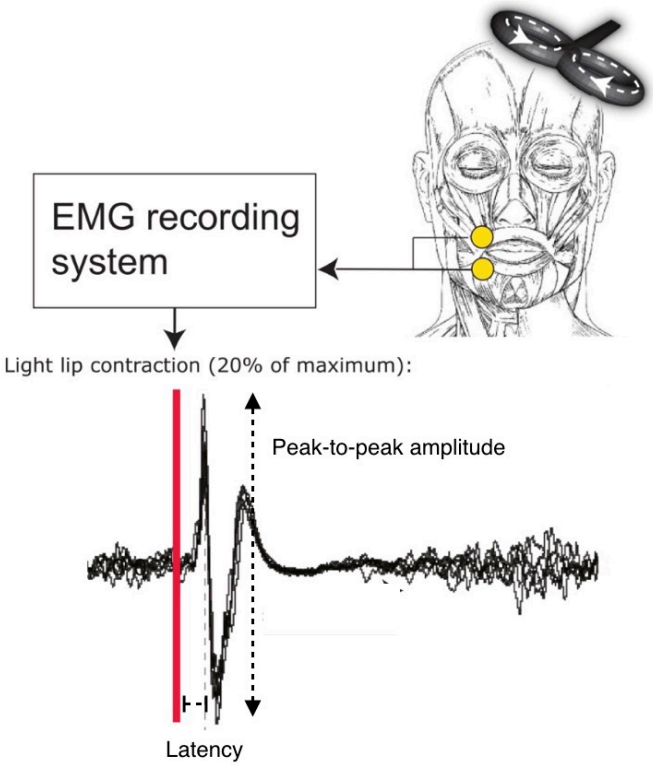


Figure 2-3 Using TMS to elicit MEPs in the lip muscle (adapted from Möttönen et al., 2014).

A figure-8 coil is placed over the representation of the lip motor cortex. The MEPs were recorded from one participant while she contracted the lips about 20% of the maximum.

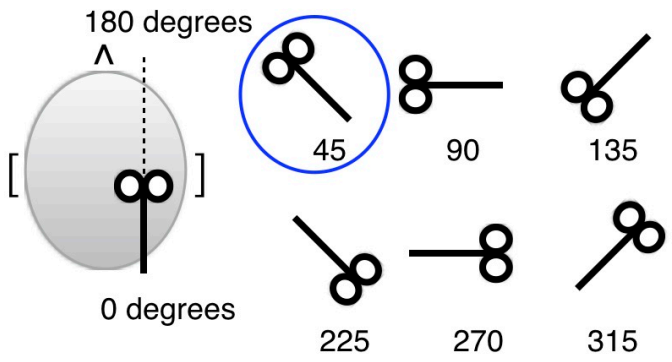


Figure 2-4 TMS coil orientation adapted from (Brasil-Neto et al., 1992).

Eight coil orientations for localizing "hot spot" for hand or articulatory representations in M1. Orienting the handle of the coil 45 degrees from the midline has been found to be optimal (highlighted by blue circle).

The size of the MEP is also influenced by the state of the motor system. In particular, it can be facilitated by voluntary contraction of the target muscle (Di Lazzaro et al., 1998). Pulses of a constant intensity elicit larger MEPs when the target muscle is contracted compared to when the muscle is relaxed. Facilitation of MEP size can also be observed by motor imagery (Kasai, Kawai, Kawanishi, & Yahagi, 1997), action observation (Kiers, Fernando, & Tomkins, 1997), and a variety of other cognitive manipulations (Pascual-Leone, Grafman, & Hallett, 1994). For example, related to the work carried out in this thesis, stimulation of the lip-representation elicits larger MEPs when listening to speech and viewing speech-related lip movements compared to when listening to non-verbal sounds or viewing eye movements (Watkins et al., 2003).

The latency of MEP, expressed in milliseconds, refers to the time taken by descending TMS pulses to reach the target muscle (Figure 2-3). MEP latencies, therefore, vary depending on the distance between the stimulation site and the

peripheral target muscle, as well as properties of the white matter fibres, such as axonal diameter, and myelination. MEPs recorded from the hand muscles have a latency (i.e. their maximal peak) approximately 24 msec after the pulse, whereas MEPs recorded from the lip muscles have a latency of approximately 15 msec after the pulse (see Figure 2.3 for the lip MEP recorded from a single representative participant) (Adank et al., 2018; Möttönen et al., 2014).

In sum, MEP measurements can be used to localize the cortical representations of different muscles in a participant. In addition, the magnitude of the MEP elicited by a fixed threshold or output of the stimulator provides a direct measure of the excitability of motor system and changes in this excitability in different conditions or states

2.2.2.2 Defining the threshold for corticomotor excitation: active motor threshold and resting motor threshold

As we mentioned above, the size of an MEP is determined by the local excitability of the spinal motoneurons and the strength of the descending cortical input (e.g. the intensity of stimulator output). Therefore, it is important to define a fixed stimulator output or threshold for stimulating each participant when using the size of MEP to evaluate the excitability of motor system at different time points or under different experimental conditions.

The first general procedure to define the motor threshold (MT) was introduced by a committee of the International Federation of Clinical Neurophysiology (IFCN) in 1994 (Rossini et al., 1994). According to Rossini et al. (1994), the MT is defined

as the lowest stimulus intensity that is required to induce an MEP in at least 5 out of 10 trials (Rossini et al., 1994). The MT can be estimated either at rest (rMT: resting motor threshold) or under active conditions (aMT: active motor threshold). In case of a relaxed target muscle, a reliable MEP response has been defined as an MEP with ⩾50 μV peak-to-peak amplitude because it is difficult to differentiate MEPs with a peak-to-peak amplitude of less than 50 μV from noise (Rossini et al., 1994). The aMT is estimated when the target muscles are contracted. Under active conditions, a cut-off value is usually set to 200 μV peak-to-peak in order to avoid classifying voluntary electromyographic activity as an MEP (note also that a silent period frequently follows an MEP, which can help to identify them from the background contraction). As mentioned above, MEPs can be facilitated by voluntary contraction of the target muscle (Di Lazzaro et al., 1998). Therefore, it is important to train the participant to maintain a constant level of contraction, if MEPs are recorded from contracted muscles.

2.2.2 Repetitive transcranial magnetic stimulation (rTMS)

Repetitive TMS (rTMS) refers to a train of TMS pulses delivered at the same intensity over a brain area. The frequency of rTMS trains typically ranges between 1 Hz (one pulse per second) and 50 Hz (50 pulses per second). When applied to the motor cortex or other cortical regions of the brain, rTMS may induce effects that outlast the stimulation duration. Such TMS-induced perturbation effects have been studied to gain insight into the causal relations between brain activity and behaviour.

The direction of the effect induced by rTMS, either facilitation or suppression, is largely dependent on the frequency of stimulation (inter-pulse interval). For example, higher rTMS frequencies of 5–20 Hz tend to increase cortical excitability (Maeda, Keenan, Tormos, Topka, & Pascual-Leone, 2000; Pascual-leone, Valls-solé, Wassermann, & Hallett, 1994), while low-frequency (1 Hz) stimulation leads to a suppression of cortical excitability (R. Chen et al., 1997). Here, I focus on low-frequency rTMS, which was the protocol used in Study 2 and Study 3.

Low-frequency rTMS (1 Hz or below) is a commonly used protocol to investigate structure-function relationships even though the phenomena induced by such low-frequency rTMS in the cortex are complex and their physiological mechanisms are not well understood. Chen et al., (1997) showed that stimulation at 0.9 Hz over the hand representation in M1 for 15 minutes (810 pulses) led to a mean reduction in MEP amplitude of 19.5%. Such a decrease in cortical excitability lasted for at least 15 minutes after the end of the 0.9 Hz stimulation. Similar observations were reported by several other groups (e.g. Muellbacher, Ziemann, Boroojerdi, & Hallett, 2000; Siebner et al., 1999). Muellbacher et al., (2000) further showed that the suppressive effect was restricted to the hand motor representation, which was the target of the rTMS, and absent in the adjacent representation of the biceps. This indicates that such rTMS protocols allow relatively high precision in terms of spatial focality.

However, the aforementioned studies used trains of rTMS at suprathreshold intensity (e.g. 115% of the rMT in Chen et al., 1997) so that twitches in

contralateral muscles were evoked during the train. Thus, it is difficult to exclude that somatosensory feedback or some refractory period locally at the neuromuscular junction, rather than the direct rTMS cortical effects, may account for the findings. To address this, a later study applied subthreshold (90% of rMT) 1-Hz rTMS (10 min) over the hand representation and assessed the effect of such subthreshold rTMS on cortical excitability. They found that 1-Hz rTMS applied at sub-threshold intensity (which did not evoke motor responses during the train) also led to a lasting suppression of MEP amplitudes, which persisted for about a further 10 min (Romero, Anschel, Sparing, Gangitano, & Pascual-Leone, 2002).

Subthreshold low-frequency rTMS has been used to temporarily disrupt the articulatory representations in M1 (Möttönen et al., 2014; Mottonen & Watkins, 2009; Smalle, Rogers, & Möttönen, 2015). In these experiments, monophasic pulse sequences were applied at frequency of 0.6 Hz, due to hardware limitations. The highest repetition rate of monophasic pulses provided by a Magstim 200 stimulator is 3s (i.e. one pulse every 3s). By alternating pulses from two Magstim 200 stimulators through a 70-mm figure-8 coil connected via a BiStim module (Magstim, Dyfed, UK), an effective rTMS train can be delivered at 0.6 Hz (i.e. one pulse every 1.5s).

2.2.3 Effect of low-frequency rTMS on cortical excitability of tongue representation

In study 2, we aimed to explore whether temporary disruption of the articulatory

representation in left M1 by low-frequency rTMS impairs speech adaptation. As mentioned above, the effectiveness of low-frequency rTMS in reducing cortical excitability of hand and lip representations was demonstrated previously (e.g. Chen et al., 1997; Mottonen & Watkins, 2009). The MEP size recorded from hand or lip muscles elicited by supra-threshold single-pulse TMS over the primary motor representation of the target is significantly reduced after a 15-min train of low-frequency subthreshold rTMS (Mottonen & Watkins, 2009). Such reduction was maintained for at least another 15 min after the end of rTMS (Möttönen and Watkins 2009). We wanted to confirm that the same low-frequency subthreshold rTMS protocol could also suppress the cortical excitability of the tongue representation, which was the target area in our experiment (Study 2) evaluating the effects of such disruption on the perturbation task developed in Study 1. It was not possible to demonstrate this in the participants undergoing the speech adaptation in each experiment as electrodes attached to the tongue would affect speech production. Therefore, to confirm the effect of the same form of rTMS over tongue representation on motor excitability, we recruited a small group of participants to participate in a small proof-of-principle study.

2.2.3.1 Methods

Five healthy participants gave informed consent approved by The Central University (of Oxford) Research Ethics Committee and then were stimulated (26.82 ± 7.72 years, 2 females). All TMS pulses were monophasic, generated by two Magstim 200s and delivered through a 70-mm figure-8 coil connected through a BiStim module (Magstim, Dyfed, UK). Electromyographic (EMG) activity was recorded using a CED 1902 amplifier, a CED 1401 analog-to-digital

converter, and a computer running Spike2 (Cambridge Electronic Design). To record EMG activity from tongue muscles, two disposable electrodes were mounted on a noseclip (used by swimmers, Cressi, Italy) and attached to the dorsal and ventral surfaces of the right side of the tongue (Panouillères & Möttönen, 2018). The ground electrode was attached to the forehead.

Using single pulses of TMS, we first determined the active motor threshold (aMT: the minimum intensity at which TMS elicited at least 5 out of 10 MEPs with an amplitude of at least 200 μV when the target muscle was contracted at 20–30% of the maximum) and the threshold for determining MEPs (the lowest percentage of stimulator output that elicited MEPs in the tongue muscle of peak-to-peak amplitude of at least 1 mV on 10 consecutive trials during muscle contraction) for each participant. The mean stimulator intensity (percentage of maximum output) for the aMT, which was used for the repetitive TMS, was 50.0 % (SE ± 2.2 %). The mean MEP threshold was 57.4 % (SE ± 2.3 %). A low-frequency rTMS train (0.6 Hz, 12 min) was delivered at the aMT over the tongue of M1 cortex while the tongue was relaxed (i.e. stimulation was subthreshold). EMG activity was measured throughout to ensure that no MEPs were elicited. The success of disruption induced by rTMS of the tongue representation was evaluated by comparing the size of MEPs elicited by single pulses of TMS over the tongue representation before rTMS (pre-rTMS), immediately after (post-rTMS 1) and 10 mins after the end of the rTMS train (post-rTMS 2). At each time point, 20 TMS pulses were applied at the MEP threshold. Peak amplitudes were measured automatically within a time window of 10-32 ms in Matlab. The automatic windowing was checked manually by the experimenter. The mean MEP amplitudes for each recording session were then compared using one-tailed

paired-t tests as the direction of the MEP change was predicted to be a decrease (reduced excitability).

2.2.3.2 Results

Figure 2-5 presents the MEP amplitude changes before (pre-rTMS) and after (post-rTMS 1 and post-rTMS 2) disruptive rTMS, recorded from tongue. The mean peak-to-peak amplitude of tongue MEPs obtained pre-rTMS was 1.75 mV (±SE = 0.38). The mean amplitude was suppressed to 1.31 mV (±SE = 0.24) on average after 12 min rTMS over tongue representation (23% reduction in post-rTMS 1; $t_4 = 2.66$, $p = .026$, $d = 0.60$) and to 1.41 mV (±SE = 0.26) on average 10 mins after the end of the low-frequency rTMS train (17% reduction in post-rTMS 2; t4 = 2.17, p = .048, d = 0.46).

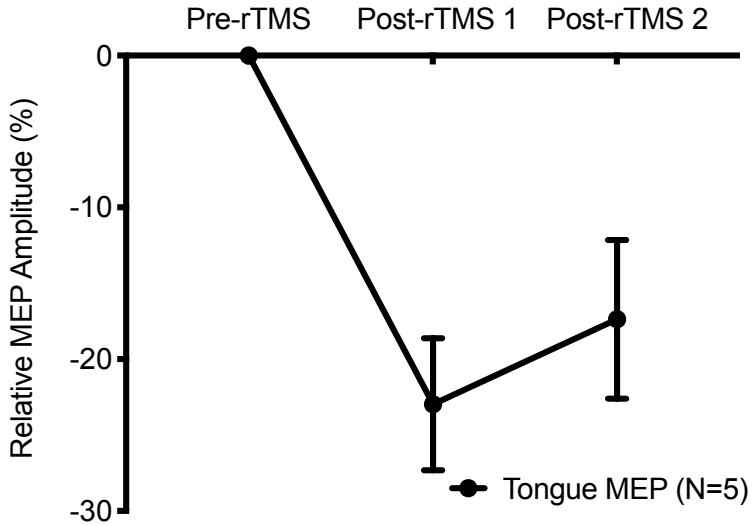


Figure 2-5 Effect of rTMS on cortical excitability of tongue motor cortex.

The figure shows mean reduction in peak-to-peak amplitudes (± SE) of post-rTMS tongue MEPs (post-rTMS 1 and post-rTMS 2) relative to pre-rTMS tongue MEPs.

2.2.3.3 Conclusion

Our results confirmed that low-frequency rTMS over the tongue representation in M1 was effective in reducing excitability as seen previously for the lip and the hand representations. This TMS-induced disruptive effect due to 12 min low-frequency subthreshold rTMS was maintained for at least 10 mins after the end of the rTMS train; this was sufficient to be effective for the duration of the whole learning phase of the speech adaptation experiment in Study 2.

2.2.4 Effect of low-frequency rTMS on cortical excitability in Mandarin speakers

In study 3, we aimed to use the same form of rTMS to temporarily disrupt the lip representation in both tonal (Mandarin) and non-tonal language speakers. Our own previous studies using this protocol involved groups of English-speaking participants comprising mainly individuals who were ethnically White and European. Anecdotal evidence in our own lab and previous work suggested that motor thresholds were higher in Chinese participants possibly due to differences in skull shape and thickness (Yi et al., 2014). We wished to evaluate whether our low frequency sub-threshold rTMS protocol would successfully reduce motor excitability in Chinese participants as these would be the Mandarin speakers in Study 3. Due to the limited duration of the TMS-induced disruption (see above), we were unable to obtain MEPs and record sufficient EEG data in the same participants. Therefore, we recruited a small group of additional participants who were speakers of Mandarin and ethnic Chinese to participate in a second small proof-of-principle study.

2.2.4.1 Methods

We stimulated the lip motor cortex in five Chinese participants (26.60 ± 1.67 years, 3 females). Participants gave informed consent approved by The Central University (of Oxford) Research Ethics Committee. All TMS pulses were monophasic, generated by two Magstim 200s and delivered through a 70-mm figure-8 coil connected through a BiStim module (Magstim, Dyfed, UK). Disposable electrodes were attached on the right (3) or left (2) orbicularis oris muscle (the upper and lower lip contralateral to the stimulation site) and forehead (ground) to record electromyography (EMG) signals using a CED 1902 amplifier, a CED 1401 analog-to-digital converter, and a PC running Spike2 (Cambridge Electronic Design).

After localization, single pulses of TMS were applied over the lip representation to determine the active motor threshold (aMT: the minimum intensity at which TMS elicited at least 5 out of 10 MEPs with an amplitude of at least 200 μV when the lip muscle was contracted at 20–30% of the maximum) and the threshold for MEP recordings (the lowest percentage of stimulator output that elicited MEPs in the lip muscle of peak-to-peak amplitude of at least 1 mV on 10 consecutive trials during muscle contraction) for each participant. The mean stimulator intensity (±SE) used for eliciting MEPs was 67.8% (±1.4%); for the repetitive TMS (aMT) it was 60.2% (±1.0%). A low-frequency rTMS train (0.6 Hz, 15 min) was delivered at the aMT over the lip representation of M1 cortex while the lip was relaxed (i.e. subthreshold). EMG activity was measured throughout to ensure that no MEPs were elicited. The success of disruption induced by rTMS of the lip representation

in Chinese participants was evaluated by comparing the size of MEPs elicited by single pulses of TMS over the lip: 1) before rTMS, 2) immediately after 15-mins rTMS, and 3) 15 mins after the end of the rTMS train. At each time point, 20 TMS pulses were applied at the MEP threshold. Peak-to-peak amplitudes were calculated automatically in a window of 11 - 32 ms after the TMS pulse using Matlab. The automatic windowing was checked manually by the experimenter. The 20 MEPs were then averaged at each time point and compared using one-tailed paired-t tests as the direction of the MEP change was predicted to be a decrease (reduced excitability).

2.2.4.2 Results

Figure 2-6 presents the change in mean peak-to-peak amplitude MEPs at three time points (pre-rTMS, post-rTMS, and 15 mins later after the end of rTMS) recorded from lip muscles in response to single-pulse TMS over the lip representations in M1. The mean amplitude of lip MEPs obtained pre- rTMS was 1.18 mV (±SE = 0.07). The mean amplitude was suppressed to 0.92 mV (±SE = 0.05) on average after 15 min rTMS over lip representation (21% reduction in post-rTMS 1; $t_4 = 2.77$, $p = .025$, $d = 1.91$) and remained suppressed to 0.95 mV (±SE = 0.04) on average 15 mins after the end of the low-frequency rTMS train (19% reduction in post-rTMS 2; $t_4 = 4.41$, $p = .006$, $d = 1.77$).

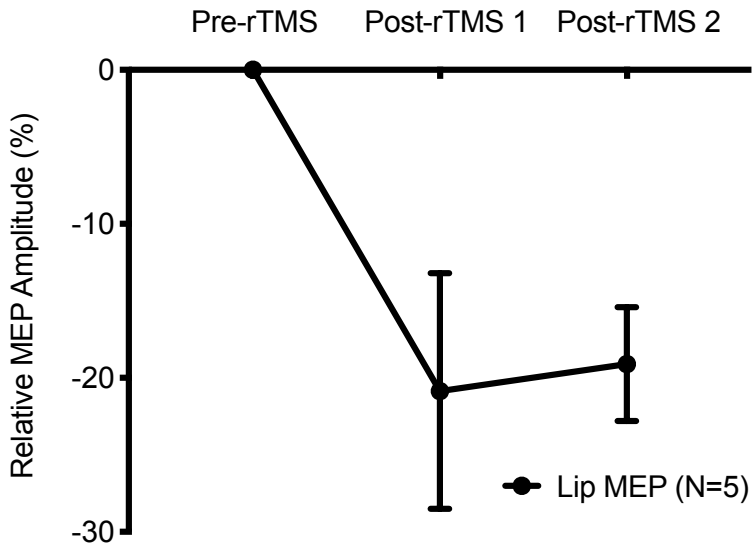


Figure 2-6 Effect of rTMS on cortical excitability in Mandarin speakers.

The figure shows mean reduction in peak-to-peak amplitudes (± SE) of post-rTMS lip MEPs (post-rTMS 1 and post-rTMS 2) relative to pre-rTMS lip MEPs.

2.2.4.3 Conclusions

Here, we confirmed that cortical excitability of speech motor cortex can be reduced in Chinese (N=5) participants using sub-threshold low-frequency rTMS and MEPs can be elicited from target muscles by applying single pulse TMS, which provides a direct and objective assessment of the motor system.

2.3 Electroencephalography (EEG)

EEG is an electrophysiological monitoring method to record electrical activity of the brain. It is typically non-invasive, with the electrodes placed on the scalp surface. It provides high temporal resolution, allowing researchers to detect activity generated by the brain at millisecond timescales.

The EEG reflects thousands of simultaneously ongoing brain processes, which means that the brain response to a single stimulus or event of interest is not usually visible in a single EEG recording trial. To observe the brain's response to a stimulus or event, the experimenter must conduct many trials and average them together, which results in random brain activity (noise) averaging out and the brain response of interest remains, called the event-related potential (ERP). ERPs refer to averaged EEG responses that are time-locked to sensory, motor or cognitive events.

2.3.1 Mismatch negativity (MMN)

The mismatch negativity (MMN) is a negative component of the ERP obtained by subtracting voltage changes in response to hearing a frequent stimulus (standard) from those to hearing an infrequent stimulus (deviant) in an oddball sound sequence (for a review, see Näätänen, Paavilainen, Rinne, & Alho, 2007). Previous studies have shown that the behavioural ability to discriminate speech sounds is related to the MMN amplitude. For example, phonemic contrasts of the native language that are easily detectable elicit large MMNs, whereas phonemic contrasts that do not exist in the native language are hard to detect and elicit smaller MMNs (Naatanen et al., 1997).

We used the MMN as an index of auditory discrimination in study 3 where we investigated the asymmetry of motor contributions to auditory discrimination in speakers of tonal and non-tonal languages. We chose the MMN because of three important features of MMN responses: the MMN is an ***automatic*** and ***preattentive*** ERP component, that is generated primarily in the ***auditory cortex***,

with maximal amplitudes over ***fronto-central scalp regions***. Below, I elaborate on each of these features and why they were important for our proposed study.

An important characteristic of the MMN in auditory oddball paradigms is the fact that it can be observed even when the participant is not paying attention. The MMN can be elicited during non-attentive states such as sleep (Sallinen, Kaartinen, & Lyytinen, 1994) or even in coma (Fischer et al., 1999). In addition, reliable MMN responses have been recorded in newborns (Cheour et al., 1998; Cheour, 2002) and even anesthetized rats (Ahmed et al., 2011). This demonstrates the brain's capacity to perform discrimination between consecutive sounds automatically (Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). Previous studies indicate that the involvement of motor areas in speech perception might be modulated by attention or task complexity (Sato et al., 2009). In order to eliminate the possible influence of attention or task complexity, we used the MMN response, which is preattentive, to evaluate participants' abilities to automatically detect phoneme and tone changes in sequences of syllables

Several studies have proposed mechanistic accounts of how the MMN might be generated (see Garrido, Kilner, Stephan, & Friston, 2009 for a review). Two most common explanations are: (i) the model adjustment hypothesis, which suggests that the MMN is a marker for error detection, caused by deviation from a learned regularity (Näätänen & Winkler, 1999; Winkler, Karmos, & Näätänen, 1996); and (ii) the adaptation hypothesis, which suggests that the MMN results from a simple mechanism of local neuronal adaptation at the level of the auditory cortex, causing attenuation and delay of the N1 response (Jääskeläinen et al.,

2004). Each of these hypotheses predicts that the major generators of MMN are located in the auditory cortex, and this claim is supported by several source reconstruction studies (e.g. Alho, 1995; Giard, Perrin, Pernier, & Bouchet, 1990; Hari et al., 1984). Localisation of sources to auditory cortex has been consistently reported across a variety of different modalities, including EEG (Jemel, Achenbach, Müller, Röpcke, & Oades, 2002; Marco-Pallarés, Grau, & Ruffini, 2005), MEG (Tiitinen et al., 2006), combined EEG with MEG (Rinne, Alho, Ilmoniemi, Virtanen, & Näätänen, 2000), fMRI (Molholm, Martinez, Ritter, Javitt, & Foxe, 2005; Rinne, Degerman, & Alho, 2005), and combined EEG-fMRI (Liebenthal et al., 2003).

The MMN peaks at about 100–250 ms from the onset of the change (here a syllable onset) and exhibits the strongest amplitude in fronto-central regions of the topographic scalp map (Näätänen et al., 2007; Näätänen et al., 1997). In study 3, we aimed to use low-frequency rTMS to temporarily disrupt the left or right motor cortex by placing a 70-mm figure-8-coil over the target area. The TMS coil limited the number of electrodes that could be used, since the electrodes could not be placed directly under the coil. Therefore, we used a custom electrode cap with 11 electrodes (Fz, FCz, Cz, CPz, Pz, F1, F2, P1, P2, left, and right mastoids), which provided maximal coverage over sites of interest. As previously, we also found that MMN responses were maximal over FCz so focussed our statistical analyses on measurements from this electrode. Data from all electrodes is available, however, at https://osf.io/en7uq/.

To sum up, measuring the MMN offers benefits over performance measures in tasks, in which it is uncertain whether the rTMS-induced disruptions affect task-related processes rather than modulating auditory processing. Task performance also requires the listener to attend to the relevant features of the speech sounds, whereas the processing of speech sounds measured by the MMN is preattentive. As such, the choice of MMN to determine the discriminability of speech-sound stimuli in odd-ball sequences offers a “purer” measure of this auditory function relative to task-based measures.

Chapter 3 Altered auditory feedback induces coupled changes in formant frequencies during speech production

3.1 Abstract

Here, we used real-time manipulations of speech feedback to test the independence of speech control of individual formants. During repetitive productions of “head”, “bed”, and “dead”, either the first (F1) or the second formant (F2) of vowels was shifted and fed back to participants. We then examined whether changes in production in response to these alterations occurred for only the perturbed formant or both formants. In Experiment 1, results showed that participants who received increased F1 feedback decreased the frequency of F1 production in compensation, but also increased the frequency of F2. The combined F1-F2 change moved the utterances closer to the vowel category /ɪ/ as for “hid”, “bid”, “did”. In Experiment 2, we further showed that a downshift in the frequency of F2 feedback also induced significant compensatory changes in both the perturbed (F2) and unperturbed formant (F1) that were in opposite directions. Taken together, the results demonstrate that a shift in auditory feedback of a single formant drives predictable changes in it and in related formants. The results suggest that, although formants can be controlled independently, the speech motor system may favour a strategy in which changes in formant production combine to fall within specific regions of the vowel space corresponding to existing speech-sound categories.

3.2 Introduction

Speaking is one of the most complicated motor behaviours. Fluent speech production requires fine coordination of many muscles and integration with auditory and somatosensory feedback. It is clear that auditory feedback is particularly important in speech development; pre-lingually deaf children have difficulties with speech articulation (Svirsky et al., 2004). It is also important for the maintenance of speech intelligibility, which can deteriorate in post-lingually deaf adults (Waldstein, 1990). Such individuals can rely on somatosensory feedback from orofacial muscles to achieve normal speech articulation (Nasir & Ostry, 2008). Even among individuals with normal hearing, there are differences in the extent to which somatosensory feedback is favoured over auditory feedback when monitoring speech production (Lametti, Nasir, & Ostry, 2012; Tremblay et al., 2003).

It is possible to examine the specific role of auditory feedback in speech motor control using a learning paradigm involving speech adaptation (Houde & Jordan, 1998). This paradigm was based on visuomotor adaptation in the limb movements, which has been studied extensively. The basic speech adaptation paradigm involves altering frequencies that characterize vowel sounds, feeding these back to the speaker in near real time (e.g., with less than a 40-ms delay), and measuring the compensatory changes in the vowel sounds the speaker subsequently produces (Houde & Jordan, 1998, 2002; Lametti, Smith, Watkins, & Shiller, 2018; Mollaei, Shiller, & Gracco, 2013; Purcell & Munhall, 2006b, 2006a). In speech, vowel sounds are defined by vocal tract resonances known as formants. Formants are distinctive frequency components evident in the

acoustic signal reflecting the size and shape of different chambers in the vocal tract. The first two formants (F1 and F2) in a vowel sound contain most of the acoustical energy and are sufficient to disambiguate vowel sounds. By examining the relationship between the changes in the first two formants of vowels produced when speech feedback is altered, we can gain insight into how speech movements are controlled. For example, if the frequency of F1 for the vowel in “head” is increased, it sounds closer to the F1 frequency of the vowel in “had”. Participants can compensate by lowering the frequency of F1 in their subsequent productions, which would counteract the acoustic change applied and result in an F1 that with the frequency shift applied and fed back sounds closer to that of the intended vowel in “head”. Alternatively, participants could change the frequency of both F1 and F2 in their productions; a combination of decreasing F1 frequency whilst simultaneously increasing F2 could move their production closer to that of the vowel in “hid”, a familiar region of the vowel space (see Figure 3-1). Determining whether adaptive speech movements are characterized by specific changes along separate dimensions, (e.g., F1, which corresponds to tongue height, and F2, which corresponds to front-back position of the tongue body; Ladefoged, 2001) or are controlled more globally by the production of vocal tract shapes that involve both dimensions, would inform our understanding of the nature of representations in models of speech motor control.

Previous studies perturbed both F1 and F2 feedback and showed that speakers compensated by changing production of both F1 and F2 frequencies in directions opposite to that of the applied shift (Houde & Jordan, 2002; MacDonald, Goldberg, & Munhall, 2010; Munhall et al., 2009). However, a more recent study altered F1

and F2 feedback simultaneously and found that both children and adults, with or without a speech problem (stuttering), only compensated by changing their F1 production (Daliri, Wieland, Cai, Guenther, & Chang, 2018). These results indicate that the speech motor system is able to modify the shape of the vocal tract to affect production of individual formants.

The independence of speech control of individual formants was more carefully tested by shifting the frequency of single formants in feedback and examining the compensatory responses in production of the shifted and unshifted formants. Again, findings from previous studies are inconsistent, some providing evidence for changes limited to F1 production during adaptation to F1 altered feedback, and others finding combined changes involving compensatory responses in both F1 and F2 feedback. For example, previous work involving adaptation to gradual shifts in F1 showed compensation in F1 production but no significant change in F2 (Shum, Shiller, Baum, & Gracco, 2011; Villacorta et al., 2007). Another study found changes in both F1 and F2 when feedback in F1 was changed abruptly (MacDonald, Purcell, & Munhall, 2011). In this case, the direction of change in F2 was the same for both upwards and downwards shifts of F1 and occurred more gradually than the change in F1, which suggested that independent processes governed the different formant changes. On the other hand, findings from several studies support the idea that formants are not modified independently. For example, reliable F2 changes in response to shifts in F1 were observed in the opposite direction to the compensatory changes in F1 (Lametti, Smith, Freidin, et al., 2018; Rochet-Capellan & Ostry, 2011).

In the current study, we revisit the idea that speakers counteract errors in formant production by only changing the altered formant. In the first experiment, we increased the frequency of F1 and examined compensatory changes in both F1 and F2. Based on previous findings (see above), we predicted that participants would decrease their F1 to offset the perceived acoustical error. We also predicted that increases in F2 would accompany the F1 decrease to move speech closer to a known region of the vowel space (Lametti, Smith, Freidin, et al., 2018; Rochet-Capellan, Richer, & Ostry, 2012). In other words, when participants produced “head”, “bed”, and “dead” and heard an upward shift in F1, an increase in F2 concurrent with the compensatory decrease in F1 would place the new production closer to “hid”, “bid”, “did” in vowel space (see Figure 3-1). In a second experiment, we decreased the frequency of F2 and examined compensatory changes in both F2 and F1. We predicted that increases in F2 would be observed to offset the induced acoustical error, and, as for responses to increased F1, decreases in the unaltered formant (this time F1) would accompany these F2 increases to move speech closer to a known region of the vowel space (Figure 3-1). Such a result would support the idea that, although F1 and F2 can be controlled independently (MacDonald et al., 2011), the speech motor system favours a strategy that moves formant production to specific regions of the vowel space.

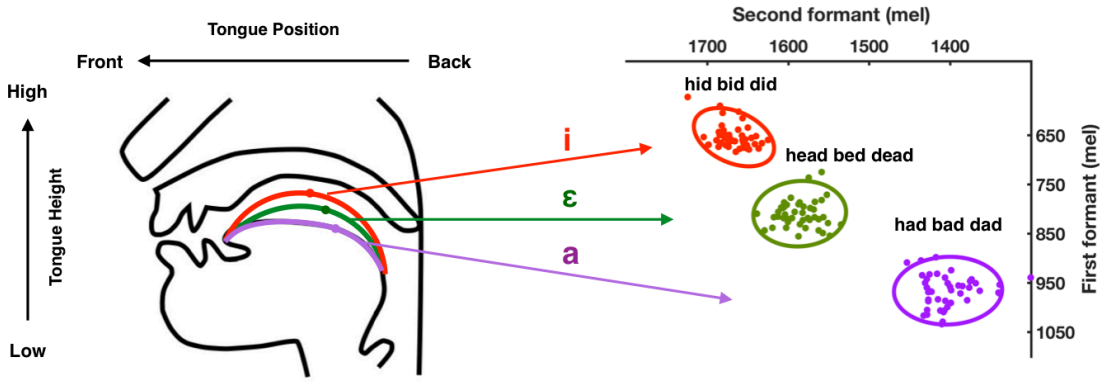


Figure 3-1 Vowel production.

Schematic showing the relationship between tongue height and position in the vocal tract (left) and the corresponding frequencies of the first and second formants of vowels (/ɪ/, /ɛ/, and /æ/) in vowel space from a single participant in the current study (right). Data points correspond to F1 and F2 values of single utterances during the vowel exploration phase in a representative participant. The ellipses represent a 90% confidence interval around the data points of same colour.

3.3 Methods

3.3.1 Participants

Forty native English speakers between the ages of 18 and 40 years, with no reported history of speech, language, or hearing disorders took part in the study. The Central University Research Ethics Committee (at the University of Oxford) approved the experimental protocol and participants gave informed consent. There were 20 participants in Experiment 1 (21.55 ± 5.27 years; 18 females) and 20 different participants in Experiment 2 (20.15 ± 3.59 years, 14 females).

3.3.2. Apparatus

The experiment was conducted in a quiet room with participants seated in front of a computer screen, which was used to present words. During the experiment, participants wore over-ear headphones (Sennheiser, HD 280 Pro) and a head-mounted microphone (Shure, WH20) that was placed approximately 4-7 cm away from the right corner of the participant's mouth. A Matlab Mex-based program Audapter (Cai, Boucek, Ghosh, Guenther, & Perkell, 2008; Tourville, Cai, & Guenther, 2013) was used to alter speech and play it back to participants with an unnoticeable delay.

3.3.3 Procedure

In both experiments, consonant-vowel-consonant words were presented on the computer screen for 1500 ms, one at a time. The inter-trial-interval was 750 ms. Participants were told that they would have to read out the words on the screen, and that they would hear their own voice through headphones.

First, participants completed a vowel exploration session with normal feedback in which they produced nine different words (dead, bed, head, dad, bad, had, did, bid, hid, containing three different vowels /ɛ/, /æ/ and /ɪ/ correspondingly) 15 times each. This session allowed participants to explore the vowel space. It also provided us with estimates of F1 and F2 frequencies for each vowel category in each individual. Vowel exploration was followed by three experimental phases: Baseline, Learning, and Unlearning (Figure 3-2). During the baseline phase, three different words "dead", "bed", and "head", containing the same vowel /ɛ/, were produced 15 times each by participants with normal feedback (2 min). In

the learning phase (10 min), participants produced “dead”, “bed”, “head” 75 times each with altered auditory feedback designed to induce sensorimotor learning (see Real-Time Feedback Alteration). In the final phase, Unlearning, “dead”, “bed”, and “head” were produced 30 times each with normal feedback (4 min). The order of words was randomized within each phase.

Three short noise-masked sessions preceded each of these experimental phases. During these noise-masked sessions, participants produced the words “dead”, “bed”, and “head” five times each but, instead of hearing feedback of their speech, they heard masking noise that scaled with the amplitude of the produced speech signal (i.e. the signal-to-noise ratio was 0 dB). White noise was also added to block participants’ ability to hear their own speech as much as possible.

The three noise-masked sessions were carried out at the following time-points: 1) Pre-Baseline 2) Before-Learning, and 3) After-Learning. The aim of these sessions was to measure whether feedforward changes in speech production associated with sensorimotor learning would persist in the absence of auditory feedback. Figure 3-2 shows a schematic of the Experimental Procedure.

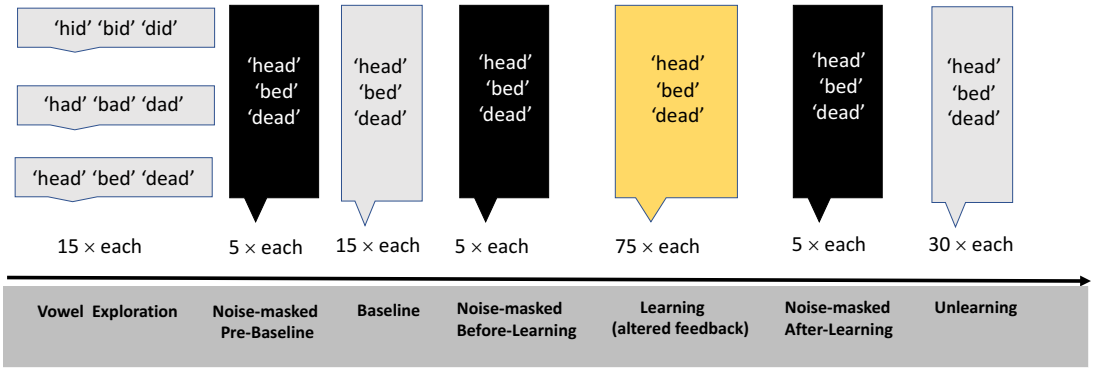


Figure 3-2 Experimental procedure.

The study involved Vowel Exploration, Baseline, Learning, and Unlearning, interleaved with three short noise-masked sessions.

3.3.4. Real-Time Feedback Alteration

To induce sensorimotor learning in speech, the frequency of either F1 or F2 was altered and played back to participants with an imperceptible delay (~18ms). The formant alteration was applied in mels — a perceptual scale characterizing changes in frequency that are judged by listeners to correspond to equal changes in pitch (Stevens, Volkmann, & Newman, 1937). The transformation from Hz to mel used was:

$$mel = 1127.01048 \times \log(1 + Hz / 700)$$

Participants were randomly assigned to one of two groups. The F1+110 group (Experiment 1) experienced a 110-mel increase in first formant (F1) production, while the F2-110 group (Experiment 2) experienced a 110-mel decrease in second formant (F2) production. A feedback shift magnitude of 110 mel was selected for this study because it is comparable to the F1 feedback perturbation used previously (Lametti, Smith, Freidin, et al., 2018). Both the increase in F1 frequency (F1+110) and the decrease in F2 frequency (F2-110) of the vowel /ɛ/ in the produced words resulted in the percept of speech with formant frequencies that were closer to F1 or F2 frequencies respectively of the vowel /æ/ in the vowel space (Fig 3-1). Compensatory responses were expected to be produced by speakers to counteract the perceived shift, either by reducing F1 (Experiment 1: F1+110 group) or increasing F2 (Experiment 2: F2-110 group) (Fig 3-1). We was also interested in the extent to which speakers altered their production of the unshifted formant in each experiment.

3.3.5 Auditory analysis

Speech was recorded at 16000 Hz in MATLAB and analysed using custom-written MATLAB scripts. First, the F1 and F2 frequencies of each vowel were extracted from the data output of Audapter, which estimates the first two formant frequencies using linear predictive coding. Then the average formant frequency was calculated from a 30-ms segment at the centre of each vowel. Data from one participant during the final noise-masked session were failed to record because of a technical problem. Utterances that contained production errors or contained formant frequencies that were greater than three standard deviations from a participant's mean F1 and F2 values in each phase of the experiment were excluded from further analyses. This resulted in approximately 3% of the data being excluded.

3.3.6 Statistical analysis

In each participant, the F1 and F2 frequency data for utterances produced during the learning and unlearning phases of the experiments were normalised by subtracting the mean and dividing by the standard deviation of the formant frequencies produced during baseline (with normal feedback) to create a standardised Z value for each datapoint. Data normalised by subtracting the mean values for F1 or F2 at baseline from the values for each utterance in the learning and unlearning phases were also statistically compared. The pattern of results did not change when the data were normalised in this traditional way (results for raw data are presented in Appendix 1).

Following Lametti, Smith, Freidin, et al., (2018), we assessed the significance of compensatory changes in the shifted and unshifted formant frequencies in each experiment using a one-sample t-test on the average Z values of the last 30 trials in the learning phase (against Z = 0, no change from baseline). As the directions of the compensatory change to the formant feedback shifts were predictable, we used directional t-test with a significance level of $p < .05$. We also averaged the Z values across all the trials in the learning phase to capture differences in rate of learning that are not captured by only analysing the end point of this phase. In the current study, we also used noise-masked sessions before the baseline, before and after the learning phase. Data obtained in these sessions allowed us to compare the effect of the perturbation on speech production under the same conditions when feedback was blocked. The pre-baseline session was used to normalise the before- and after-learning sessions as described above (we first confirmed that there was no change in F1 and F2 in the sessions performed either side of the baseline; $p > .05$). Then average change from before-learning session in produced F1 and F2 during after-learning session was calculated and compared with zero using directional t-tests to examine if there was a significant change from before to after-learning.

Following MacDonald et al., (2011), we assessed the significance of the return-to-baseline changes in the shifted and unshifted formant frequencies in the unlearning phase of each experiment. The average Z values of the last 15 trials in the unlearning phase were compared using a one-sample t-test (against Z = 0, i.e., the baseline average).

For each participant, we also calculated the Euclidean distance between the centre of the vowel clouds for “hid”, “bid”, “did” produced during the vowel exploration phase of the experiment and “head”, “bed”, “dead” produced at baseline and at the end of the learning phase (last 30 utterances) (Lametti, Smith, Freidin, et al., 2018). Learning-related changes in this distance were then compared to zero (no change) with one-sample t-tests to examine whether there was a significant change in the distance induced by F1 (experiment 1) or F2 (experiment 2) feedback perturbation. Distance changes were also compared between groups (experiment 1 and experiment 2) using a two-tailed t-test. With sample size of 20 participants in each experiment, we had 80% power to detect a medium effect size (Cohen’s d = 0.58) using an alpha of .05 for the one-sample t-test (one-tailed).

3.3.7 Unplanned Analyses

In an exploratory analysis, we compared the direction of compensation in Experiment 1 and Experiment 2. To do this, we calculated the angle between the vector that represents the direction of actual compensatory change from baseline and the vector representing the direction to move from “head”, “bed”, “dead” to “hid”, “bid”, “did” for each participant. The angles were then compared to zero (no difference from required angle) with one-sample t-tests to examine whether there was a significant difference between actual movement direction and the direction required in each Experiment separately. The angles were also compared between groups (experiment 1 and experiment 2) using a two-tailed t-test.

3.4 Results

3.4.1 Experiment 1: F1 shift of +110 mel

Figure 3-3 shows how F1 and F2 produced by participants, normalized as Z-scores, changed during the sensorimotor learning and unlearning phases in response to a 110-mel increase in F1. By the end of the learning phase, participants had decreased their F1 on average by 30 mel to compensate for the increase in F1 and increased F2 on average by 15 mel (Figure 3-3B). Directional t-tests showed that both F1 and F2 frequency at the end of the learning phase (last 30 utterances of learning phase) were significantly different from zero (baseline average) (F1: t_{19}= 6.24, p < .001, d = 1.40; F2: t_{19} = 2.88, p = .005, d = 0.64). Analysis of the Z normalised values for F1 and F2 averaged across the whole learning phase similarly showed a significant difference from baseline (F1: t_{19}= 5.82, p < .001, d = 1.30; F2: t_{19} = 2.10, p = .05, d = 0.47).

Compensatory responses of F1 and F2 production were also evaluated under noise-masked conditions, which were used as a measure of changes from before to after learning in the same context (i.e. with no feedback). Figure 3-3C shows the change in normalized F1 and F2 for heavily noise-masked speech from before and after learning. After learning, F1 significantly decreased (t_{18} = 4.12, p < .001, d = 0.95) and F2 significantly increased (t_{18} = 3.00, p = .004, d = 0.69).

By the end of unlearning phase, F2 production returned to baseline (one-sample t-test against baseline average, Z = 0: t_{19} = 1.42, p = .172), while F1 remained

significantly reduced relative to baseline (t_{19} = 2.33, p = .031, d = 0.52) after these 90 trials with normal feedback (Figure 3-3D).

In sum, in response to an F1 feedback perturbation, both F1 and F2 in production were observed to change though the magnitude of the change in the shifted formant (F1) was larger than in the unshifted one (F2). The results provide support for the idea that a single formant perturbation will induce a combination of changes in the shifted formant as well as the unshifted formant. Such combined changes could place the new production closer to a learnt pattern of speech production used to produce another vowel sound (“hid”, “bid”, “did” in current study; Figure 3-5A). This hypothesis was tested by comparing the magnitude (Euclidean distance) and direction of the compensatory shifts in F1 and F2 in vowel space between the pre- and post-learning productions of (“head”, “bed”, “dead”) and the production of “hid”, “bid”, “did” during the vowel exploration phase of the experiment before learning had occurred.

As shown in Figure 3-5B, participants in Experiment 1, responding to an increase in F1 feedback, moved their productions of “head”, “bed”, “dead” closer to the vowel space occupied by “hid”, “bid”, “did” by a magnitude of 32 mel on average (one-sample t-test against zero, t_{19} = 5.44, p < .001, d =1.22). The direction of this change was not significantly different from that predicted (Figure 3-5C; t_{19} = 1.45, p = .164). The same results were obtained by analysis of the data from the noise-masked session after learning.

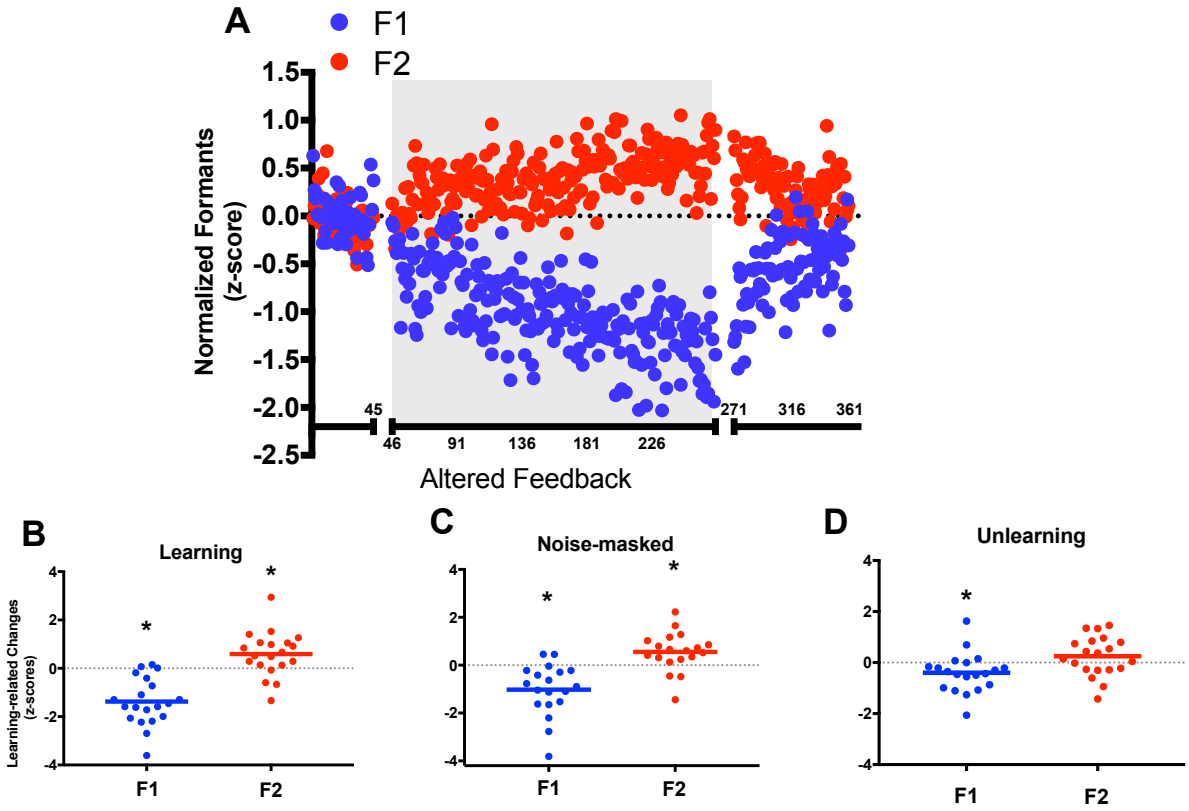


Figure 3-3 Experiment 1: F1 shifted upwards by 110 mel. Change in normalized formant frequencies during speech production.

*Values for F1 are shown in blue and F2 in red. The dashed line reflects no change from baseline. The graph in the top panel (A) shows the group average changes across learning and unlearning phases of the experiment relative to baseline; solid circles indicate the group average for each trial; the shaded area represents the trials where the altered feedback was applied. The graphs in the bottom panel (B, C and D) show average changes in normalised formant frequencies for individual participants for the Learning (B), Noise-masked (C) and Unlearning (D) phases of the experiment relative to baseline; the solid lines represent the group mean; * indicates significant change from baseline.*

3.4.2 Experiment 2: F2 shift of -110 mels

In Experiment 2, we gave participants feedback with a downshift in F2 and measured the changes made in production of F1 and F2. We predicted that participants would show a conceptually similar compensatory pattern to that

observed in Experiment 1: F2 frequency would increase, to offset the perceived acoustical error, but F1 would also change (decrease) to move production changes closer to “hid”, “bid”, “did” (see Fig 3-1).

Figure 3-4 shows how F1 and F2 produced by participants, normalized as Z-scores, changed during the sensorimotor learning and unlearning phases in response to a 110 mel decrease in F2. By the end of the learning phase, participants had increased their F2 on average by 50 mel to compensate for the decrease in F2 and decreased F1 on average by 14 mel (Figure 3-4B). Directional t-tests showed that both F1 and F2 frequency at the end of the learning phase (last 30 utterances of learning phase) were significantly different from zero (baseline average) (F2: $t_{19} = 7.79$, $p < .001$, $d = 1.74$; F1: $t_{19} = 2.30$, $p = .017$, $d = 0.51$). Analysis of the Z normalised values for F1 and F2 averaged across the whole learning phase similarly showed a significant difference from baseline (F1: $t_{19}= 2.08$, $p = .026$, $d = 0.47$; F2: $t_{19} = 7.07$, $p < .001$, $d = 1.58$).

Figure 3-4C shows the change in normalized F1 and F2 frequency for heavily noise-masked speech from before and after learning in response to a decrease in F2 feedback. There was a significant increase in F2 production ($t_{19} = 7.12$, $p < .001$, $d = 1.59$), and a significant reduction in F1 ($t_{19} = 1.99$, $p = .030$, $d = 0.45$) from before to after learning.

By the end of the unlearning phase, F2 frequency returned to baseline (one-sample t-test against baseline average, Z = 0: $t_{19} = 1.95$, $p = .066$), while F1

frequency remained significantly reduced relative to baseline (t_{19} = 2.31, p = .032, d = 0.52) after these 90 trials with normal feedback (Fig 3-4D).

The results of Experiment 2 provide further support for the idea that feedback perturbations induce changes not only in the shifted formant, but also in the unshifted formant. As in Experiment 1, we evaluated whether these combined changes served to move the produced speech token closer to another vowel sound. The vector analysis confirmed a learning-related reduction in the Euclidean distance between “head”, “bed”, “dead” and “hid”, “bid”, “did” in F1-F2 space Experiment 2 (Figure 3-5B). By the end of learning, the productions of “head”, “bed”, “dead” were 36 mel closer to the productions of “hid”, “bid”, “did” in response to the F2 shift-down (one-sample t-test against zero, t_{19} = 5.96, p < .001, d = 1.33). However, the direction of the change was significantly different from the angle predicted (Figure 3-5C; t_{19} = 6.86, p < .001, d = 1.53). Thus, although the results of Experiment 2 met our predictions—a decrease in F2 drove an increase in F2 and a decrease in F1—changes in production were not as well directed towards the “hid”, “bid”, “did” part of vowel space as they were in Experiment 1 (Figure 3-5A). The same results were obtained by analysis of the data from the noise-masked session after learning.

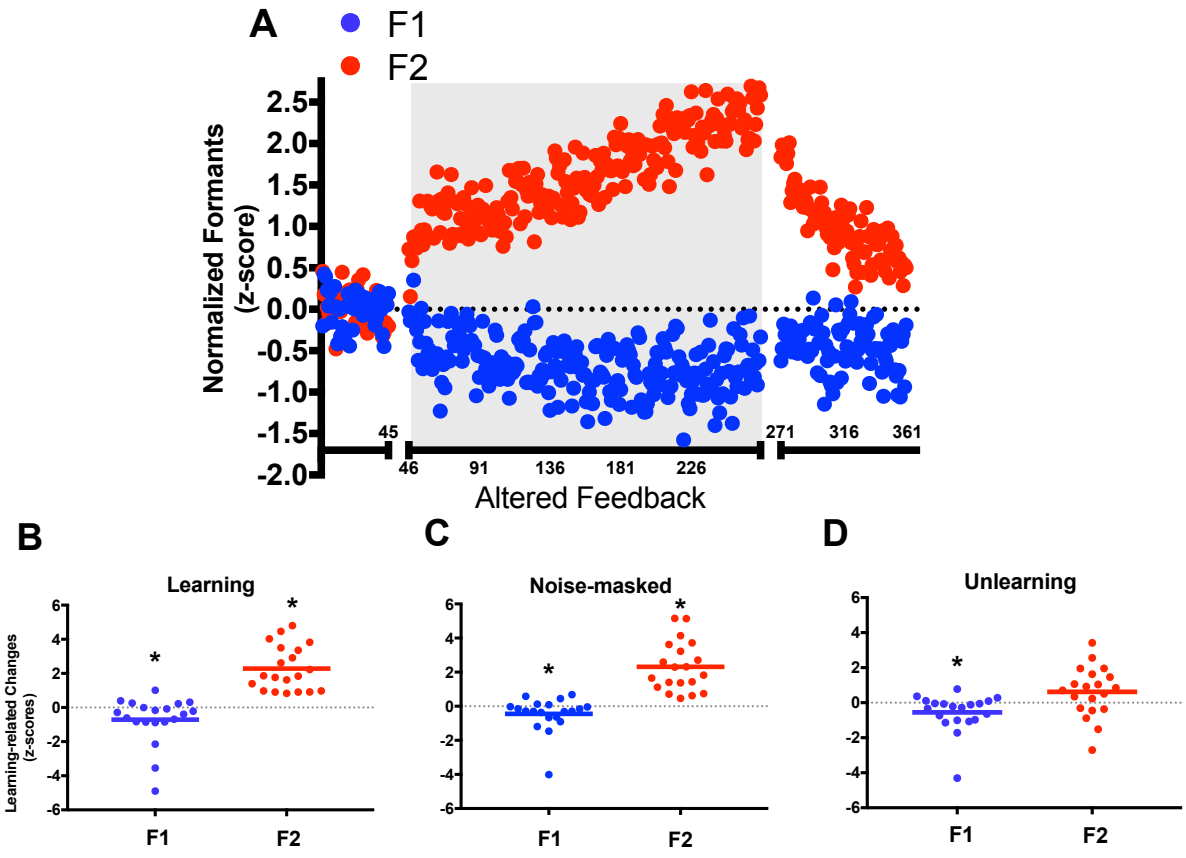


Figure 3-4 Experiment 2: F2 shifted downwards by 110 mel.

See legend to Fig. 3-3 for details.

3.4.3 Comparison of Experiment 1 and 2

We compared the changes in shifted and unshifted formants between the two experiments. The data were rectified as the directions differed between the two experiments (all data were made positive) and analysed using a two-tailed t-test. For the shifted formants, the magnitude of the learning-related change (last 30 utterances of the learning phase) was significantly greater in the F2-shift experiment (Exp 2) compared with the F1-shift experiment (Exp 1) (t_{38} = 2.48, p = .018, d = 0.79, Fig. 3-3B and 3-4B). However, for the unshifted formants, the magnitudes of the shift at the end of the learning phases of Experiments 1 and 2 did not differ (t_{38} = 0.99, p = .329, d = 0.31; Fig 3-3B and 3-4B). In the vector

analysis, there was no significant difference in the magnitude of the learning-related decrease in distance between Experiments 1 and 2 (t_{38} = 0.52, p = .610, d = 0.16; Fig 3-5B). However, the differences between the actual and predicted direction of the changes in each experiment were significantly different (t_{38} = 4.385, p < .001, d = 1.39), with a larger difference from the predicted direction for the actual direction of changes in Experiment 2 (Fig 3-5C).

Taken together, these findings suggest that single formant perturbation induced combined F1-F2 changes, which served to move the new production towards a known vowel category. The significantly larger amount of compensation in the shifted formant (F2) observed in Experiment 2 (in response to a shift down in F2) may explain the discrepancy between the actual and predicted direction required to move from the “head”, “bed”, “dead” to the “hid”, “bid”, “did” vowel space for that experiment. Nevertheless, the magnitude of the difference between the two vowel productions was significantly reduced following learning in both experiments. It is worth noting that when participants were debriefed after the study, 75% of participants who experienced the F2 shift indicated they were aware of the changes, while only 20% of the participants who experienced the shift upward in F1 noticed them.

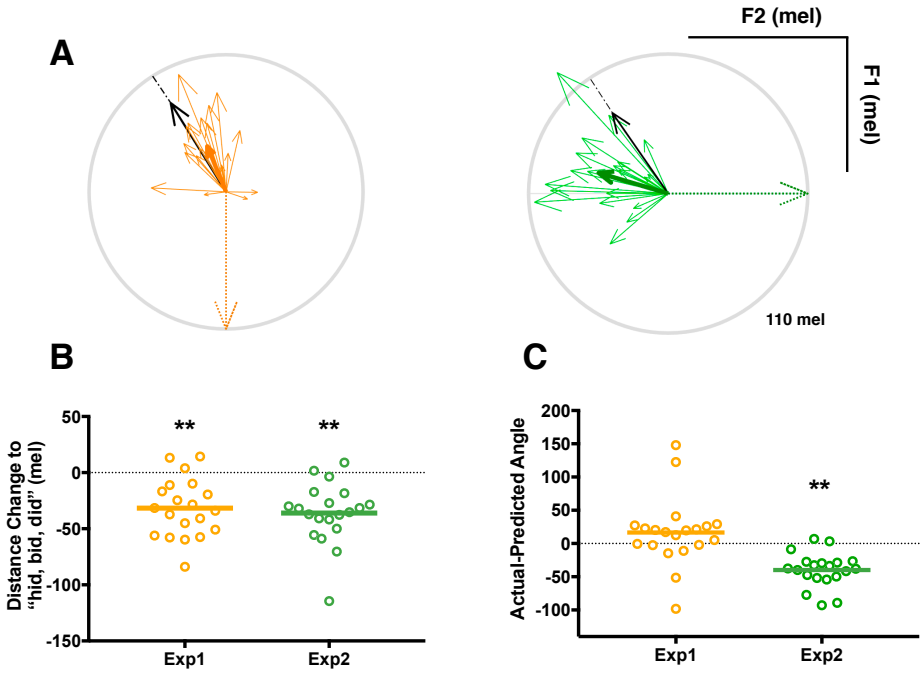


Figure 3-5 Comparison of changes in vowel production in Experiment 1 and Experiment 2.

Top panel: (A) Vector figures of experiment 1 (yellow) and experiment 2 (green). The vectors represent the direction and magnitude of the compensatory change from baseline in vowel space. The thin vectors represent data from individual participants and the thick vectors represent the group average for each experiment. The black arrow represents the average of the direction required to move from "head", "bed", "dead" to "hid", "bid", "did", while the dotted arrow represents the formant frequency shift we applied (F1+110 mel in Experiment 1 and F2-110 mel in Experiment 2). The light grey circle indicates 110 mel. Bottom panel: (B) Learning-related reduction in Euclidean distance between "head", "bed", "dead" and "hid", "bid", "did" in F1-F2 space. Negative values indicate that production of "head, bed, dead" by the end of learning has moved closer to "hid", "bid", "did". Bottom panel: (C) The difference between the angles of the direction of actual compensatory movement and predicted movement (to "hid", "bid", "did") for each participant. Open circles represent individual participants, the solid lines

*represent the group means. Distance changes (B) and Actual-Predicted angles (C) were compared to zero in each experiment. * $p < 0.05$ and ** $p < 0.01$*

3.5 Discussion

The current study investigated how speakers altered their production of the first two formants of vowels by manipulating the auditory feedback for F1 and F2 in two separate experiments. Our hypothesis was that, even though individual formants can be controlled independently to compensate for specific acoustic changes in feedback, the speech motor system may favour a strategy in which changes in formant production are coupled to maintain speech production within specific regions of the vowel space corresponding to existing speech-sound categories. We predicted that altered auditory feedback of a single formant during speech production would induce changes not only in the shifted formant, but also in the unshifted formant. Furthermore, We predicted that these combined changes would result in a pattern of speech production closer to the frequencies of an existing vowel category.

Consistent with our first prediction, the results of Experiments 1 and 2 showed that by the end of the adaptation (learning) phase, participants had significantly changed production of both the shifted and the unshifted formants. Specifically, those who received feedback with an increase in F1, produced speech with a significant decrease in F1 as well as a significant increase in F2. Those who received feedback with a decrease in F2, produced speech with a significant increase in F2 and a significant decrease in F1. The changes relative to speech prior to the application of the shift were evident both in the last 30 trials of the learning phase and in the noise-masked blocks that followed the learning phase.

The results of Experiment 1, in which F1 feedback was increased and participants compensated by decreasing F1 and increasing F2 in their productions, are consistent with previous findings (Lametti, Smith, Freidin, et al., 2018; Rochet-Capellan et al., 2012). However, the findings of both experiments in the current study contrast with previous work in which participants showed independent changes to the perturbed and unperturbed formants when feedback was gradually shifted in either F1 or F2 (MacDonald et al., 2011; Shum et al., 2011; Villacorta et al., 2007). In these studies, compensatory F1 and F2 changes were observed in the opposite direction to the applied shift in F1 and F2 respectively, as seen in our study. However, F1 did not change during F2 feedback perturbation, which contrasts with the findings for Experiment 2 in our study. Small F2 changes induced by F1 feedback shift were observed by MacDonald et al (2011) but these changes were decreases in F2 when F1 feedback was shifted both up and down. In contrast, in our study, F2 production increased when F1 was shifted up. Furthermore, the small reductions in F2 induced by perturbed F1 in the previous study were found to persist even after the feedback was returned to normal, whereas in our experiments, F2 returned to baseline values but the F1 changes while reduced were not quite returned to baseline (MacDonald et al., 2011).

The differences in results between our study and previous studies may stem from large variability among participants, smaller sample sizes, and differences in experimental designs. Unlike in studies of sensorimotor adaptation in limb movement, where nearly every participant shows compensatory responses after receiving sensory perturbation (Krakauer, Pine, Ghilardi, & Ghez, 2000), a fairly

large proportion of people in auditory studies exhibit little or no compensation for auditory perturbations (10-40%; Lametti et al., 2012; Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014; Rochet-Capellan et al., 2012). Furthermore, meta-analyses of auditory studies indicated that on average 22% of participants altered production in a pattern that follows (instead of opposes) the direction of the F1 shift (MacDonald et al., 2011). These inter-individual differences in behaviour complicate work aiming to observe the compensatory responses induced by single formant perturbation, especially for changes in the unperturbed formants, which are generally smaller. Even though relatively large variability is found among participants during speech motor learning, a sample size of 12 (or less) per group is typical for these studies. We propose that a larger sample size (N=20 in current study) might be necessary to observe group compensatory patterns induced by single formant shift. We also suggest that a longer learning phase during which feedback is perturbed is important to observe changes in unperturbed formant. For example, in one study (MacDonald et al., 2011), only 50 utterances were produced during the learning phase, which was assumed to be sufficient for speech adaptation to reach a plateau. In the current study, we witnessed changes in both the perturbed and unperturbed formants that continued to occur after 50 utterances (see Figs. 3-3 and 3-4). Therefore, we suggest that, in order to observe reliable compensatory responses to a single formant perturbation, it is important to a) have larger sample sizes, and b) increase the duration of the learning phase in which auditory feedback is altered. In addition, in the current study, we quantified learning-related changes by averaging the normalized F1 and F2 frequencies (Z-score) of the last 30 trials in the learning phase or across the whole learning session. In our view, the latter

method allows us to capture both rate and magnitude of adaptation during learning rather than just focus on a limited (and variable) number of utterances at the end of learning phase, which was used in more recent adaptation study (Lametti et al., 2020).

Current models of speech production suggest that the relationship between the movements used to articulate speech sounds and the sensory consequences is learned during speech development and used to build internal models that encode these sensorimotor mappings and guide production. Internal models allow the brain to predict the sensory consequences of actions, perform online correction and rapidly detect errors (Houde & Nagarajan, 2011; Tourville & Guenther, 2011). Our findings are compatible with the idea that speech movement parameters can be represented in internal models as coordinated and combined sets to achieve specific sensory goals rather than as precise one-to-one mappings between individual motor dimensions and acoustic outcomes. However, as we discussed above, results in current study differ from those found previously using a similar paradigm (MacDonald et al., 2011). On the basis of their findings, these authors concluded that the nervous system is capable of independent control of F1 and F2. Even so, the authors suggested that such precise control of single formant may be observed only during feedback error correction and this may not extend to general speech motor control. In our view, the representations in the internal models tested by feedback perturbation paradigms are likely to be the same as those used in general speech motor control. Based on the results of this study and previous studies, we speculate that the inverse and forward internal models may differ in the extent to which they

exert independent and precise articulatory adjustments to each acoustic parameter during feedback perturbation. A previous study showed that stimulation to the cerebellum affected only compensatory responses in F1 production in response to an F1 perturbation, whereas stimulation to the motor cortex affected the amount of compensation in both F1 and F2 in response to the same F1 perturbation (Lametti, Smith, Freidin, et al., 2018). This would suggest that the forward model, which is used to make rapid ongoing adjustments to speech production can independently adjust movements in single dimensions corresponding to separate acoustic changes. On the other hand, the inverse model may encode sensorimotor mappings that represent the combination or combinations of articulatory movements required to achieve a sensory target that exists in the speech repertoire.

Certainly, the combined F1-F2 changes observed in the current study could also be due to physical or articulator constraints. In other words, the change in unshifted formant could be a by-product of the result of compensation for the shifted formant, and the direction of unshifted formant change is decided by physical constraints. In the current study, the vowels ɪ-ɛ-æ are on the edge of the vowel space, which might suggest there is a constraint on the extent to which speakers can modify the vocal tract shape to simultaneously increase or decrease F1 and F2 formants (e.g. see acoustical simulation in (Badin, Boë, Sawallis, & Schwartz, 2014). We consider this to be unlikely, however, because the direction of combined F1-F2 changes induced by either F1 or F2 feedback shift met our prediction, which was to move the new productions closer to a known vowel category, rather than in any other direction (Figure 3-5). In

experiment 2, the direction of combined F1-F2 changes did not perfectly match in the direction to the expected vowel category (Figure 3-5). However, the discrepancy between the actual and predicted direction required to move from the “head”, “bed”, “dead” to the “hid”, “bid”, “did” is mainly driven by the significantly larger amount of compensation in the shifted formant (F2) observed in Experiment 2 (in response to a shift down in F2). For the unshifted formants, the magnitudes of changes at the end of the learning phases of Experiments 1 and 2 did not differ ($p = .329$; Figs. 3-3 and 3-4). We need to explain why the shift in F2 induced larger compensation in F2, compared to F1 compensation induced by the same amount of F1 shift. It is possible that the change in F2 produces a qualitatively different percept than the equivalent change in F1. Although we applied the same amount of shift (110 mel) to both F1 and F2, 75% of participants who experienced the F2 shift down indicated they were aware of the changes, while only 20% of the participants who experienced the shift upward in F1 noticed the changes. There may be more explicit processes at play in the F2-110 mel experiment (though see Munhall et al., 2009). Further work comparing compensation to different magnitudes of shift could clarify this point.

To summarize, we used altered auditory feedback to show that adaptive responses to feedback shifts applied to a single formant resulted in changes in production extending from the shifted to the unshifted formant. These combined changes moved productions closer to the frequencies of nearby existing categories in the vowel space. We conclude that although formants can be controlled independently, the speech motor system may prefer to make coupled changes in formant production. This latter strategy would serve to maintain

speech production within specific regions of the vowel space corresponding to existing speech-sound categories.

Chapter 4 Disruption of speech motor adaptation with repetitive transcranial magnetic stimulation of the articulatory representation in primary motor cortex

4.1 Abstract

When auditory feedback perturbation is introduced in a predictable way over a number of utterances, speakers learn to compensate by adjusting their own productions, a process known as sensorimotor adaptation. Despite multiple lines of evidence indicating the role of primary motor cortex (M1) in motor learning and memory, whether M1 causally contributes to sensorimotor adaptation in the speech domain remains unclear. Here, we aimed to assay whether temporary disruption of the articulatory representation in left M1 by repetitive transcranial magnetic stimulation (rTMS) impairs speech adaptation.

To induce sensorimotor adaptation, the frequencies of first formant (F1) were shifted up and played back to participants while they produced “head”, “bed”, and “dead” repeatedly (the learning phase). A low-frequency rTMS train (0.6 Hz, subthreshold, 12 min) over either the tongue or the hand representation of M1 was applied before participants experienced altered auditory feedback in the learning phase. Participants who received rTMS over the hand representation showed the expected compensatory response for the upwards shift in F1 by significantly reducing F1 and increasing the second formant (F2) frequencies in their productions. In contrast, such compensatory changes were largely abolished by rTMS applied over the tongue representation. Critically, rTMS

(subthreshold) over the tongue representation did not affect vowel production, which was unchanged from baseline. These results provide direct evidence that the articulatory representation in left M1 causally contributes to sensorimotor learning in speech. Furthermore, these results also suggest that M1 is critical to the network supporting a more global adaptation that aims to move the altered speech production closer to a learnt pattern of speech production used to produce another vowel.

4.2 Introduction

When we speak, we continuously monitor auditory and somatosensory feedback. If a mismatch between the expected outcome and actual sensory feedback is detected, we use the mismatch to gradually modify our subsequent movements. This form of adaptation is one of the hallmarks of the human speech motor control system, which plays a crucial role in the development and maintenance of speaking skills (Daliri, Wieland, Cai, Guenther, & Chang, 2018; Maas et al., 2008; Villacorta, Perkell, & Guenther, 2007).

Speech adaptation can be assessed using a real-time perturbation paradigm which has been utilized extensively over the past two decades. A basic type of speech adaptation paradigm involves altering the frequencies of the first (F1) or second (F2) vowel formants or both and playing these back to participants in near real time (Houde & Jordan, 1998; MacDonald et al., 2010). With vowels, the frequencies of the first two formants can largely determine which vowel is perceived. Figure 4-1a shows the relative positions of three vowel sounds (/ɛ/, /æ/ and /ɪ/) in vowel space. In a typical speech adaptation experiment, if the

frequency of F1 for the vowel /ɛ/ in “head” is increased, it is perceived to be more similar to the F1 frequency of the vowel /æ/ in “had”. After receiving such altered auditory feedback, participants learn to compensate by, for example, lowering F1 frequencies in their subsequent productions over the course of many repetitions.

Production changes resulting from such adaptive learning are retained for a period after the alteration is removed (Lametti, Smith, Freidin, et al., 2018; Lametti, Smith, Watkins, et al., 2018; Purcell & Munhall, 2006a), suggesting speech adaption involves modification of the feedforward motor commands that guide speech. According to the DIVA (Directions Into Velocities of Articulators) model, the primary motor cortex (M1) receives stored motor commands originating from the feedforward control system (putatively located in the left premotor cortex) and the corrective commands originating from the feedback control system (putatively located in the right premotor cortex and cerebellum), and configure the vocal tract to produce speech sounds (Tourville & Guenther, 2011).

The role of M1 in speech adaptation has received little attention; in contrast, it has been studied extensively in visuomotor adaptation. One recent study explored the possible effects of cerebellar and motor cortex tDCS on speech adaptation induced by real-time feedback perturbation (Lametti, Smith, Freidin, et al., 2018). In response to feedback perturbation of F1, anodal tDCS stimulation over the cerebellum increased compensation in F1, while anodal stimulation of the ventral motor cortex led to increased adaptation in both F1 and F2. This pattern of findings suggested that the M1 is involved in generating a global

adaptation that altered speech production towards the production of another vowel sound rather than compensation simply to offset the acoustic error, which was the pattern of behaviour facilitated by cerebellar stimulation. The electrodes used to deliver tDCS in the previous study were large, however, and covered ventral M1, as well as ventral premotor cortex and may have extended to stimulate portions of posterior inferior frontal cortex.

The current study used repetitive TMS to examine whether the articulatory representation in M1 causally contributes to speech adaptation. The perturbation approach provided by rTMS allows greater precision in terms of spatial focality relative to tDCS. It also offers a more convincing test of causality in terms of temporarily interfering with the function of an area whereas anodal tDCS induces facilitatory effects. We applied a low frequency rTMS train to either the tongue or the hand representation in left M1 for 12 min prior to speech adaptation. This train of subthreshold pulses successfully reduces cortical excitability for a further 10-15 min (Mottonen & Watkins, 2009). We predicted that inhibitory TMS over the tongue representation, but not hand representation would impair adaptive learning induced by the auditory feedback perturbation. Furthermore, we predicted that temporary disruption of the articulatory representation would impair speech adaptation not only in the shifted formant (F1), but also in the unshifted formant (F2). This prediction was based on findings from previous work indicating that altered auditory feedback of a single formant during speech production induced changes not only in the shifted formant, but also in the unshifted formant (MacDonald, Purcell, & Munhall, 2011; study 1), and that

neuromodulatory stimulation by tDCS over motor cortex enhances such combined changes (Lametti, Smith, Freidin, et al., 2018).

4.3 Method

4.3.1 Participants

Forty right-handed native English speakers between the ages of 18 and 40 years participated. Twenty participants received stimulation to the tongue area in left M1 and another twenty participants received stimulation to the hand area in left M1. Demographics of the twenty participants in each of the stimulation groups are given in Table 4-1. An additional 5 participants participated in a small proof-of-principle study to confirm the effect of low frequency rTMS over tongue representation on motor excitability as it was not possible to obtain this data at the same time as speech data in the main experiment; see Chapter 2 for details. All participants reported no hearing problems and no personal or family history of seizures or other neurological disorders. The Central University (of Oxford) Research Ethics Committee approved the experimental protocol. Participants gave informed consent and received payment for their participation.

4.3.2 Procedure

During the experiment, participants were seated in front of a computer screen. They were instructed to produce consonant-vowel-consonant words that were presented on the computer screen, into a head-mounted microphone (Shure, WH20) located approximately 4-7 cm away from the right corner of mouth. Each word was presented for 1500 ms, one at a time, with a 750 ms inter-trial-interval.

They heard their own voice through over-ear headphones (Sennheiser, HD 280 Pro).

Participants underwent speech adaptation as shown in Figure 4-1b. Before the adaptation experiment, participants produced “dead”, “bed”, “head”, “dad”, “bad”, “had”, “did”, “bid”, “hid”, 15 times each with normal feedback to explore the vowel space. Then, during baseline, participants produced three different words “dead”, “bed”, “head”, which contain the same vowel /ɛ/, 15 times each, again with normal feedback. To induce sensorimotor learning, auditory feedback was altered (see Real-Time Feedback Alteration) while participants produced “dead”, “bed”, “head” 75 times each; we refer to this as the learning phase (10 min). In the final phase, referred to as unlearning, participants produced “dead”, “bed”, “head” 15 times each again with normal feedback as in the baseline.

Participants received the 12-min rTMS train (see below) before the learning phase. Measures of motor excitability were also obtained in the hand group after baseline, before and after learning; the tongue group rested for the equivalent time. Furthermore, two noise-masked sessions were performed after the first and third motor excitability measurements (or equivalent rest), during which participants produced the same words as they produced in the vowel exploration but feedback of their speech was masked by noise that scaled with the amplitude of the produced speech signal (i.e. the signal-to-noise ratio was 0 dB). Unfortunately, the masking was not successful due to a technical problem, so these data were excluded from the experiment.

4.3.3 Real-Time Feedback Alteration

We used a Matlab Mex-based program Audapter (Cai et al., 2008; Tourville et al., 2013) to alter speech and play it back to participants in near real time (e.g., approximately a 18-ms delay). Participants experienced a 110-mel increase in the first formant (F1) production during the learning phase.

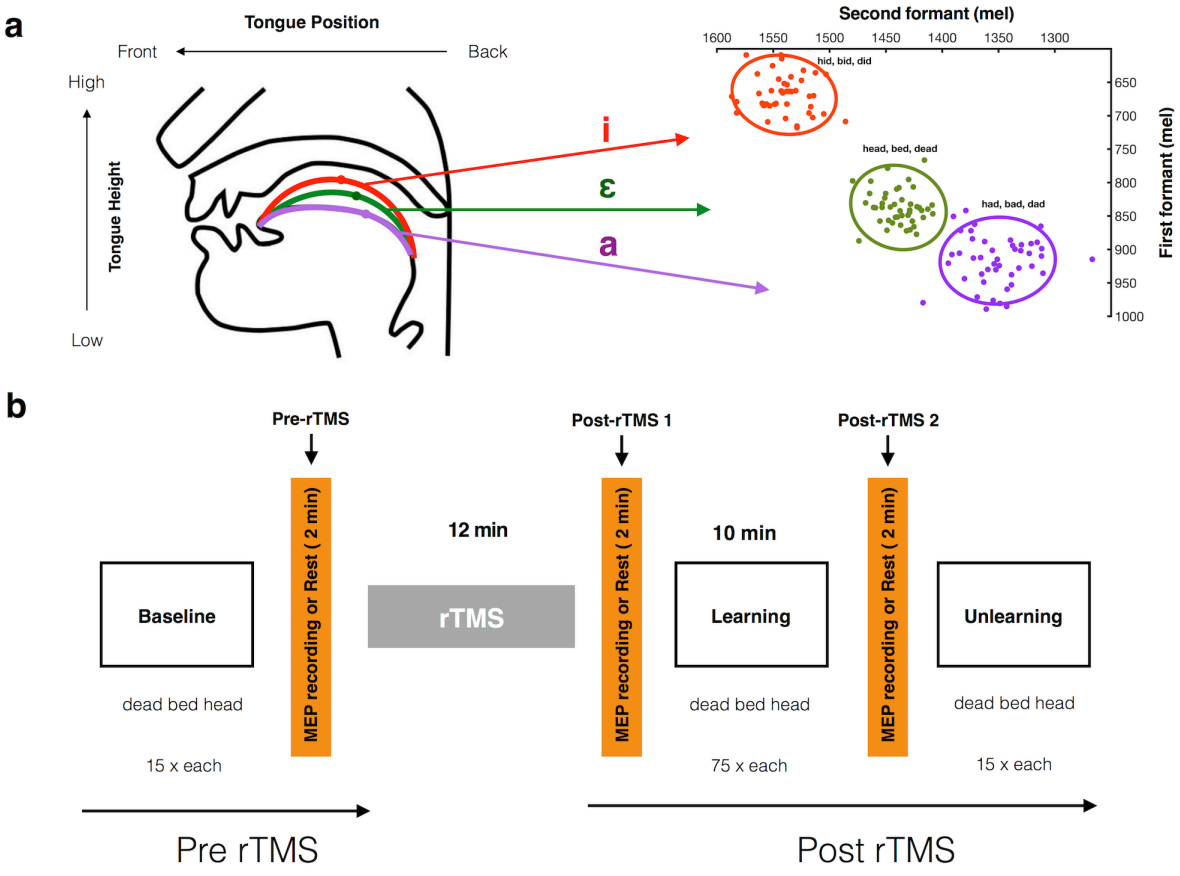


Figure 4-1 Vowel spaces and experimental procedure.

(a) Vowel spaces. Schematic showing the relationship between tongue height and position in the vocal tract (left) and the corresponding frequencies of the first and second formants of three vowels (/ɛ/, /æ/ and /ɪ/) in vowel space from a single representative participant in the current study (right). Data points correspond to F1 and F2 values of single utterances during the vowel exploration. The ellipses represent a 90% confidence interval around the data points of same colour. (b) Experimental procedure. There were three phases of the adaptation experiment: baseline, learning, and unlearning. 12min rTMS was applied before participants

experienced altered auditory feedback in the learning phase. In the hand group only, the changes in cortical excitability induced by the rTMS were measured using single pulses of TMS over the representation of the hand muscle at three time-points: 1) after baseline (pre rTMS), 2) immediately after repetitive TMS (post rTMS 1), and 3) following the learning phase (post rTMS 2) (orange bars). In the tongue group, participants did not receive TMS single pulses and rested for the equivalent time period. In addition, two noise-masked sessions occurred before rTMS and after the final MEP recording/rest block but these data were not included in the analysis (see text for details).

4.3.4 Transcranial Magnetic Stimulation

A low-frequency rTMS train (0.6 Hz, subthreshold, 12 min) was delivered over either the tongue or the hand representation of M1 cortex before the learning phase. All TMS pulses were monophasic, generated by two Magstim 200s and delivered through a 70-mm figure-8 coil connected through a BiStim module (Magstim, Dyfed, UK).

After baseline, the hand or tongue representation in the motor cortex was localised using single pulses of TMS. For the hand group, the representation of the first dorsal interosseus muscle in the left hemisphere was localised as previously described (Mottonen & Watkins, 2009). The coil was placed tangential to the skull, to induce a current flow from posterior to anterior under the junction of the two wings of the figure-eight coil. The position of the coil over the lateral scalp was adjusted until a robust motor-evoked potential (MEP) was observed in the contralateral target muscle. After localization, single pulses of TMS were applied over the hand representation to determine the active motor threshold, that is, the minimum intensity at which TMS elicited at least 5 out of 10 MEPs

with an amplitude of at least 500 μV when the target muscle was contracted at 20–30% of the maximum. For the tongue group, we were unable to use tongue MEPs for localisation and to measure changes in excitability before and after the learning phase (as in the hand group), because attaching electrodes to the tongue muscles to obtain electromyography (EMG) signal would affect a participant's speech production. We, therefore, first localised the hand and lip representations (using EMG electrodes to record from the orbicularis oris muscle; Möttönen et al., 2009) and established the active motor threshold for the lip representation, that is, the minimum intensity at which TMS elicited at least 5 out of 10 MEPs with an amplitude of at least 200 μV when the target muscle was contracted at 20–30% of the maximum. The mean active motor thresholds of each stimulation group are given in Table 4-1. To estimate the location of the tongue representation, we moved the coil ventrally from the lip representation for half of the distance between hand and lip representations along the trajectory of the line connecting the two. The distance across the scalp between the "hot spots" for tongue and lip representations varied among participants from 12.5 mm to 22.5 mm. The intensity of each participant's active motor threshold was then used for 12 min rTMS (the intensity of active motor threshold for lip representation was used for tongue representation). During the 12-min rTMS train, participants watched a silent nature film and were instructed to stay still and relax their muscles (hence the stimulation is subthreshold). During repetitive TMS, the EMG signals were carefully monitored to ensure that muscles were relaxed, and no MEPs were evoked in the target muscle. The coil was changed at the start and halfway through repetitive TMS to prevent overheating.

The success of disruption induced by rTMS in the hand group was evaluated by comparing the size of MEPs elicited by single pulses of TMS over the representation of the target muscle after baseline (pre-rTMS), immediately after the end of the rTMS train (post-rTMS 1), and after learning phase (post-rTMS 2). At each time point, 20 TMS pulses were applied. The intensity used for single pulses TMS was established following localization, as the lowest percentage of stimulator output that elicited MEPs in the target muscle of peak-to-peak amplitude of at least 2 mV on 10 consecutive trials during muscle contraction (see table 4-1 for the mean stimulator intensity used to elicit MEPs in the hand muscle). Because we did not record MEPs from tongue muscles in the tongue group (due to interference from electrodes on speech production), single pulse TMS recordings at the three time-points described above were replaced by two minutes breaks (rest).

4.3.5 Electromyography recordings

Using a CED 1902 amplifier, a CED 1401 analog-to-digital converter, and a PC running Spike2 (Cambridge Electronic Design), the EMG signal was sampled (5000 Hz), filtered (1 Hz–1 kHz bandpass) and displayed on a computer screen. Participants used this display to practise producing a constant level of contraction of target muscles (20–30% of the maximum). Disposable electrodes were attached to the right orbicularis oris muscle or the first dorsal interosseous muscle of the right hand, to record EMG signal from lip or hand respectively.

4.3.6 MEPs Analysis

MEPs were analysed trial-by-trial using custom-written MATLAB scripts. For hand MEPs, peak-to-peak amplitude was automatically computed from 15 to 39 ms post-TMS. The automatic windowing was checked manually by the experimenter. The absolute value of the background muscle activity was averaged across the 100ms before the TMS pulse and MEPs were excluded from further analysis if a mean absolute value of background muscle activity was greater than the mean plus two standard deviations for each MEP recording session. This resulted in approximately 1.5% of the data from hand group being excluded. The mean MEP amplitudes for each recording session were then compared using one-tailed paired-t tests as the direction of the MEP change was predicted to be a decrease (reduced excitability). The MEP data during post-rTMS 2 for one participant in the hand group was missing.

4.3.7 Analysis of speech data

F1 and F2 frequencies of each utterance were extracted from the data output of Audapter. We confirmed that the Audapter successfully applied F1 frequency increase without changing the fundamental frequency or F2 frequency. Then the average formant frequency was calculated from a window of 30-ms placed over the centre of each utterance. First and second formant values greater than three standard deviations from a participant's mean F1 and F2 values in each phase of the experiment were excluded from further analyses. This resulted in approximately 4.2% of the data being excluded. Four participants (three in the hand group and one in the tongue group) did not complete the unlearning phase due to time constraints.

No significant difference was found between hand and tongue groups in terms of the frequency of F1 or F2 productions during baseline (see Table 4-1). In each participant, the F1 and F2 frequency data for utterances produced during the experiment were then normalised to Z-scores, following the method we used in study 1 reported in Chapter 3 (i.e. subtracting the mean and dividing by the standard deviation of the formant frequencies produced during baseline with normal feedback). The normalized results for each utterance, averaged across the participants in each group, can be seen in Fig 4-3 (F1 changes) and Fig 4-4 (F2 changes).

4.3.8 Statistical Analysis

The normalized F1 and F2 frequencies (Z-score) during the whole learning phase were averaged for each participant (as for Chapter 3) and used in the following analyses. The average normalized F1 and F2 changes of the learning phase were compared against zero (the baseline average) in the hand and tongue group separately, using a one-sample t-test. The average normalized F1 and F2 changes of the learning phase were compared between hand and tongue groups, using independent t-tests. We also assessed the significance of the return-to-baseline changes in F1 and F2 frequencies in the unlearning phase. The average normalized F1 and F2 changes of the last 15 trials in the unlearning phase were compared against zero (the baseline average) in the hand and tongue group separately, using a one-sample t-test. Data normalised by subtracting the mean baseline frequencies from the frequencies in other phases of the experiment were also analysed in the same way. The pattern of results obtained using this

normalisation method was the same as the results obtained here (see Appendix 2 for details).

For each participant, we also calculated the Euclidean distance between the centre of the vowel clouds for “hid”, “bid”, “did” produced during the vowel exploration phase of the experiment and for “head”, “bed”, “dead” produced at the baseline and learning phase (Figure 4-1a). Learning-related changes in distance were then compared between groups (Tongue vs Hand) using a two-tailed t-test.

Table 4-1

Group details (Means ± SD)

	Hand (N=20)	Tongue (N=20)
Age (years)	22.7 ± 5.2	22.5 ± 5.0
Female/Male	10/10	10/10
Baseline F1 (Mel)	711 ± 86.7	692 ± 83.26
Baseline F2 (Mel)	1432 ± 81.0	1448 ± 66.6
Active Motor Threshold used for rTMS	48.8 ± 5.6	58.0 ± 5.7
Stimulator intensity used to elicit MEPs	53.6 ± 61	/

Stimulator intensity is given as a percentage of maximum stimulator output.

4.4 Results

4.4.1 Effect of rTMS on cortical excitability

Figure 4-2 presents the MEP changes before (pre-rTMS) and after (post-rTMS 1 and post-rTMS 2) disruptive rTMS, recorded from the hand in response to single-pulse TMS over the target representations in left M1. The mean amplitudes of hand MEPs obtained pre- rTMS and post-rTMS 1 were 2.71mV (±SE = 0.17) and 2.13 mV (±SE = 0.12) respectively. A paired t-test comparing the peak-to-peak amplitude of hand MEPs pre and post-rTMS revealed a significant effect of low frequency rTMS on cortical excitability (19% reduction in post-rTMS 1; $t_{19} = 3.55$, $p = .001$, $d = 0.88$). Moreover, the inhibitory effect of low frequency rTMS was maintained after the end of learning phase (post-rTMS 2) in the hand group (11% reduction in post-rTMS 2; $t_{18} = 2.56$, $p = .007$, $d = 0.50$).

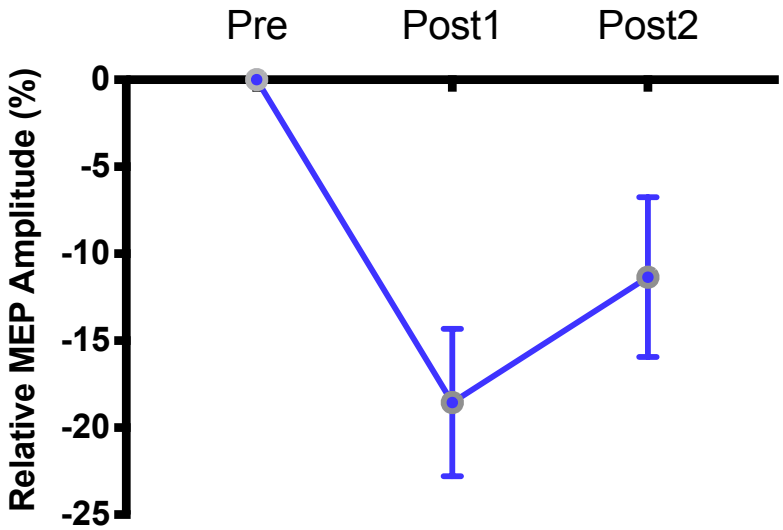


Figure 4-2 Effect of rTMS on cortical excitability.

The figure shows mean peak-to-peak amplitudes (± SE) of post-rTMS MEPs (post-rTMS 1 and post-rTMS 2) in relation to pre-rTMS MEPs. MEPs obtained from hand are shown in blue (N=20).

4.4.2 Effect of rTMS on speech adaptation

Figure 4-3a shows average F1 frequency produced for each utterance, normalized as a Z-score, for the whole adaptation experiment for hand and tongue groups. During the learning phase, a compensatory decrease in F1 production was clearly observed in the hand group following the introduction of F1 upwards shift, while such compensatory change was disrupted by rTMS applied over the tongue representation, and this effect persisted for the remaining course of learning phase. An independent t-test was conducted to compare the learning-related changes across the whole learning phase between groups and showed there was significant difference between hand and tongue group (Figure 4-3b, t_{38} = 2.65, p= .012, d = 0.84). Critically, rTMS over the tongue representation before the learning phase, resulted in no F1 change from baseline productions, which means the subthreshold rTMS impaired speech motor learning without affecting movement execution (one-sample t-test against baseline average, Z = 0; t_{19} = 1.38, p= .18). At the end of the unlearning phase, F1 frequency of both groups had returned to baseline (Figure 4-3b, one-sample t-test against baseline average, Z = 0; tongue group: t_{18} = 0.18, p= .856; hand group: t_{16} = 0.35, p= .730).

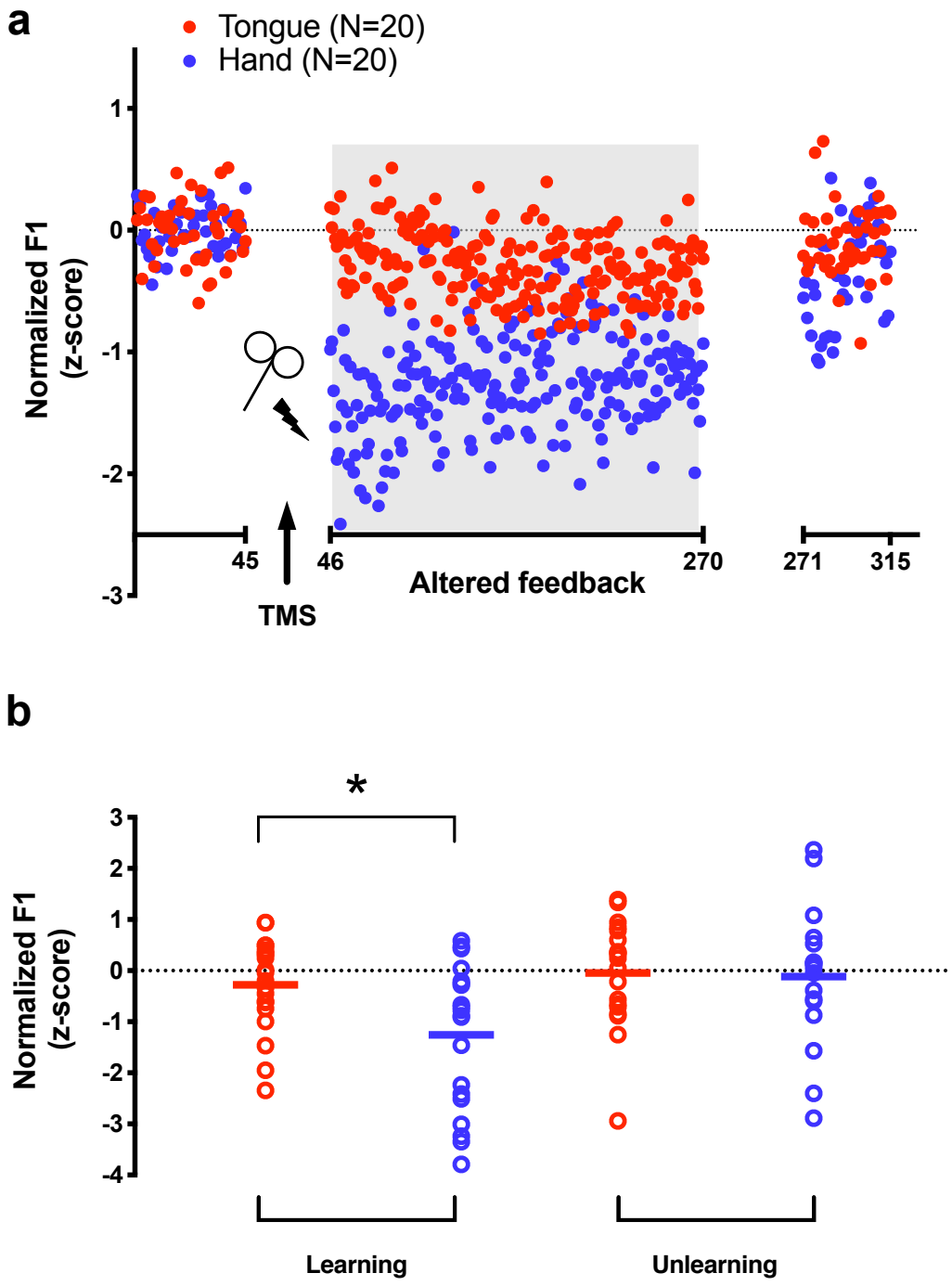


Figure 4-3 Changes in normalized F1 production (Z-score) related to sensorimotor learning for the hand and tongue groups.

(a) Changes in normalized F1 frequency during baseline, learning, and unlearning phases of the experiment. Solid dots indicate the group average for each utterance. The shaded area represents the trials where the altered feedback was applied. The dotted line at zero is the baseline average. (b) Mean F1 changes (normalized) for the whole learning phase (utterances 46-270) and

*the end of the unlearning phase (utterances 301-315). Open circles represent individual participants, the solid lines represent the group means Values for hand group are shown in blue and values for tongue group are shown in red. Asterisks (*P < 0.05) indicate reliable between group differences.*

Figure 4-4a shows average F2 frequency produced for each utterance, normalized as a Z-score, for the whole experiment for hand and tongue groups. In response to F1 shift up, the hand group showed expected response in F2 by increasing the frequency of their F2 productions. This response was expected based on previous work using this manipulation (Lametti, Smith, Freidin, et al., 2018; MacDonald et al., 2011; Rochet-Capellan & Ostry, 2011; also see Chapter 3). In contrast, the tongue group exhibited a slight tendency to shift F2 downwards, although F2 changes in tongue group did not significantly differ from zero (baseline production, $t_{19} = 1.76$, p= .095). Again, we compared F2 changes across the whole learning phase between hand and tongue group. A significant difference between these two groups was found in F2 learning (Figure 4-4b, $t_{38} = 3.35$, p = .002, d = 1.06). These results confirm that as above, the rTMS over the tongue representation interferes with adaptive responses but does not impair or alter speech production more generally.

During the end of unlearning phase, F1 frequencies did not differ from F1 frequencies produced during the baseline in both groups (Figure 4-4b, one-sample t-test against baseline average, Z = 0; tongue group: $t_{18} = 0.86$, p = .185; hand group: $t_{16} = 0.86$, p = .404).

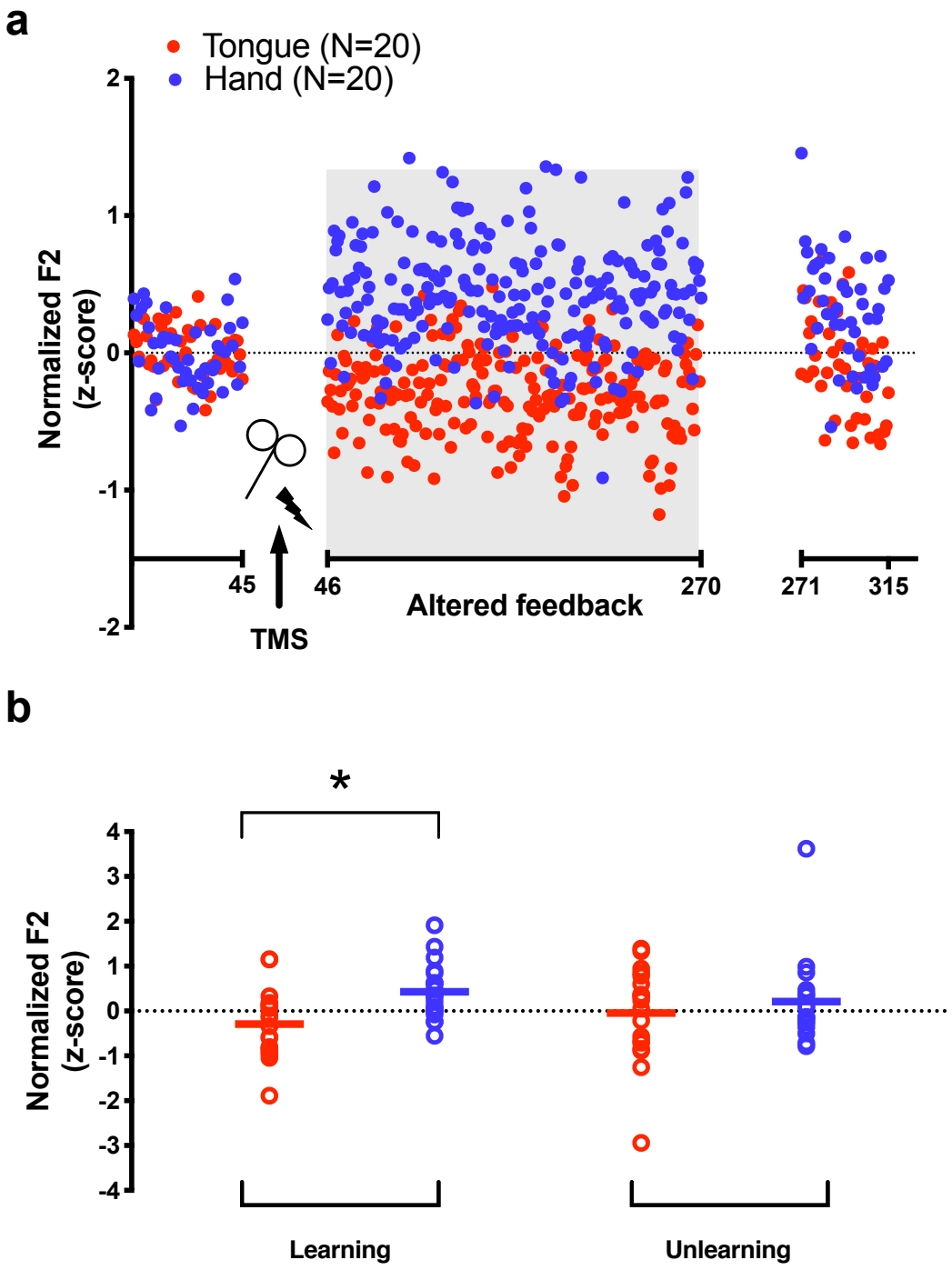


Figure 4-4 Changes in normalized F2 production (Z-score) related to sensorimotor learning for hand and tongue groups.

(a) Changes in normalized F2 frequency during baseline, learning, and unlearning phases of the experiment. (b) Mean F2 changes (normalized) for the learning phase (utterances 46-270) and the end of the unlearning phase (utterances 301-315). See legend to Fig. 4-3 for details.

In sum, rTMS over the tongue representation in primary motor cortex severely disrupted the expected compensatory responses in production of F1 and F2 in response to an F1 feedback perturbation; rTMS over the hand representation had no effect.

Combined changes in F1 and F2, specifically a decrease in F1 coupled with an increase in F2, place utterances closer to values used to produce another vowel sound, namely the /ɪ/ vowel as in “hid”, “bid”, “did” (Figure 4-1a). This pattern of compensation was observed previously using a similar paradigm in participants who received sham brain stimulation (Lametti, Smith, Freidin, et al., 2018) and by us using the same paradigm in a separate study without stimulation (see Chapter 3). To evaluate how closely the new productions during learning approximated another vowel in the vowel space of individual participants, we calculated the Euclidean distance between the centres of the vowel clouds for “hid”, “bid”, “did” produced during the vowel exploration phase of the experiment with those for “head”, “bed”, “dead” produced at baseline and during the learning phase. In response to the F1 shift-up, the productions of “head”, “bed”, “dead” (across the whole learning phase) in the hand group were 32.17 mel closer to the productions of “hid”, “bid”, “did” in F1-F2 space (t_{19} = 4.32, $p < .001$). In contrast, for the tongue group, the change in distance (4.27 mel, t_{19} = 0.60, $p = .554$) was not significantly different from zero (no change). The difference between the hand and the tongue groups in the magnitude of change in distance was significant (t_{38} = 2.71, $p = .001$, $d = 0.86$; Fig 4-5a). Figure 4-5b shows the vectors for individual participants indicating the direction and magnitude of compensatory changes in response to the F1 upwards shift (downwards dotted

arrow). Participants in the hand group showed the expected combined reduction in F1 and increase in F2 frequencies in response to the altered feedback shifting in the expected direction towards the /ɪ/ vowel (blue arrows in Fig 4-5b). In contrast the compensatory changes in the tongue group were abolished by rTMS resulting in (mostly) small and inconsistently orientated vectors for individual participants (red arrows in Fig 4-5b).

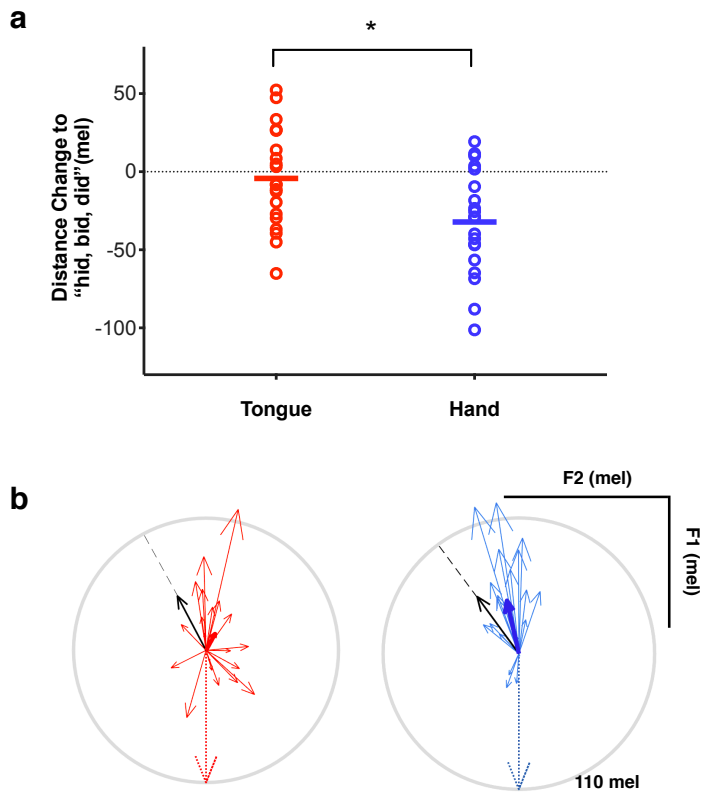


Figure 4-5 Coupled F1-F2 changes in vowel space for production during F1 feedback perturbation.

(a) Learning-related reduction in Euclidean distance between "head", "bed", "dead" and "hid", "bid", "did" in relation to sensorimotor learning and rTMS site. Negative values indicate that production of "head", "bed", "dead" has moved closer (in vowel space, see Fig. 4-1) to "hid" "bid" "did" in response to F1 feedback

*perturbation. Open circles represent data from individual participants, the filled squares (± SE) represent the group means. Asterisks (*P < 0.05) indicate reliable between-group differences. (b) Vector figures of tongue (red) and hand (blue) stimulation groups. The vectors represent the direction and magnitude of the compensatory change from baseline in vowel space. The thin vectors represent data from individual participants and the thick vectors represent the group average. The black arrow represents the average of the direction required to move from "head" "bed" "dead" to "hid", "bid", "did", while the dotted arrow represents the formant frequency shift we applied. The light grey circle indicates 110 mel (i.e. the magnitude of the shift applied).*

4.5 Discussion

In the current study, we asked whether disruption of the articulatory representation in left M1 by repetitive TMS would interfere with speech adaptive learning. We found that disruption of the tongue representation in M1 almost abolished the expected compensatory response for the upwards F1 shift in auditory feedback, namely a decrease in F1 frequency alongside an increase in F2 frequency. Participants who received rTMS over the hand representation showed this expected response. For the participants in the hand group, combined changes in F1 and F2 served to move their productions towards the vowel space representation of the /ɪ/ vowel as previously observed in studies using this paradigm (Lametti, Smith, Freidin, et al., 2018; study 1 in chapter 3). The combined F1 and F2 changes made by the tongue group were smaller in magnitude and random in terms of their direction within the vowel space. This is the first study to address the key question as to whether cortical motor areas that control movements of articulators causally contribute to speech motor adaptation.

Our results showed that disruption of articulatory representations in left M1 disrupted compensatory responses in the F1 induced by F1 feedback perturbation. This is consistent with findings from two previous studies that have recently examined the impact of ventral motor cortex or cerebellar tDCS on speech adaptation (Lametti, Smith, Freidin, et al., 2018; Scott et al., 2020). In those studies, anodal tDCS stimulation over the ventral motor cortex, which is assumed to increase cortical excitability, enhanced compensation in F1 in response to F1 feedback perturbation. Taken together, the results from these three studies confirm an important role for left speech motor cortex in speech adaptation. The results of the current study extend this previous work by using an interference method, which we suggest is a better test of the causal contribution of the motor cortex to speech adaptation. Furthermore, the TMS used in this study has greater focality than the tDCS study using large electrodes (Lametti, Smith, Freidin, et al., 2018) thereby confirming that disruption of tongue representation in the left M1 specifically impaired the expected compensatory response in F1, while disruption of the nearby hand representation (unrelated control area) had no affect on the same adaptive process.

We observed that compensatory changes in F2 were also affected by rTMS applied over the tongue but not the hand representation. This is in line with the previous study in which anodal tDCS over ventral motor cortex enhanced compensatory changes in both F1 and F2 in response to a single formant perturbation (Lametti, Smith, Freidin, et al., 2018). In the current study, these coupled F1-F2 changes produced by participants in the hand group moved new productions closer to values used to produce another vowel sound (i.e. the /ɪ/

vowel as in “hid”, “bid”, “did”; Fig 4-1a), a pattern of compensation that was again observed in previous studies using a similar paradigm (Lametti, Smith, Freidin, et al., 2018; also see chapter 3). We suggest, therefore, that the articulatory representation in left M1 is involved in generating a form of adaptation that aims to move speech production closer to a learnt pattern of speech production used to produce another vowel. It is worth noting, however, that compensatory changes in F2 were not seen in another study under either sham or anodal motor stimulation (Scott et al., 2020). It is unclear whether this difference reflects small differences between the paradigms, for example the size of the feedback shift or the duration of the learning phase or the large variability among participants. Unlike in studies of visuomotor adaptation, where nearly every participant exhibits compensatory responses after experiencing feedback error (Krakauer et al., 2000), a fairly large proportion of participants in auditory studies shows little or no compensation for auditory perturbations (10-40%; Lametti, Nasir, & Ostry, 2012; Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014; Rochet-Capellan, Richer, & Ostry, 2012). These relatively large inter-individual differences suggest that larger sample sizes might be necessary to observe compensatory responses induced by single formant perturbation, especially for changes in the unperturbed formants, which are generally smaller.

In contrast to the speech motor adaptation literature, the contribution of the motor cortex to visuomotor adaptation has been extensively studied using non-invasive brain stimulation. Findings are inconsistent, however. Several studies corroborate the notion that M1 is uninvolved in initial motor adaptation but is important for the retention of learned movement patterns after feedback

perturbations are removed (Galea et al., 2011; Hadipour-Niktarash et al., 2007; Richardson et al., 2006). For example, disruption of M1 by repetitive TMS before learning left adaptation to a force field unaffected, but significantly impaired participants' performances in the same force field 24h later (Richardson et al., 2006). Similarly, anodal tDCS, which increases neuronal excitability, over M1 had no effect on visuomotor adaptation, but increased the persistence of adapted movements when perturbations were removed (i.e. longer after-effects; Galea et al., 2011). Conversely, anodal tDCS delivered to M1 during force-field adaptation significantly increased the initial movement trajectory error (in the first 150 ms of movement), which suggests that M1 is involved in the adaptation of reaching movements of human participants (Hunter et al., 2009). A later study revealed that M1 may contribute to reaching adaptation in different phases; disruptive TMS applied to M1 had no effect on the rapid adaptation phase (approximately the first 20 trials when the perturbation was abruptly applied), but reduced adaptation at the plateau later in the learning phase (Orban de Xivry, Criscimagna-Hemminger, & Shadmehr, 2011). The findings of the latter study concord with those of studies with non-human primates, which found learning-related changes in the M1 during the late phase of adaptive learning (Paz, 2005; Paz, Boraud, Natan, Bergman, & Vaadia, 2003).

Here, we found that the articulatory representation in left M1 contributed to speech motor adaptation even in the early stages of learning. In what ways might the motor cortex contribute to speech adaptation? One possibility is that the left articulatory motor cortex is necessary for the motor adjustment to the induced error. Current models of speech production, such as the Directions into Velocities

of Articulators (DIVA) model have attempted to explain the neural network involved sensorimotor learning in the speech domain (Tourville & Guenther, 2011). According to this model, production of a phoneme or syllable starts with activation of the speech sound map which is assumed to lie in the left ventral premotor cortex (vPMC) and serves to store the motor commands for common speech sounds. Activation of the speech sound map results in predictions being formed of the sensory consequences of the given motor commands. When a mismatch between the predicted and actual sensory signals occurs, an error signal is generated in the auditory and somatosensory error maps. In the updated version of the DIVA model, a feedback control map was added, which is assumed to transform the auditory and somatosensory error signals into corrective motor commands, that are later used to modify the forward model (Tourville and Guenther, 2011). Neuroimaging studies suggest that such feedback control maps are located in the right frontal or ventral premotor cortex (Golfinopoulos et al., 2011; Tourville et al., 2008). Our results suggest that the left motor cortex is also involved in transforming the auditory error signals into corrective motor commands.

The parietal cortex is another brain area considered an integral component for the formation and modification of forward and inverse internal models of motor control based on the limb movement literature (Della-Maggiore, 2004; Wolpert & Kawato, 1998). Accordingly, one previous study used repetitive TMS to demonstrate the contribution of inferior parietal cortex to sensorimotor integration for speech motor learning also (Shum et al., 2011). Participants who received rTMS over parietal cortex exhibited a diminished adaptive response, while the

adaptive response of participants who received sham stimulation was unimpaired. It is suggested that the left inferior parietal cortex is also involved in the process of transferring the speech error signal into motor corrective commands, therefore, especially when sensory perturbation is applied.

It is possible that TMS over the tongue representation disrupted not only the target region, but also adjacent areas, including premotor and prefrontal cortex. To control for this, and for more general effects associated with TMS, we also stimulated the nearby hand representation in M1 in a separate group of participants; this group maintained the expected compensatory responses to the auditory perturbation. Even so, although the effects of TMS are focal and maximal at the stimulated area, regions located more distally that are functionally connected to the stimulated site can be affected (Paus et al., 1997). Our results indicate that the primary motor representation is a key part of such a functional network. Additional work combining TMS with measures of functional connectivity would shed light on the extent of this network but based on the literature reviewed thus far, it would include inferior, ventral premotor cortex bilaterally, and the cerebellum. Interference at different nodes in this network might produce different types of impairment in speech motor adaptation, which should be the focus of further study.

The effectiveness of repetitive low-frequency TMS in creating a “virtual lesion” in the area directly under the coil was demonstrated in several previous studies (Chen et al., 1997; Mottonen & Watkins, 2009). The MEP size recorded from hand or lip muscles in response to single-pulse TMS over target primary motor

representation is significantly reduced after a 15-min train of low-frequency subthreshold rTMS (Mottonen & Watkins, 2009). Our results confirmed the effectiveness of this form of rTMS on temporal disruption of hand motor excitability. We were unable to demonstrate a similar pattern of suppression of motor excitability for the tongue representation concurrent with task performance as placing electrodes on the tongue would have affected speech production. We therefore estimated the position of the tongue representation by localising the hand and lip representations in individual participants. As a proof-of-principle that this estimation was successful and that a 12-min train of subthreshold low frequency rTMS suppresses cortical excitability of tongue representation, we collected MEP data in a small group of five participants (see Chapter 2). The results confirmed that the estimation of the location was accurate as MEPs were elicited from the tongue and also the rTMS protocol successfully reduced MEPs size as observed for the hand group.

It should be noted that our results have limitations, which could be addressed in future work. First, we could not rule out the possibility that disruption of the tongue representation in M1 impaired sensorimotor learning by impairing speech perception. Growing evidence (including our own) demonstrates a role for M1 in speech perception (see Liebenthal & Möttönen, 2018 for review). Recent TMS studies showed disruption of motor cortex impaired participants' ability to perform a phonetic discrimination task and suppressed their brain responses to changes in speech sounds (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Möttönen, Dutton, & Watkins, 2013; Smalle, Rogers, & Möttönen, 2015). In the current study, it is possible that rTMS over the articulatory representation impaired participants'

perceptual acuity, which is positively related to compensatory adaptation (Villacorta et al., 2007; Villacorta, Perkell, & Guenther, 2005); perhaps they did not perceive the effect of the perturbation in the feedback they received. Current models of speech production predict that people with better auditory discrimination ability have more distinguishable auditory goals and greater adaptation to auditory perturbations (Houde & Nagarajan, 2011; Perkell, 2012; Tourville & Guenther, 2011). However, if perception of feedback was impaired by the perturbation, we might expect that production would not have been maintained as at baseline but that it would have started to deviate or that more random production was produced. Further work exploring the effect of rTMS-induced disruption over the tongue representation on auditory acuity and how it relates to adaptation could clarify this point.

In the current study, we observed an unusually rapid adaptation in the hand group at the beginning of learning phase (see Fig 4-3). Typically, adaptation happens quickly over the first 20 shifted utterances (Lametti, Smith, Freidin, et al., 2018; MacDonald et al., 2011). Then, participants continually adjust their productions at a slower rate. The very rapid learning observed in the current study could be interpreted as a baseline shift induced by brain stimulation. However, further analysis indicated this is unlikely. Firstly, the F1 frequencies of production when the feedback shift was applied were the same as the F1 frequencies for the baseline for the first three trials in participants in the hand group. Secondly, at the end of the unlearning phase when the auditory feedback perturbation was removed, the frequency of F1 productions by the hand group were once again the same as at baseline (see Fig 4-3). These results suggest rTMS over the hand

area did not induce a baseline shift in speech productions. Nevertheless, our protocol appears to have led to more rapid adaptation at the beginning of learning. One possible explanation of this effect is that TMS has a general arousing or alerting effect leading to faster adaptation in the group receiving stimulation over the hand area, and this effect was abolished by stimulation of the tongue area. A more likely explanation of the rapid adaptation involves the time interval between baseline and learning phase in this experiment, which did not occur in previous studies (Lametti, Smith, Freidin, et al., 2018; MacDonald et al., 2011; Rochet-Capellan & Ostry, 2011; Tang, Lametti, & Watkins, 2019). In one study, a similar length interval between baseline and learning was used to allow completion of a speech perception task (Lametti et al., 2014); participants showed a similarly rapid adaptation to the feedback perturbation. In the current study, the interval between baseline and learning was about 25 mins due to TMS localization, thresholding and the 12-min rTMS train during which they were silent. It is possible that the lack of recent normal feedback 'evidence' confirming that the sensorimotor system is functioning accurately resulted in a more rapid adaption to the feedback perturbation.

In summary, we investigated whether the articulatory representation in M1 causally contributes to speech motor adaption. We found that inhibitory TMS over the tongue, but not the hand representation, significantly impaired compensatory changes in both the shifted formant (F1) and the unshifted formant (F2). We posit that the articulatory representation in left M1 is critical to the network supporting adaptation.

Chapter 5 Asymmetry of auditory-motor speech processing is determined by language experience

5.1 Abstract

Speech processing relies on interactions between auditory and motor systems and is asymmetrically organized in the human brain. The left auditory system is specialized for processing of phonemes, whereas the right is specialized for processing of pitch changes in speech affecting prosody. In speakers of tonal languages, however, processing of pitch (i.e., tone) changes that alter word meaning is left-lateralized indicating that linguistic function and language experience shape speech processing asymmetries. Here, we investigated the asymmetry of motor contributions to auditory speech processing in male and female speakers of tonal and non-tonal languages. We temporarily disrupted the right or left speech motor cortex using transcranial magnetic stimulation and measured the impact of these disruptions on auditory discrimination (mismatch negativity) responses to phoneme and tone changes in sequences of syllables using electroencephalography. We found that the effect of motor disruptions on processing of tone changes differed between language groups: disruption of the right speech motor cortex suppressed responses to tone changes in non-tonal language speakers, whereas disruption of the left speech motor cortex suppressed responses to tone changes in tonal language speakers. In non-tonal language speakers, the effects of disruption of left speech motor cortex on responses to tone changes were inconclusive. For phoneme changes, disruption of left but not right speech motor cortex suppressed responses in both language

groups. We conclude that the contributions of the right and left speech motor cortex to auditory speech processing are determined by the functional roles of acoustic cues in the listener's native language.

5.2 Introduction

Over the past two decades, a growing number of studies have emphasized the importance of interactions between auditory and motor systems for speech perception (see Liebenthal & Möttönen, 2018; Skipper, Devlin, & Lametti, 2017). The dorsal pathway connecting auditory and motor systems is strongly left-lateralized for phonemic processing (Hickok & Poeppel, 2007). In contrast, auditory-motor processing of prosody is lateralized to the right hemisphere (Ethofer et al., 2006; Sammler, Grosbras, Anwander, Bestelmeyer, & Belin, 2015; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006).

Asymmetry in auditory processing is thought to reflect differences in the temporal and spectral sensitivities of the left and right auditory cortex to acoustic inputs (e.g. Zatorre and Belin, 2001). An extensive body of work indicates that the left auditory cortex is specialized for processing rapidly changing temporal information, which is necessary to resolve phonemic changes, i.e. to discriminate small acoustic differences in speech sounds that alter word meaning. In contrast, the right auditory cortex is sensitive to more slowly changing auditory information and the spectral content of the same acoustic signals, which is necessary to discriminate prosodic changes in speech, such as the use of pitch for stress, or to convey emotion (Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002). Although this view has gained lots of support, it should be noted that it has been

primarily tested in speakers of non-tonal languages (Albouy et al., 2020; Boemio et al., 2005; Flinker, Doyle, Mehta, Devinsky, & Poeppel, 2019; Jamison, Watkins, Bishop, & Matthews, 2006).

For as many as 70% of the world's languages, supra-segmental speech elements such as changes in pitch (i.e. tones) are also used to alter word meaning (Yip, 2002). For example, in Mandarin Chinese, the word "ma" produced with a high-level tone (ma1) means "mother" and when produced with a rising-falling tone (ma3) means "horse". Auditory processing of these lexical tones in tonal language speakers is also left lateralized (Gandour et al., 2000; Hsieh, Gandour, Wong, & Hutchins, 2001; Klein et al., 2001) even though it does not require rapid temporal processing of the acoustic signal. In this case, the left lateralization reflects the linguistic function of the acoustic signals and thereby demonstrates how language experience can alter hemispheric lateralization of speech processing.

It is likely therefore that asymmetry in speech processing is determined by bottom-up properties in the acoustic signal as well as by top-down influences (Albouy et al., 2020; Zatorre & Gandour, 2008). We hypothesised that one such influence could come from the speech motor system. We have previously shown that TMS-induced disruption in the articulatory motor cortex affects behaviour, i.e., discrimination and identification of speech sounds (Mottonen & Watkins 2009, Smalle et al., 2015; Rogers et al., 2014). Behavioural tasks may, however, engage mechanisms that are not normally used in speech perception and such tasks require using special stimulus material (e.g., acoustic continua between

phonemes). Auditory discrimination can be examined without behavioral tasks by using EEG to record responses elicited by infrequent changes in sound sequences (Näätänen et al. 2001). These mismatch negativity (MMN) responses are automatic and preattentive, which is generated primarily in the auditory cortex (see chapter 2 for detail). Previous studies have shown that the behavioural ability to discriminate speech sounds is related to the MMN amplitude. For example, phonemic contrasts of the native language that are easily detectable elicit large MMNs, whereas phonemic contrasts that do not exist in the native language are hard to detect and elicit smaller MMNs (Naatanen et al., 1997). By combining rTMS and EEG, Möttönen and colleagues (2013) successfully demonstrated the automatic involvement of left speech motor cortex in phoneme perception. They found that disruption of the lip, but not hand representation in left M1 significantly suppressed MMN responses to changes in sequences of speech sounds (i.e. sequences of phonemes), and did not affect those elicited by changes in sequences of non-speech sounds (i.e. piano tones). However, less is known about the contribution of the right speech motor cortex to auditory speech processing (both phoneme and lexical tone) and the between-hemisphere asymmetry of such motor contributions.

Here, we aimed to determine whether speech processing in the auditory system is affected by lateralized interactions with the motor system. We investigated the asymmetry of auditory-motor speech processing in both tonal and non-tonal language groups capitalizing on the fact that the same low-level acoustic signal (e.g. pitch) serves different linguistic functions in each.

We used low-frequency repetitive TMS to temporarily disrupt the speech motor cortex and EEG to record MMN responses to changes in a sequence of frequent “da” syllables (da1) that were either a phoneme (ba1) or a tone change (da4). If hemispheric differences in motor contributions to auditory processing relate to low-level acoustics, then disruption of the right speech motor cortex should modulate tone processing and disruption of the left speech motor cortex should modulate phoneme processing in both language groups equally. In contrast, if these differences arise due to the functional roles of the acoustic cues, which differ between tonal and non-tonal languages for tone processing, then the effect of motor disruptions on processing of tone changes would differ between tonal and non-tonal language groups.

5.3 Materials and Methods

5.3.1 Participants

We recruited 32 adult human participants (16 male, 16 female) who were speakers of non-tonal languages and aged between 18 and 35 years and 32 tonal language speakers (16 male, 16 female) aged between 18 and 34 years. None of the 64 participants had received any formal musical training that had lasted longer than 2 years. The native languages of the group of non-tonal language speakers were: English (21), French (2), Spanish (2), Czech (2), German (1), Italian (1), Romanian (1), Polish (1), and Dutch (1). None of the non-tonal language speakers had ever learned a tonal language but most had learned another language to varying levels of proficiency. All the tonal language speakers were native speakers of Mandarin and had at least a basic level of English; 29

were studying at an English-speaking university and all were living in the UK. The ages at which they first started to learn English ranged from 7 to 15 years old. Male and female participants in each language group were randomly assigned to receive either left or right hemisphere stimulation, so that groups were gender balanced. All participants were right-handed and reported no hearing problems and no personal or family history of seizures or other neurological disorders. The Central University (of Oxford) Research Ethics Committee approved the experimental protocol and participants gave informed consent. EEG Data from two non-tonal language speakers (both native English speakers; 1 male, 1 female) were excluded from the analyses because their mean mismatch negativity (MMN) amplitudes were extreme (greater than three standard deviations from the group mean MMN amplitude). The mean ages and gender balance for the four groups that contributed data to the analysis are as follows: non-tonal language, left hemisphere stimulation group: mean age 21.60 years, 8 males, 7 females; non-tonal language, right hemisphere stimulation group: mean age 22.60 years, 7 males, 8 females; tonal language, left hemisphere stimulation group: mean age 25.69 years, 8 males, 8 females ; tonal language, right hemisphere stimulation group: mean age 26.13 years, 8 males, 8 females. There was no significant difference among groups in terms of age.

5.3.2 Procedure

For all participants, EEG during a 16.5-min oddball sequence was recorded during two sessions: a no-TMS baseline and a post-TMS session. The oddball sound sequence comprised infrequent tone and phoneme changes, which were expected to elicit MMN responses. During no-TMS baseline sessions, event-

related potentials (ERPs) were also recorded to two control sequences (see below). The order of presentation of oddball and control sequences was counterbalanced. Participants watched a silent movie and were instructed to ignore the sound sequences during all recordings. The order of no-TMS and post-TMS sessions was counterbalanced: half participants started with the no-TMS session followed by repetitive TMS (rTMS) and then an immediate post-TMS session, while the other half received rTMS first, the post-TMS session immediately after, and then a no-TMS session at least 50 minutes later. TMS involved 15-min of low-frequency (0.6 Hz) sub-threshold repetitive stimulation over the lip representation (henceforth, speech motor cortex) in either the left or the right primary motor cortex. This protocol temporarily inhibits excitability in the target representation for a further 15 mins approximately (Mottonen & Watkins, 2009).

5.3.3 Stimuli

The oddball sequence contained a total of 1800 stimuli, including infrequent ‘ba1’ and ‘da4’ (probability = 0.1 for each) and frequent ‘da1’ syllables (probability = 0.8). Control sequences consisted of 400 repetitions each of the infrequent syllables in the oddball sequence: two control sequences, one for ‘ba1’ and one for ‘da4’. Syllables ‘da1’ (tone 1: high-level tone) and ‘da4’ (tone 4: high falling tone) shared the same phoneme but carried different tone contours, while ‘da1’ and ‘ba1’ had the same high-level tone (tone 1) but contained different phoneme information. The original stimuli (‘da1’, ‘da4’,’ ba1’) were recorded from a female native Mandarin speaker. Using Praat software (Institute of Phonetic Sciences, University of Amsterdam, The Netherlands), we digitally normalized them to

equal durations (150 ms) and matched them for fundamental frequency (F0) for the first 35 ms of the syllable. The 'da4' syllable differed from the other two in the F0 trajectory after 35ms, which is clearly seen in Figure 5-1. The inter-stimulus interval was 400 ms in all sequences. Stimulus presentation was controlled via custom scripts in MATLAB (MathWorks). Participants wore TMS-compatible insert earphones (Etymotic, Elk Grove Village, IL, USA); stimuli were played at a comfortable listening level ~70 dB.

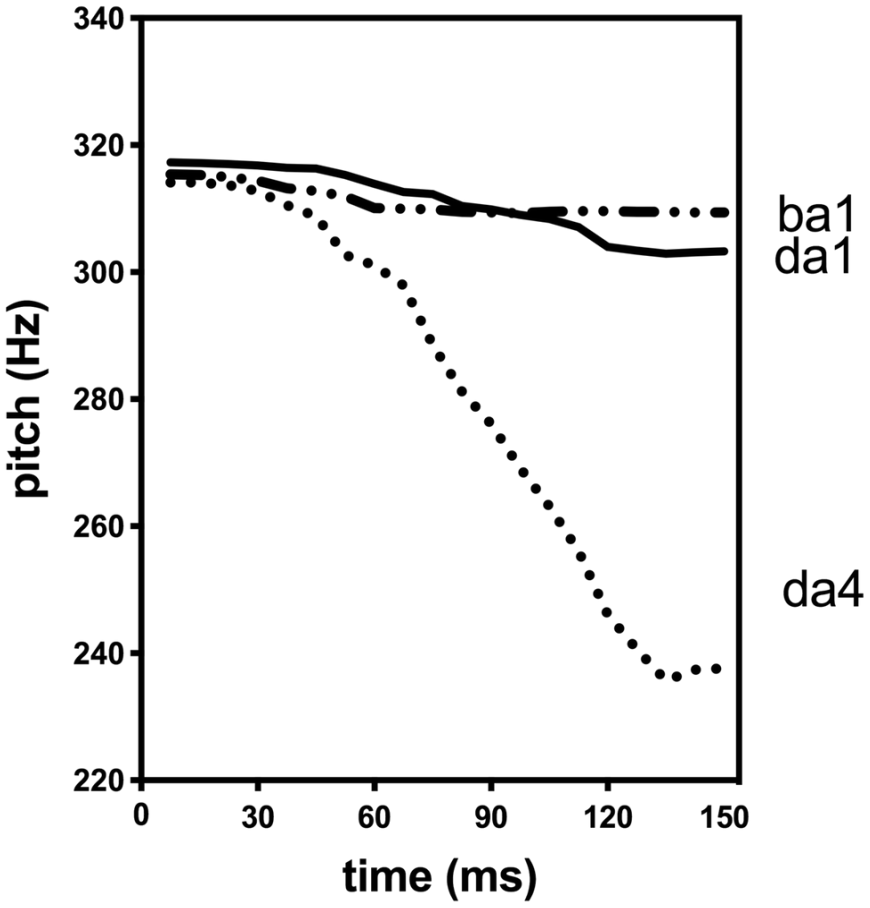


Figure 5-1 The trajectories of fundamental frequency (F0) of the three syllables used in current study.

Ba1 is a "ba" syllable with tone 1, which is a high-level tone; da1 is a "da" syllable also with tone 1; da4 is the same "da" syllable with tone 4, which is a falling tone.

5.3.4 Transcranial Magnetic Stimulation

All TMS pulses were monophasic, generated by two Magstim 200s and delivered through a 70-mm figure-8 coil connected through a BiStim module (Magstim, Dyfed, UK).

Before rTMS, single TMS pulses were used to localize the lip representation, following the procedure previously described (Möttönen et al., 2014). The coil was placed tangential to the skull to induce a current flow from posterior to anterior under the junction of the two wings of the figure-eight coil. The position and orientation of the coil over the lateral scalp (starting with the handle held horizontally) was adjusted until a robust motor-evoked potential (MEP) was observed in the contralateral target muscle. Then, once the target area had been located, we established the active motor threshold, that is, the minimum intensity at which TMS elicited at least 5 out of 10 motor-evoked potentials (MEPs) with an amplitude of at least 200 μV when the target muscle was contracted at 20–30% of the maximum. The mean active motor threshold (percentage of maximum stimulator output, ±SD) for the left and right stimulation groups respectively was 63.1% (±7.8) and 62.1% (±7.1) in the tonal language speakers, and 58.8% (±6.1) and 57.47% (±7.4) in the non-tonal speakers (no significant differences among the four groups; $F_{3,58} = 2.71$, $p = .101$). The intensity of each participant's active motor threshold was used for 15-min rTMS while the muscles were relaxed (hence the stimulation was subthreshold). In total, each participant received 600 pulses at 0.6Hz (rTMS), and between 30 and 60 single pulses for thresholding and localizing. During rTMS, participants watched a silent nature documentary and were instructed to stay still. Electromyography (EMG) recordings were

carefully monitored throughout stimulation to ensure that muscles were relaxed, and no MEPs were evoked in the lip muscles. The coil was changed halfway through rTMS to prevent overheating.

5.3.5 Electromyography (EMG) recordings

Disposable electrodes were attached on the right or left orbicularis oris muscle (the upper and lower lip contralateral to the stimulation site) and forehead (ground) to record electromyography (EMG) signals. The EMG signals were amplified, sampled at 5000 Hz, filtered (1 Hz–1 kHz bandpass) using a CED 1902 amplifier, a CED 1401 analog-to-digital converter, and a PC running Spike2 (Cambridge Electronic Design). A power bar displayed on a computer screen allowed participants to practise producing a constant level of contraction of the lips (20–30% of the maximum).

5.3.5.1 Electroencephalographic (EEG) recording parameters

We acquired EEG using Synamps amplifiers and Curry 7 data acquisition software. We used a custom electrode cap with 11 electrodes (Fz, FCz, Cz, CPz, Pz, F1, F2, P1, P2, left and right mastoids), which allowed us to position the coil over sites of interest on the lateral convexity of the scalp. The ground and reference electrodes were placed on the right upper arm and the tip of the nose, respectively. During acquisition, data were low-pass filtered (400 Hz cutoff), digitized at 1000 Hz, and stored for offline analysis. Electrode impedances were reduced below 10 kΩ before recording.

5.3.5.2 EEG data Analyses

EEG signals were preprocessed using the EEGLab Toolbox (Delorme & Makeig, 2004). The raw data (down sampled to 500 Hz) were first re-referenced to the mean of the two mastoids to improve the signal-to-noise ratio of the MMN responses. Then the data were digitally filtered (low-pass filter of 30 Hz, high-pass filter of 1 Hz), baseline corrected, and segmented into epochs of 400 ms comprising a 100-ms pre-stimulus baseline and a 300-ms interval after the onset of the stimuli. The epochs containing amplitude fluctuations exceeding ± 70 μV were considered as artifacts and removed before averaging. In addition, the epochs for the first 10 stimuli in each sequence and for the first standard stimulus after each deviant were removed.

Both tone and phoneme changes were expected to elicit MMNs, a fronto-central negative component with sources primarily in the left and right auditory cortex, which reflects the automatic detection of a deviant stimulus in a passive auditory odd-ball paradigm (Näätänen et al., 1997; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Winkler et al., 1999). The amplitude of the MMN provides an index of an individual's ability to detect the change in the stimulus sequence, which requires discrimination between deviant and standard stimuli (Näätänen, 2008). For example, phonemic changes of the native language that are easily detectable elicit large MMNs, whereas phonemic changes that do not exist in the native language are hard to detect and elicit smaller MMNs (Näätänen et al., 1997).

For the baseline (no TMS) sessions, MMN responses were calculated in two ways: the traditional method (subtracting the ERPs evoked by the frequent stimuli from those evoked by the infrequent stimuli in the oddball sequence), and the same-stimulus method (subtracting ERPs evoked by the stimuli presented frequently in the control sequence from those evoked by identical sounds presented infrequently in the oddball sequences). The advantage of the traditional method is that the responses are recorded for frequent and infrequent stimuli during the same sequence. The disadvantage is that we are comparing responses to sounds that differ acoustically. The same-stimulus method gets around this potential confound by comparing the response to acoustically identical stimuli presented either in the control sequence (i.e. with other identical stimuli) or in the oddball sequence as an infrequent sound (i.e. with other acoustically different stimuli). Identity MMN responses obtained using the same-stimulus method therefore reflect differences due to the automatic discrimination of speech sounds presented infrequently in the context of more frequent but acoustically different speech sounds (Jacobsen & Schröger, 2003; Jacobsen & Schröger 2001; Kujala, Tervaniemi, & Schröger, 2007). Identity MMN responses were calculated for the baseline sessions to confirm that the responses were due to discrimination of speech sounds (i.e. that they were genuine). For the post-TMS sessions, MMN responses were calculated using the traditional method only as there was insufficient time to obtain control sequences during the TMS-induced disruption, which is thought to last less than 20 minutes.

5.3.6 Statistical analyses

Statistical analyses were pre-planned and followed the methods used in Möttönen et al., (2013), thereby allowing comparison between studies. MMN responses calculated using both the traditional or the same-stimulus (identity MMNs) method were maximal at the FCz electrode, as seen previously (Möttönen et al., 2013; Näätänen et al., 1997, 2001). MMN responses at FCz were selected for statistical analyses therefore; data for all electrodes are available at https://osf.io/en7uq/.

To evaluate whether the TMS-induced disruption of the motor cortex modulated MMN responses, we compared MMN responses (calculated using the traditional method) in the no-TMS and post-TMS conditions. Paired t-tests (2-tailed) were calculated at each time point from 0 to 300 ms after the onset of the syllable to determine whether any portion of the MMN responses differed significantly between conditions and the latencies of these differences. To control for false-positives, the MMN differences were considered significant when p-values were lower than 0.05 for 10 (= 20 ms) or more consecutive time points (Guthrie & Buchwald, 1991; Möttönen et al., 2013). These analyses were conducted for the phoneme and tone contrasts and each language group separately.

In a separate set of analyses, mean MMN amplitudes were calculated as the mean voltage across a 40-ms window centred at the peak latency for each participant. Peak amplitudes were calculated in the same way for the identity MMNs obtained in the baseline (no-TMS) session. To test whether significant MMN responses were elicited in the no-TMS and post-TMS sessions, we

compared the peak amplitudes against zero (no change) using one-sample t-tests. The amplitudes and latencies of these responses were compared between groups using two-sample t-tests. To test the influence of language experience on the TMS-induced disruptions of speech motor cortex on the peak amplitude of the MMN responses elicited by phoneme and tone contrasts, separate repeated measures analyses of variance (ANOVAs) were conducted with TMS (no-TMS vs. post-TMS) as a within-subjects factor and language group (tonal vs. non-tonal language speakers) as a between-subjects factor. Interactions were explored using post-hoc paired t-tests (no-TMS vs. post-TMS) for each group separately and to provide Cohen's d effect sizes for future studies.

5.4 Results

5.4.1 Effects of TMS-induced disruption in the right speech motor cortex

We stimulated the right speech motor cortex in 15 non-tonal language speakers and 16 tonal language (Mandarin) speakers and tested the effect of this disruption on discrimination of tone and phoneme contrasts measured by the MMN. We recorded EEG while participants listened to an oddball sequence in no-TMS (baseline) and post-TMS sessions. ERPs to frequent sounds ('da4') were subtracted from responses to infrequent phoneme ('ba1') and tone ('da4') changes to obtain the MMN responses. To evaluate the effect of TMS-induced disruption on these measures of phoneme and tone discrimination in each language group, we first compared the MMN responses in the two sessions using

sequential paired t-tests. This analysis determined the latencies of any significant differences.

TMS-induced disruption of the right speech motor cortex had no effect on the MMN response elicited by the phoneme change (Fig. 5-2A; ‘ba1’) in either the tonal or the non-tonal language group. In contrast, MMN responses elicited by the tone change were significantly suppressed by TMS-induced disruption of the right speech motor cortex at a latency of 158–230 ms after stimulus onset in the non-tonal language group but had no effect on these responses in the tonal language group at any latency (Fig. 5-2B; ‘da4’). In the non-tonal language group there was a significant difference between no-TMS and post-TMS responses to the phoneme change at a latency of 264-300 ms (Fig. 5-2A; ‘ba1’); this late discriminatory negativity occurred after the MMN response and possibly indicates prolonged processing due to the right hemisphere disruption (Cheour, Korpilahti, Martynova, & Lang, 2001).

Peak amplitudes for MMN responses were calculated by averaging the voltage across a 40-ms window centred at the peak latency in each participant for each contrast. We used one-sample t-tests against zero to confirm that significant MMN responses were elicited in both no-TMS and the post-TMS sessions by the phoneme and tone changes in both groups (Table 5-1). Comparison of MMN responses calculated using the traditional method with identity MMNs calculated using the same-stimulus method, confirmed that these changes were due to discrimination of speech sounds (Tables 5-1 and 5-2). Furthermore, during the no-TMS session, we confirmed that the latencies and amplitudes of the MMN

responses elicited by either tone or phoneme changes did not differ between language groups (two-sample t-tests: all $p > .05$).

To test whether the effects of TMS-induced disruption of the right speech motor cortex on the MMN responses were sensitive to the language experience of the participants, we compared the groups using two-way ANOVAs for the phoneme and tone contrasts separately. For the MMN responses elicited by the phoneme change, the TMS-induced disruption in the right speech motor cortex had no effect in either language group (i.e. the main effect of TMS was not significant: $F_{1,29} = 1.02$, $p = .322$, $\eta_p^2 = .03$; see Fig. 5-3A; the main effect of group was also not significant: $F_{1,29} = 0.03$, $p = .858$ $\eta_p^2 < .01$). For the MMN responses elicited by the tone change, there was a significant main effect of group ($F_{1,29} = 8.32$, $p = .007$, $\eta_p^2 = .22$), indicating that tone changes elicited larger MMN responses in tonal language speakers. More importantly, there was a significant interaction between TMS and group ($F_{1,29} = 8.57$, $p = .007$, $\eta_p^2 = .23$; see Fig. 5-3B; the main effect of TMS was not significant: $F_{1,29} = 4.04$, $p = .061$, $\eta_p^2 = .12$). Post hoc paired t-tests indicated that the disruption of the right speech motor cortex suppressed MMN responses to the tone change in the non-tonal ($t_{14} = 3.41$, $p = .004$, $d = .88$) but not in the tonal language group ($t_{14} = .65$, $p = .526$, $d=.16$).

In sum, our results showed that in both language groups, TMS-induced disruption of the right speech motor cortex had no significant effect on MMN responses elicited by a phoneme change. In contrast, the effect of the right-hemisphere disruption on MMN responses elicited by a tone change differed according to language experience: it significantly suppressed MMN responses in the non-

tonal language group but had no effect on MMN responses in the tonal language group.

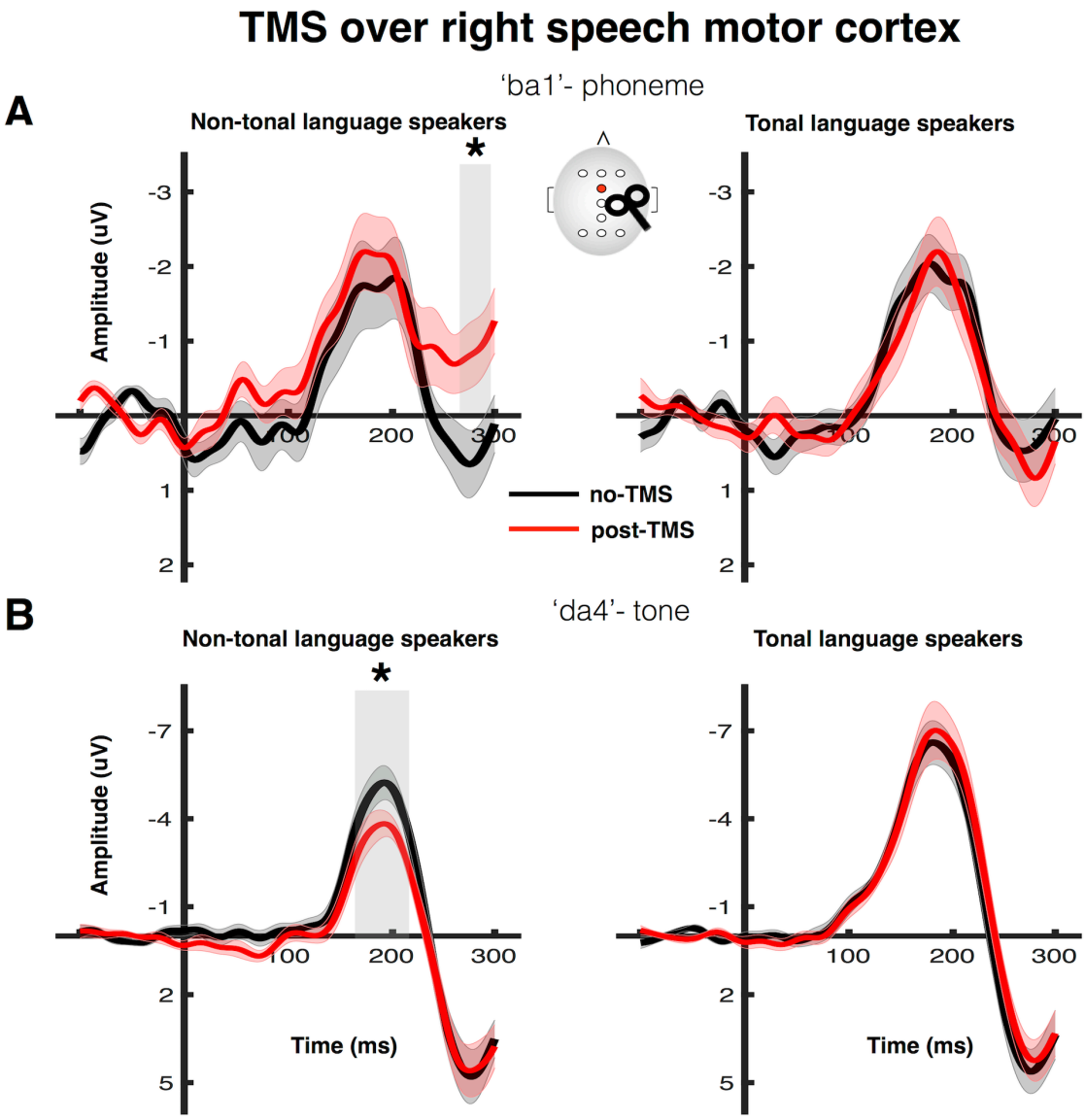


Figure 5-2 Effect of rTMS over right speech motor cortex on MMN responses elicited by the phoneme (A) and tone (B) changes in tonal and non-tonal language speakers.

The MMN responses recorded from the FCz electrode were obtained by subtracting the responses to frequent 'da1' syllable from the responses to

*infrequent 'ba1' (A) or 'da4' (B) syllables. Shaded regions around the MMN responses indicate standard error of the mean (solid line) across participants. The grey rectangle indicates the time windows during which the no-TMS (black) and post-TMS (red) MMN responses differed significantly from each other (sequential t-tests). *p < .05 (2-tailed). Note the different scales used for A and B; MMN responses to the tone change were robustly higher than for the phoneme change.*

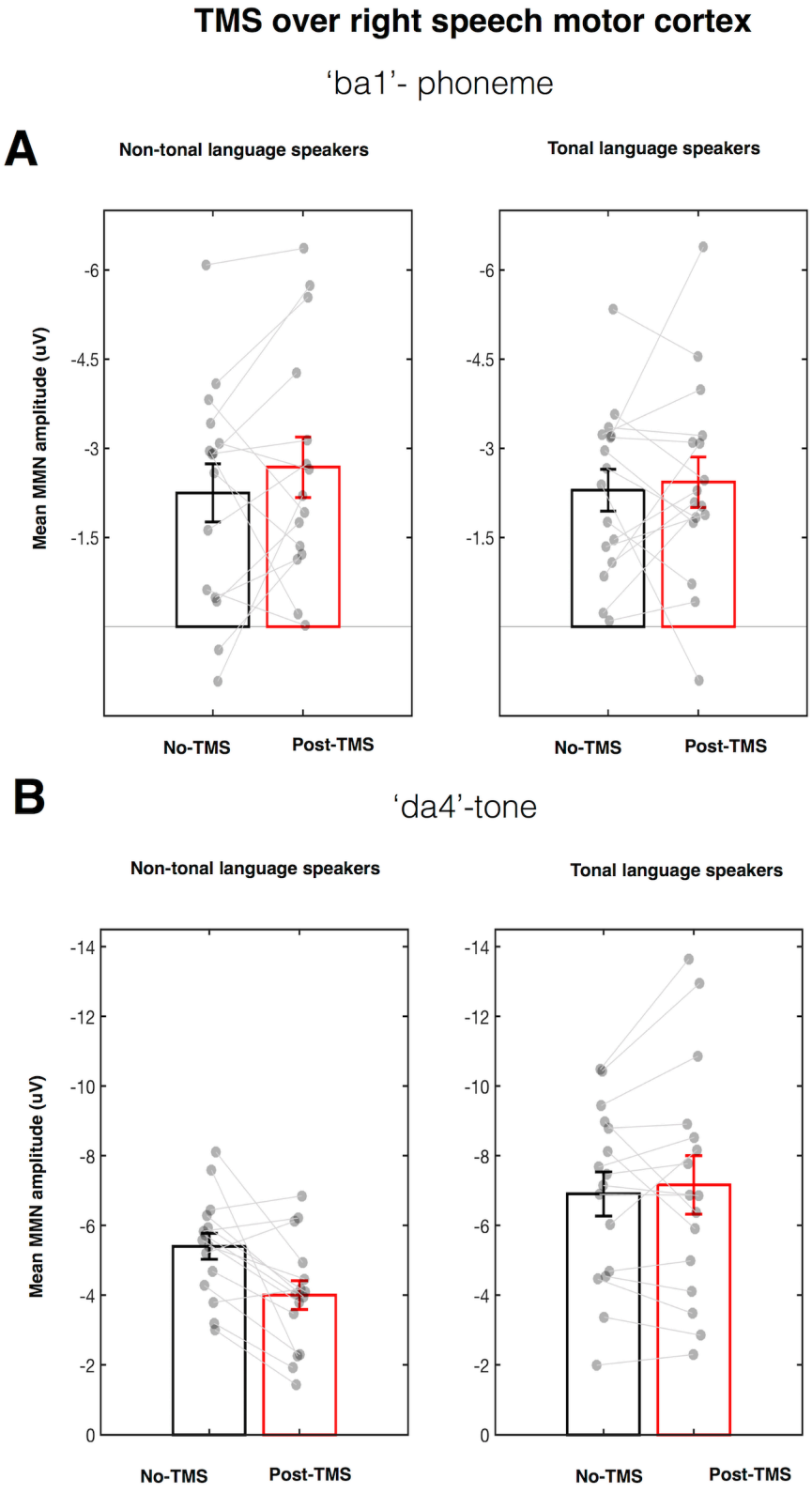


Figure 5-3 Comparison of amplitudes of MMN responses elicited by the phoneme (A) and tone (B) changes in tonal and non-tonal language speakers.

The grey circles show the mean amplitudes of MMN responses obtained during no-TMS and post-TMS recordings for individual participants connected by a grey line. The group mean MMN amplitudes during no-TMS and post-TMS recordings are shown in black and red unfilled bars respectively. Error bars represent standard error of the mean.

5.4.2 Effects of TMS-induced disruption in the left speech motor cortex

We stimulated the left speech motor cortex in another group of 15 non-tonal language speakers and a further 16 tonal language (Mandarin) speakers and tested the effect of this disruption on discrimination of tone and phoneme contrasts measured by the MMN. Data were analysed as described for the right motor cortex.

TMS-induced disruption of left speech motor cortex significantly suppressed the amplitude of the MMN responses elicited by the phoneme change in both the non-tonal and the tonal language groups. These differences were significant at a latency of 162–224 ms after stimulus onset in the non-tonal language group and at latencies of 134–190 ms and 196-224 after stimulus onset in the tonal language group (Fig. 5-4A; 'ba1'). In contrast, MMN responses elicited by the tone change were significantly suppressed by TMS disruption of left speech motor cortex in the tonal language group at a latency of 154–224 ms after stimulus onset (Fig. 5-4B; 'da4'); in the non-tonal group, the slight suppression of MMN responses to the tone change was not significant at any latency (Fig. 5-4B; 'da4').

Analysis of the peak amplitudes (one-sample t-test against zero) confirmed that significant MMN responses were elicited in both the no-TMS and post-TMS sessions by the phoneme and the tone changes in both groups (Table 5-1). Comparison of MMN responses calculated using the traditional method with identity MMNs calculated using the same-stimulus method confirmed (as above for the groups who received right hemisphere stimulation) that these changes were due to discrimination of speech sounds (Tables 5-1 and 5-2). Furthermore, during the no-TMS session, we confirmed that the latencies and amplitudes of the MMN responses elicited by either tone or phoneme changes did not differ between language groups (two-sample t-tests: all $p > .05$).

To test whether the effects of TMS-induced disruption of left speech motor cortex on the MMN responses were sensitive to the language experience of the participants, we compared the groups using two-way ANOVAS for the phoneme and tone contrasts separately. For the MMN responses elicited by the phoneme change, the TMS-induced disruption in the left speech motor cortex significantly suppressed MMN responses in both non-tonal and tonal language groups (main effect of TMS: $F_{1,29} = 15.56$, $p < .001$, $\eta_p^2 = .35$; see Fig. 5-5A; the main effect of group was not significant: $F_{1,29} = 1.09$, $p < .306$, $\eta_p^2 = .04$). Furthermore, the magnitude of the suppression did not differ between the groups (interaction: $F_{1,29} = 0.05$, $p = .823$, $\eta_p^2 = .002$). For the MMN responses elicited by the tone change, the TMS-induced disruption in the left speech motor cortex significantly supressed peak amplitudes in both groups (main effect of TMS: $F_{1,29} = 13.80$, $p = .001$, $\eta_p^2 = .32$; see Fig. 5-5B; the main effect of group was not significant: $F_{1,29} = 1.14$, $p < .294$, $\eta_p^2 = .04$). The suppression in the tonal language group was

numerically larger (mean MMN peak amplitude change ± SE = 1.70 μV ± 0.50, d = 0.85) compared with the size of the suppression in the non-tonal language group (mean MMN peak amplitude change ± SE = 0.94 μV ± 0.50, d = 0.48; see Fig. 5-5B) but this group difference was not statistically significant (TMS x group interaction: $F_{1,29} = 1.14$, p= .294, $\eta_p^2 = .04$).

In sum, our results showed that, in both language groups, TMS-induced disruption of the left speech motor cortex suppressed MMN responses elicited by a phoneme change. In contrast, the effect of left-hemisphere disruption on MMN responses elicited by a tone change differed according to language experience: the reduction in the non-tonal language group was not significant at any latency (see Fig. 5-4B) whereas it was significantly reduced at 154-224 ms latency in the tonal language group. However, the analysis of the peak amplitude MMN responses to tone discrimination during disruption of the left speech motor cortex found that the non-significant small reduction in the non-tonal language group was statistically indistinguishable from the significant and large effect seen in the tonal language group (i.e. the interaction between group and TMS was not significant, p>.05). Examination of data shown in Fig. 5-5B indicates that some individuals in the non-tonal language group showed large reductions in the MMN amplitude elicited by the tone change (‘da4’) but this was not a reliable effect across the group.

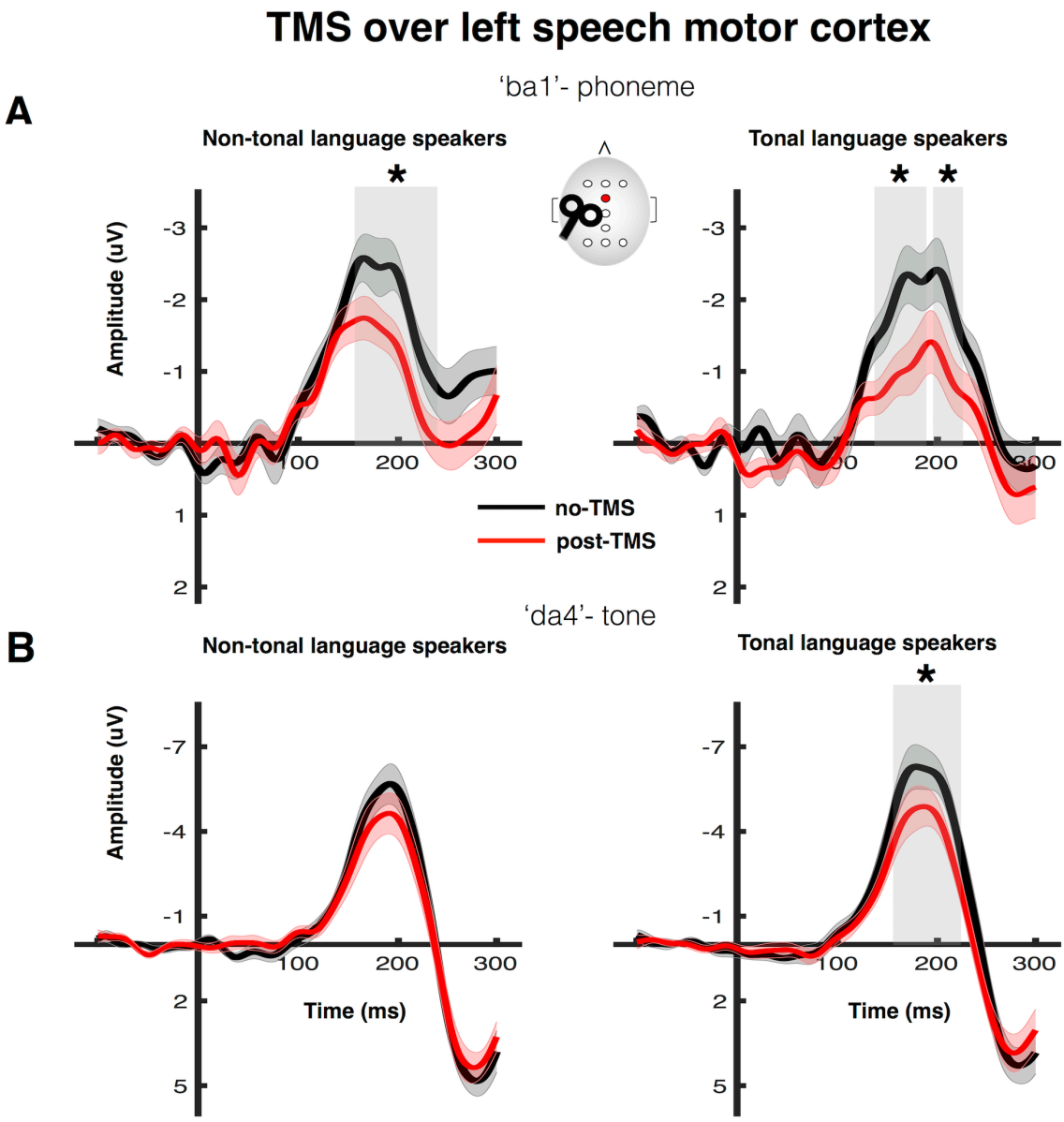


Figure 5-4 Effect of rTMS over left speech motor cortex on MMN responses elicited by the phoneme (A) and tone (B) changes in tonal and non-tonal language speakers.

See legend to Fig. 5-2 for details.

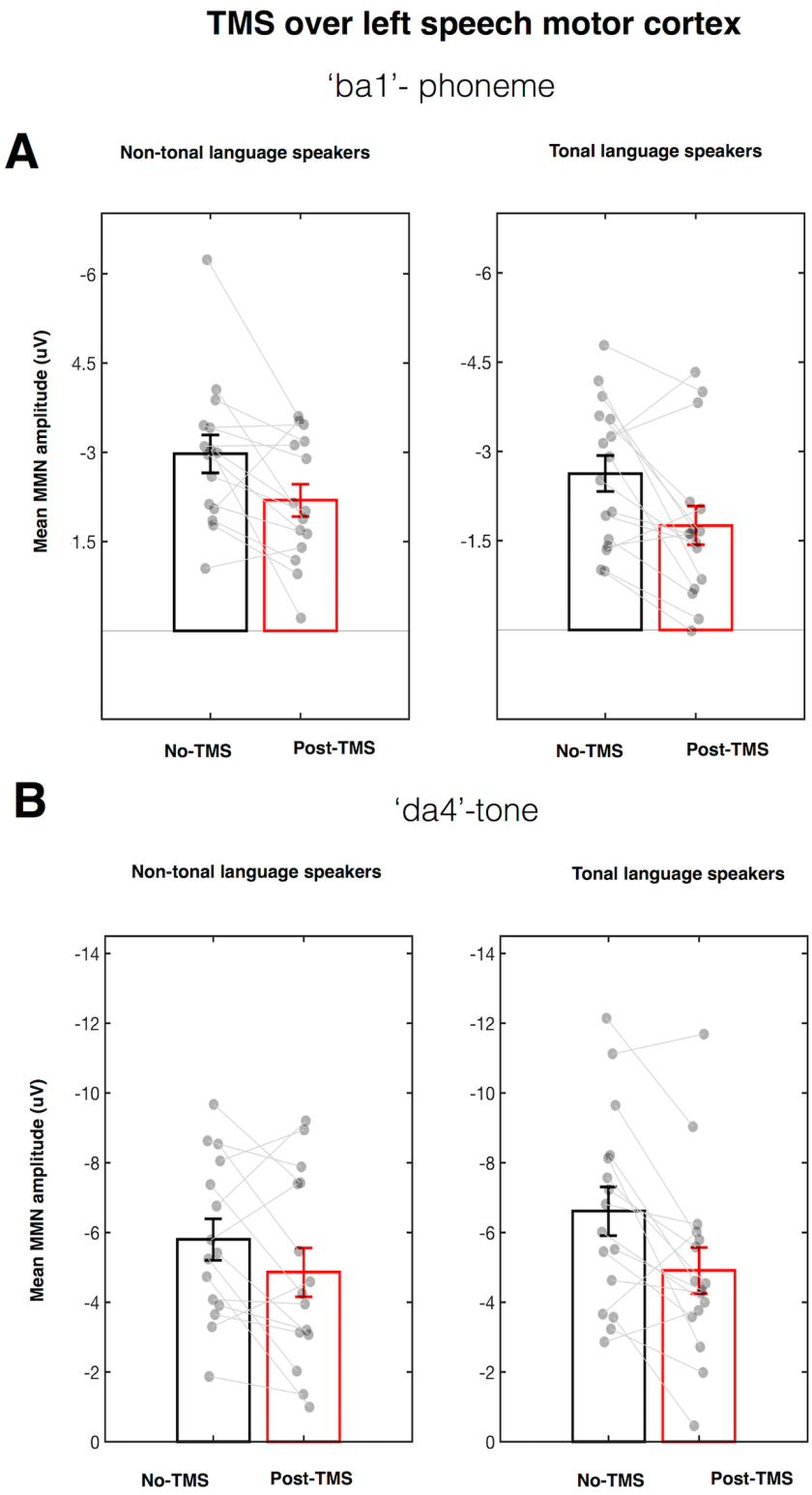


Figure 5-5 Comparison of amplitudes of MMN responses elicited by the phoneme (A) and tone (B) changes in tonal and non-tonal language speakers.

See legend to Fig. 5-3 for details.

Table 5-1

MMN peak latencies (ms) and mean amplitudes (SEM, μV) at FCz

	No-TMS			Post-TMS		
	Latency	Amplitude	t	Latency	Amplitude	t
TMS over right speech motor cortex						
Non-tonal language group, n=15						
'ba1'	182.9	-2.25(0.49)	4.58	194.3	-2.68(0.51)	5.24
'da4'	191.7	-5.40(0.38)	14.39	191.1	-4.00(0.41)	9.69
Tonal language group, n=16						
'ba1'	174.0	-2.29(0.35)	6.52	175.8	-2.43(0.43)	5.70
'da4'	182.5	-6.91(0.63)	10.93	183.3	-7.16(0.84)	8.57
TMS over left speech motor cortex						
Non-tonal language group, n=15						
'ba1'	182.8	-2.97(0.32)	9.35	190.7	-2.19(0.27)	8.07
'da4'	189.7	-5.80(0.59)	9.80	192.0	-4.86(0.70)	6.92
Tonal language group, n=16						
'ba1'	181.5	-2.63(0.30)	8.69	199.4	-1.75(0.32)	5.39
'da4'	191.6	-6.61(0.70)	9.47	187.6	-4.91(0.66)	7.41

Note: The MMN amplitudes were calculated by averaging the responses across a 40-ms window centred at this peak latency and were compared to zero with 2-tailed t-tests. All t-test results are significant with $P < 0.001$.

Table 5-2

***Identity* MMN peak latencies (ms) and mean amplitudes (μV, ±SEM) at FCz**

	Latency	Amplitude	t	p
TMS over right speech motor cortex				
Non-tonal language speakers, n=15				
'ba1'	191.9	-1.37 (0.49)	2.81	< .05
'da4'	192.3	-6.01 (0.42)	14.25	< .001
Tonal language speakers, n=16				
'ba1'	201.4	-2.29 (0.49)	4.69	< .001
'da4'	183.4	-7.50 (0.69)	10.82	< .001
TMS over left speech motor cortex				
Non-tonal language speakers, n=15				
'ba1'	210.4	-2.20(0.30)	7.41	< .001
'da4'	188.8	-7.24(0.71)	10.21	< .001
Tonal language speakers, n=16				
'ba1'	204.6	-2.64(0.41)	6.34	< .001
'da4'	193.3	-7.62(0.68)	11.28	< .001

Note: The identity MMN responses were calculated by subtracting event-related potentials evoked by the stimuli presented frequently (probability=1.0) in the control sequence from those evoked by identical sounds presented infrequently (probability= 0.1) in the oddball sequences (with no TMS). Mean amplitudes were calculated by averaging the responses across a 40-ms window centred at this peak latency and were compared to zero with 2-tailed t-tests.

5.6 Discussion

We used EEG and TMS to investigate the asymmetry of auditory-motor processing of phoneme and tone changes in speakers of tonal and non-tonal languages. Disruption of the left, but not the right, speech motor cortex impaired automatic auditory discrimination of phonemes in both tonal and non-tonal language groups. Disruption of the right speech motor cortex modulated auditory discrimination of tones in the non-tonal language group but had no effect in the tonal language group. Tone discrimination in the tonal language group was affected by disruption of the left speech motor cortex. The effect of disruption of the left motor cortex on tone discrimination in the non-tonal language group was inconclusive: in one analysis, the MMN responses to tone changes were not significantly reduced by the TMS disruption at any latency, whereas the analysis of changes in the peak amplitude found that the small suppression in the non-tonal language group was not statistically different to the larger suppression seen in the tonal language group. This is the first study to provide direct evidence that the lateralization of motor contributions to auditory speech processing is determined by the functional role of the acoustic cues in the listener's native language.

Our finding that disruption of the left motor cortex impaired auditory processing of both phoneme and tone changes in tonal language speakers is consistent with findings from previous studies. For example, early lesion studies with patients, reported impaired lexical tone production and perception in tonal language speakers following damage to the left hemisphere (Eng et al., 1996; Gandour et al., 1992; Yiu & Fok, 1995). Later, functional neuroimaging data also revealed

that in tonal language speakers processing of lexical tone evoked activity in several left-hemisphere regions, including the left posterior temporal cortex (Gandour et al., 2003, 2000; Klein et al., 2001; Pierce et al., 2014).

In the current study, disruption of the right speech motor cortex impaired tone discrimination in non-tonal language speakers. As for the findings in tonal language speakers, these results are consistent with a known pattern of right hemisphere lateralization in non-tonal language speakers for tone and prosody perception based on findings from the lesion deficit (Zatorre & Samson, 1991) and neuroimaging literature (Albouy et al., 2020; Boemio et al., 2005; Schönwiesner et al., 2005).

The motor contribution to auditory tone processing was strongly left-lateralized in tonal language speakers, while it was weakly right-lateralized in non-tonal language speakers. These findings are consistent with previous work indicating that processing of spectral modulations was weakly right lateralized compared with a strong left lateralization for processing of temporal modulations (Flinker et al., 2019). We found that the MMN responses elicited by tone changes in tonal language speakers were unchanged following right hemisphere motor disruption. For the non-tonal language speakers, right hemisphere motor disruption reduced MMN responses to tone changes by a significant amount but there was also a non-significant small reduction in these responses following left hemisphere motor disruption. The non-significant reduction following left hemisphere stimulation in non-tonal language speakers might suffer from low power, due to the relatively small sample size per group (N = 15

or 16). Further work with a larger sample size could clarify this point. In a previous study of non-tonal language speakers, behavioural discrimination of prosody in spoken questions and statements was impaired by disruption of right but not left premotor cortex (Sammler et al., 2015). It is possible that use of the MMN offers greater sensitivity compared with behavioural measures of task performance to reveal the subtle effect of left-hemisphere disruption in the non-tonal language speakers, which was considerably smaller than the effect of right-hemisphere disruption. One advantage of using the MMN is that it provides an automatic measure of auditory discrimination and, unlike task performance, does not require the listener to attend and respond to the relevant features of the sounds. Such task demands could interact with TMS-induced effects on auditory processing. It is possible that the auditory-motor processing of tone in speech is bilaterally organized with a weak right-hemisphere bias in non-tonal language speakers. However, the focus of attention on tones could enhance auditory-motor processing in the right hemisphere leading to increased lateralization.

The current study confirms the contribution of speech motor cortex to auditory processing of phoneme and tone changes in both tonal and non-tonal language speakers. The findings therefore provide further evidence for the causal role of the speech motor cortex in speech processing (Davis & Johnsrude, 2007; Möttönen et al., 2013). Note that our results provide support for the critical role of speech motor cortex in speech perception but not speech or word recognition. It is important to be aware that although the syllables used in current study (‘da1’, ‘da4’, ‘ba1’) have “lexical significance” in Mandarin, it is impossible to determine

their meaning by listening to them in isolation. For example, the syllable ‘da1’ could mean “build”, the sound of a horse’s hooves, “floppy”, and tens of other meanings dependent on the context. Thus, for both non-tonal and tonal language speakers, passive listening to the oddball sequences involved automatic discrimination of the syllables (i.e. sublexical processing) rather than access to meaning. The complementary contribution of speech motor cortex to speech perception might be observed only when contextual information is unavailable (Wilson, 2009); this could partially reconcile findings from studies in patients showing relatively intact speech comprehension after damage to frontal-motor brain areas (Hickok, 2010; Hickok et al., 2008). In real life conditions, listeners, including patients with Broca’s aphasia, can make use of contextual information to “comprehend” words even when their speech perception is partially impaired.

The auditory-motor processing of speech sounds observed in the current study was not feature-specific: the disruption of speech motor cortex involved targeting the lip representation separately in each hemisphere. The disruption affected both phoneme and tone discrimination, even though tone production is controlled by the larynx. This is consistent with our previous work showing that disruption of the left lip but not the left hand representation suppressed MMN responses evoked by infrequent lip-related ‘ba’ and tongue-related ‘ga’ sounds presented among frequent tongue-related ‘da’ sounds (Möttönen et al., 2013); the disruption had no effect on discrimination responses for non-speech sounds. However, previous behavioural studies suggested that the contribution of the motor cortex to speech perception was somatotopically organized (D’Ausilio et al., 2009; Mottonen & Watkins, 2009). The non-specific effect observed here could be due

to the close proximity of the lip and larynx representations, which makes it impossible to disrupt the lip without affecting the adjacent larynx representation (Eichert, Papp, Mars, & Watkins, 2020; Grabski et al., 2012). An alternative explanation is that the feature-specific motor contributions to speech perception depend on attention. We showed previously that when speech sounds are ignored, the motor contributions to auditory speech processing are late (>170 ms) and not feature-specific, whereas when speech sounds are attended, there is an additional early (<100 ms) motor effect on auditory speech processing, which is feature-specific (Möttönen et al., 2014).

Our findings lead us to suggest that the asymmetry in the auditory system during speech perception is modulated by the motor system. Recently, a modified version of the Asymmetric-Sampling-in-Time hypothesis was proposed, in which auditory processing is thought to be modulated by two different types of mechanism (Assaneo et al., 2019; Rimmele, Morillon, Poeppel, & Arnal, 2018): i) intrinsic auditory mechanisms, which support auditory processing in non-speech listening conditions in line with acoustic hypotheses; ii) top-down mechanisms, which might affect the lateralization of auditory processing, and the relative weight of which can increase depending on individual brain differences or specific speech-listening conditions or both. During continuous speech perception, the speech-coupled oscillations in auditory cortex are significantly modulated by top-down signals from the frontal and motor cortex (Park, Ince, Schyns, Thut, & Gross, 2015). This might partially account for left-lateralized auditory speech processing that is determined by linguistic experience and the functional role of acoustic cues. A recent study in non-tonal language speakers

investigated the hemispheric asymmetry of auditory-motor interactions during speech feedback control and found that the asymmetry was affected by spectro-temporal cues (Floegel, Fuchs, & Kell, 2020). Our results predict that language experience and the functional role of the spectro-temporal cues may also affect the involvement of left and right auditory-motor interactions in speech feedback control during speech production.

In sum, our study, for the first time, presents causal evidence that the contribution of the right and left speech motor cortex to auditory speech discrimination ability is determined by the functional role of acoustic cues in the listener's native language. This motor system asymmetry could modulate the asymmetry in the auditory system. Our study emphasizes the importance of understanding the interactions between bottom-up auditory mechanisms and top-down influences when investigating hemispheric asymmetry of speech processing. We also call for more comparative studies between speakers of different languages, since ignoring the diversity of spoken languages can to lead to an incomplete understanding of the neurobiology of speech and language.

Chapter 6 General Discussion

The overarching aim of this thesis was to improve our understanding of auditory-motor interactions during both speech production and perception. To achieve this, I conducted three studies, in which I used a multi-modal approach including real-time formant perturbation, brain stimulation, and EEG recording. The studies contributing to this thesis raised a number of interesting issues. In the following sections, I will discuss these issues in the context of limitations of our studies and also ideas for future work.

6.1 Real-time formant perturbation

6.1.1 Main findings of study 1 and study 2

Real-time feedback perturbation has been extensively used to explore auditory feedback and speech motor control. In study 1 and study 2, I characterized sensorimotor learning in speech induced by formant perturbation and the causal contribution of primary motor cortex to such learning. In study 1, results showed that feedback perturbation of single formant induced significant compensatory changes in both the perturbed and unperturbed formant that were in opposite directions by shifting F1 or F2 in Experiments 1 and 2 respectively. In study 2, I then showed that compensatory changes in both the perturbed and unperturbed formant were impaired by disruptive TMS over the tongue, but not hand representation.

6.1.2 Develop standards to test and report speech adaptation induced by real-time formant perturbation

Over the past two decades since the first introduction of this paradigm by Houde & Jordan (1998), various teams have used real-time formant perturbation to investigate speech-motor adaption. However, researchers used different methods to achieve perturbation, and also applied different procedures to analyse the data. These differences make it difficult to compare results directly across studies, and more importantly might lead to inconsistences in results reported by different teams. For example, there are some differences in our findings in study 1 and a previous study by MacDonald, Purcell, & Munhall (2011), which used a similar paradigm. These differences could be due to experimental or data analysis differences between the two studies. For example, in the previous study, only 50 utterances were produced during the learning phase, which was assumed to be sufficient for speech adaptation to reach a plateau. However, in our study, participants were instructed to produce 225 utterances in total during the learning phase. Our results showed that changes in the shifted and unshifted formants continued to occur after 50 utterances i.e. they had not reached a plateau in terms of adaptation. The second experimental difference between these two studies was the selection of training words. In our study, utterances were of three different words with the same vowel, namely “head”, “bed”, “dead”, rather than just the one word, “head”, used in MacDonald, Purcell, & Munhall (2011). Intuitively, the varied context should reduce the rate of learning (which might explain the delayed plateau) but allows for transfer or generalisation of the adapted vowel to different contexts and possibly a more stable effect by the end of learning. Finally, the magnitude of the shifts applied to feedback in the

two studies was also different. The magnitude of the shift in F1 feedback was somewhat larger in the MacDonald study (200Hz) than the equivalent of 110 mel which we applied in Experiment 1 (150Hz). However, the F2 shifts applied in the two studies were equivalent: 250Hz and 265Hz. In addition to these differences in experimental procedure, there were some differences when choosing how to quantify sensorimotor learning. In our study, instead of using raw data, we introduced a way to normalize the data according the mean frequency of baseline values and the participants' variability in those utterances produced with normal feedback. We calculated a Z-score for each utterance by subtracting the mean and dividing by the standard deviation of the formant frequencies produced during baseline (with normal feedback).

Some of the differences between our study and MacDonald, Purcell, & Munhall (2011) might have affected the results, while others might be unimportant. It is clear that some consensus on standards for experimental procedures and analysis methods are required. In our view, compared with using raw data normalised by subtraction of the mean baseline value, calculation of z-scores based on baseline mean values of frequency of formants and their variance would be more objective and allow comparison of changes between participants. Implementing this change in both study 1 and 2, did not alter the results of statistical analysis compared with analysis of data normalised by subtraction (see Appendices 1 and 2).

In our studies, we further improved the experimental design by implementing the use of noise-masked sessions before and after the learning phase. The sessions

allowed us to compare the effect of the perturbation on speech production in identical conditions (i.e, when feedback was blocked) rather than comparing production with and without the perturbation as in previous designs. In our view, this offers a “purer” measure of the effects of the perturbation. In study 1, the same results were obtained and presented by analysis of the data from the noise-masked sessions. In study 2, however, the masking was unsuccessful due to a technical problem, so these data were excluded from further analysis. Even so, we think the use of short noise-blocked sessions should become a standard in these studies.

Finally, as mentioned above, previous studies calculated learning related changes by averaging formant values for a variable number of utterances at the end of the learning phase (e.g. the last 15 compared with the last 30 utterances in Lametti et al., 2018; MacDonald et al., 2011). The choice of how many utterances to average is one of the most variable among studies even from the same research group. This choice offers too many degrees of freedom for analysis. In addition, limiting the measurement to this point of the adaptation phase ignores the possible differences among participants in the rate of adaptation. Use of the latter to quantify the learning effect presumes that the formant produced has reached relatively steady state by the end of learning phase. However, inspection of the data suggests this is not always the case and, indeed, this may continue to differ among participants even during the late learning phase. An alternative that we implemented in study 1 and study 2 is to calculate the average across the whole learning phase, which captures both rate and magnitude of adaptation during learning rather than just focus on a limited

(and variable) number of utterances at the end of learning phase. Other alternatives including learning curve fitting, which is commonly used in visuomotor learning (Smith, Ghazizadeh, & Shadmehr, 2006) and linear mixed model, which includes all trials during the learning phase into the model (Scott et al., 2020) should also be better explored.

6.1.3 Clinical applications of real-time perturbation paradigm

During the past two decades, a broad range of research topics have been investigated and addressed using formant perturbation paradigm. I believe that over the next 20 years, interest in using this paradigm will increase.

More recently, real-time perturbation paradigm has been used to investigate the mechanisms that cause instances of stuttering moments in the speech of individuals who stutter (Daliri, Wieland, Cai, Guenther, & Chang, 2018; Kim, Daliri, Flanagan, & Max, 2020; Packman & Attanasio, 2017). Max and his colleagues (Max, Guenther, Gracco, Ghosh, & Wallace, 2004) have proposed that individuals who stutter may have difficulty learning and updating accurate internal models of the complicated transformations from neuromuscular activation to vocal tract configurations and from those articulatory movements to acoustic output. By using real-time formant perturbation paradigm, they tested their hypothesis and confirmed that both children and adults who stutter exhibited significant limitations in speech adaptation. Speech adaptation for people who stutter was significantly smaller than that for age-matched controls in both the suddenly introduced and the gradually introduced formant perturbation conditions (Kim et al., 2020).

Some researchers have started to focus on the clinical applications of this paradigm (Parrell & Niziolek, 2020). For example, healthy individuals were shown to be capable of expanding their vowel space as well as the contrast between individual vowels during sensorimotor adaptation and after a further delay period (Parrell & Niziolek, 2020). Reduced vowel space and vowel contrast are common characteristics occurring in many motor disorders, and have been associated with declined speech intelligibility in these populations (Higgins & Hodge, 2002; H. Kim, Hasegawa-Johnson, & Perlman, 2011; Liu, Tsao, & Kuhl, 2005; Skodda, Visser, & Schlegel, 2011). Findings from previous study point towards the possibilities of using sensorimotor adaptation as a clinical tool to enhance speech intelligibility. However, for clinical applications, a crucial problem is “retention”: any learning-related changes must be maintained for a period after training is over. The effects of sensorimotor adaptation are often thought to be transient, which disappear a few utterances after the feedback return to normal (Lametti, Smith, Freidin, et al., 2018; Mitsuya et al., 2011). In the study of Parrell & Niziolek (2020), some effects of sensorimotor adaptation are maintained after the perturbation is removed as well as after a subsequent 10 minute silent interval, which are in line with previous results in whispered speech (Houde & Jordan, 2002). More research on retention of sensorimotor adaptation, particularly after multiple training sessions and over longer intervals, will be crucial to future clinical application of this paradigm.

6.2 The effects of TMS over M1 on the neural activities in functionally connected remote areas

In this thesis, the real-time formant perturbation paradigm was used in combination with transcranial magnetic stimulation (TMS) over the motor cortex to evaluate the contribution of speech motor cortex to sensorimotor adaptation. We first confirmed the effectiveness of low-frequency rTMS on temporal disruption of tongue (proof-of-principle for tongue) motor excitability (see Chapter 2). We also confirmed the effectiveness of the paradigm on the hand motor excitability in the study participants in the hand group (Chapter 4), as seen previously in work from our lab (Mottonen & Watkins, 2009). Our experiment showed that TMS-induced disruption of left tongue representation impaired sensorimotor learning in speech. The same adaptive learning was not affected by TMS-induced disruption of the nearby hand representation (control area), which suggest a relatively high spatial focality provided by this form of TMS. Even so, activity in regions located more remotely that are functionally connected to the stimulated site can be affected by stimulation (Paus et al., 1997). In particular, TMS effects in distant regions have been demonstrated by using either TMS alone (Bäumer et al., 2003), or in combination with other brain recording techniques, such as EEG (Paus, Sipila, & Strafella, 2001; Strens et al., 2002), PET (Paus et al., 1997; Strafella & Paus, 2001), and fMRI (Bestmann, Baudewig, Siebner, Rothwell, & Frahm, 2004; Bohning et al., 1998; Jung, Bungert, Bowtell, & Jackson, 2020). For example, Bestmann et al., (2004) measured fMRI signal changes in cortical and subcortical motor regions during high-frequency (3.125 Hz) rTMS applied over the hand representation in left M1. They found that such stimulation applied at suprathreshold intensity activated a network of areas

including M1, SMA, dorsal PM, and even the putamen and thalamus, while the same stimulation applied at subthreshold intensity elicited no MRI-detectable activity in distant areas. Similar results are reported in a study stimulating M1 using low-frequency (1Hz) rTMS, namely that higher intensity TMS caused greater activation compared with lower intensity stimulation, both locally and remotely (Bohning et al., 1999). As our protocol uses subthreshold stimulation at a low frequency, we believe that it was unlikely to affect the ongoing activity in remote but functionally connected areas.

6.3 Hemispheric asymmetry during auditory speech processing

6.3.1 Main findings of study 3

Having established with TMS a role for the speech motor cortex in sensorimotor adaptation, a form of motor learning that involved speech production, I shifted my attention for the final experiment to establishing the contribution of the speech motor cortex to speech perception. Specifically, I wished to examine interactions between auditory and motor areas during speech perception and how these differed when processing tone in tonal and non-tonal language speakers. We found that the contribution of the left and right speech motor cortex to auditory speech processing was determined by the functional role of the acoustic stimulus in the listener's language, that is, TMS-induced disruption of the right speech motor cortex suppressed MMN responses elicited by tone changes in non-tonal language speakers, while TMS-induced disruption of the left speech motor cortex suppressed MMN responses elicited by tone changes in tonal language speakers.

6.4 Top-down influences during auditory speech processing

These findings indicated that the existing asymmetry in the auditory system can be modulated by interactions with the motor system.

Note that our results only provide evidence for the contributions of motor cortex to syllable-level discrimination. In speech, pitch changes can cue contrasts at multiple levels of phonological representation (e.g., syllable, word, sentence). It allows for the possibility that different brain regions may be differentially weighted in hemispheric lateralization depending on different processing levels. Under specific processing levels, the relative weight of top-down mechanisms can increase, which might affect the lateralization of auditory processing. In study 3, we confirmed the top-down influence of motor cortex to syllable discrimination and suggested that the asymmetry of this motor contribution plays a critical role in modulating asymmetry in the auditory system during phoneme/tone discrimination.

The relative weight of top-down influences from other brain areas, such as regions in the frontal or parietal lobes, might increase when it comes to other processing levels. In a previous fMRI study, brain activity elicited by selective attention to either stress or intonation in paired sentences were examined in both Chinese and English speakers when they were required to make speeded-response discrimination judgments (Tong et al., 2005). Results showed that compared to the English group, the Chinese group exhibited significantly greater activity in the left supramarginal gyrus and posterior middle temporal gyrus for

both tasks. Similar results were reported previously, namely that a group of Chinese speakers exhibited greater activity compared with a group of English speakers in a left inferior parietal region when they selectively attended to Chinese intonation or tone (Gandour et al., 2004).

Overall, our findings emphasize the importance of understanding the interactions between bottom-up auditory mechanisms and top-down modulation when investigating hemispheric asymmetry of auditory speech processing. Linguistic status, memory, attention, and possibly other factors can modulate hemispheric asymmetry of auditory speech processing. Such modulating effects could come from motor, parietal or frontal cortex. Further investigation aimed at understanding the interactions between bottom-up auditory mechanisms and top-down modulation is likely to prove fruitful.

6.5 Implications for current speech production and perception models

6.5.1 speech production model

In chapter 1, I gave a brief introduction to the DIVA model, one of the most influential models which provides a computationally and neuroanatomically explicit account of the network of brain regions involved in speech production (Tourville & Guenther, 2011). According to the DIVA model, when a mismatch between the predicted and actual sensory signals occurs, an error signal is generated and then sent to the feedback control map in right vPMC to create a corrective motor command (Figure 1-1). The corrective motor command would be implemented via M1 projections to the articulators. The findings of study 2

(chapter 4), namely confirmation of a causal contribution of the articulatory representation in left M1 to speech motor adaption, are consistent with the processes implied in the final output stages of this model.

Specifically, our results suggest that the articulatory representation in left M1 is also involved in transforming the auditory error signals into corrective motor commands. It is possible that TMS-induced disruption of left M1 impaired the motor adjustment to the auditory error signals induced by the formant perturbation.

Another possibility is that disruption of the articulatory representation in M1 impaired sensorimotor learning by impairing speech perception. The DIVA model also proposes that the current auditory states (actual signals) are available in the auditory state map (putatively located in the Heschl's gyrus and posterior superior temporal gyrus, see figure 1-1). Previous studies (including our own) have demonstrated a casual role for M1 in speech perception (see Liebenthal & Möttönen, 2018 for review). It is possible that rTMS over the articulatory representation impaired participants' perceptual acuity (less accurate auditory signals). The DIVA model predicts that people with better auditory discrimination ability have more distinguishable auditory goals and that as a result the magnitude of adaptation to auditory perturbations is greater in such individuals (Tourville & Guenther, 2011). Further work examining the effect of rTMS-induced disruption over the articulatory representation on auditory acuity and how it affects adaptation could clarify this point.

6.5.2 speech perception models

As introduced in chapter 1, for much of the past 20 years, one of the main debates in the scientific study of speech perception has been whether the motor system plays a critical and necessary role in the initial stages of speech perception. Broadly speaking, three classes of theories/models about the auditory-motor interaction during speech perception have been proposed: a parallel, a hierarchical, and an interactive model (Liebenthal & Möttönen, 2018).

In a parallel model of auditory-motor speech processing, the auditory and motor systems function independently. In this type of model, the auditory system is responsible for speech perception while the motor system is responsible for response selection at a late phase. A core claim of parallel models is that activation of motor areas observed during speech perception solely reflects processes related to (motor) response selection and bias (Hickok, 2010; Venezia et al., 2012).

In contrast to parallel models, hierarchical models allow for a close coupling between auditory and motor systems but suggest that the primary direction of information flow is from auditory to motor areas. In a hierarchical model of auditory-motor speech processing, acoustic signals in the auditory areas are mapped into articulatory forms in the articulatory network, including frontal and motor areas, and are subsequently fed back into auditory areas. On this account, motor feedback is suggested to serve as an error prediction signal. One example of a hierarchical model is the model proposed by Rauschecker and Scott (2009), which performs both a forward mapping, transforming

acoustic inputs into motor-articulatory representations (Brodmann area 44 and ventral premotor), whose activation is transmitted to the inferior parietal lobe as an efference copy, and an inverse mapping, whereby attention- or intention-related changes affect the selection of context-dependent action programs in frontal and motor areas.

An interactive model of auditory-motor speech processing suggests that the auditory-motor interactions are bi-directional and can occur at an early phase with several iterations. In chapter 1, I introduced the dual-stream model proposed by Hickok and Poeppel (2000, 2004, 2007), which can be seen as partially consistent with an interactive model, as it proposes direct motor influences on perception via the sub-lexical dorsal pathway. However, Hickok and Poeppel's model does not predict a critical and necessary role for the motor system in speech perception under "normal" conditions. Instead, it postulates that motor influences on speech perception are limited to certain tasks (e.g. detection of subtle differences on a continuum of phonemes) or environments (e.g. noisy) that make speech perception hard (Stokes, Venezia, & Hickok, 2019; Wu, Chen, Wu, & Li, 2014).

In study 3 (chapter 5), we investigated the causal contributions of left and right speech motor cortex to automatic speech processing in both tonal and non-tonal language speakers. The results demonstrated auditory-motor interactions at an early phase (within 200 ms) of automatic speech perception, which provide support for a more comprehensive interactive model. Interestingly, Hickok and Poeppel (2007) point out that unlike the ventral stream which is

largely bilaterally organized, the dorsal stream supporting auditory-motor interactions for speech processing is strongly left-hemisphere dominant. Our results suggest that auditory-motor interactions of speech signals involve both hemispheres but vary with linguistic function and the speaker's language background.

6.6 Conclusion

The studies in this thesis investigated auditory-motor interactions during both speech perception and production. My investigations demonstrated the fruitfulness of combining TMS with (i) sophisticated behavioural paradigms aimed at testing speech motor learning and (ii) EEG measures of behaviour that do not require attention or task performance. I have demonstrated that the paradigm of low frequency rTMS is not only well tolerated by research participants (I tested approximately 120 participants in total for this thesis) but it is effective in reducing motor excitability for the expected period of time. I confirmed that as for previous studies in the hand and lip motor cortex, low frequency rTMS reduces excitability in the tongue representation and is effective also in Chinese individuals. Finally, my third study, emphasises the importance of extending research to include more diverse language speakers as ignoring the diversity of spoken languages can to lead to an incomplete understanding of the neurobiology of speech and language.

Appendices

Appendix 1: Study 1 results

In the main text, we introduced a new normalization method (Z-score). Here, I present analysis of data from Chapter 3 normalised by subtracting the mean baseline frequencies from the frequencies in other phases of the experiment, which has been the method most commonly used in previous studies. The pattern of results obtained is the same as that presented in Chapter 3 for analysis of the Z-score normalized data.

Results

Experiment 1: F1 shift of +110 mels

Figure S1A shows how F1 and F2 produced by participants, normalized by subtraction of the mean in the baseline phase, changed during the sensorimotor learning and unlearning phases in response to a 110 mel increase in F1. By the end of the learning phase, participants decreased their F1 frequency by approximately 30 mel and increased F2 by approximately 15 mel to compensate for the upward shift in F1 (F1+110 mel). Directional t-tests showed that both F1 and F2 frequency at the end of the learning phase (last 30 utterances of learning phase) were significantly different from zero (baseline average) (Figure S1B, F1: t_{19}= 5.30, p < .001, d = 1.19; F2: t_{19} = 3.27, p = .002, d = 0.73). Analysis of the normalised values (subtraction) for F1 and F2 averaged across the whole learning phase similarly showed a significant difference from baseline (F1: t_{19}= 5.67, p < .001, d = 1.27; F2: t_{19} = 1.97, p = .032, d = 0.48).

By the end of unlearning phase, both F1 and F2 production returned to baseline (Figure S1C, one-sample t-test against baseline average, Z = 0: F1: t_{19} = 1.99, p = .061; F2: t_{19} = 2.03, p = .057).

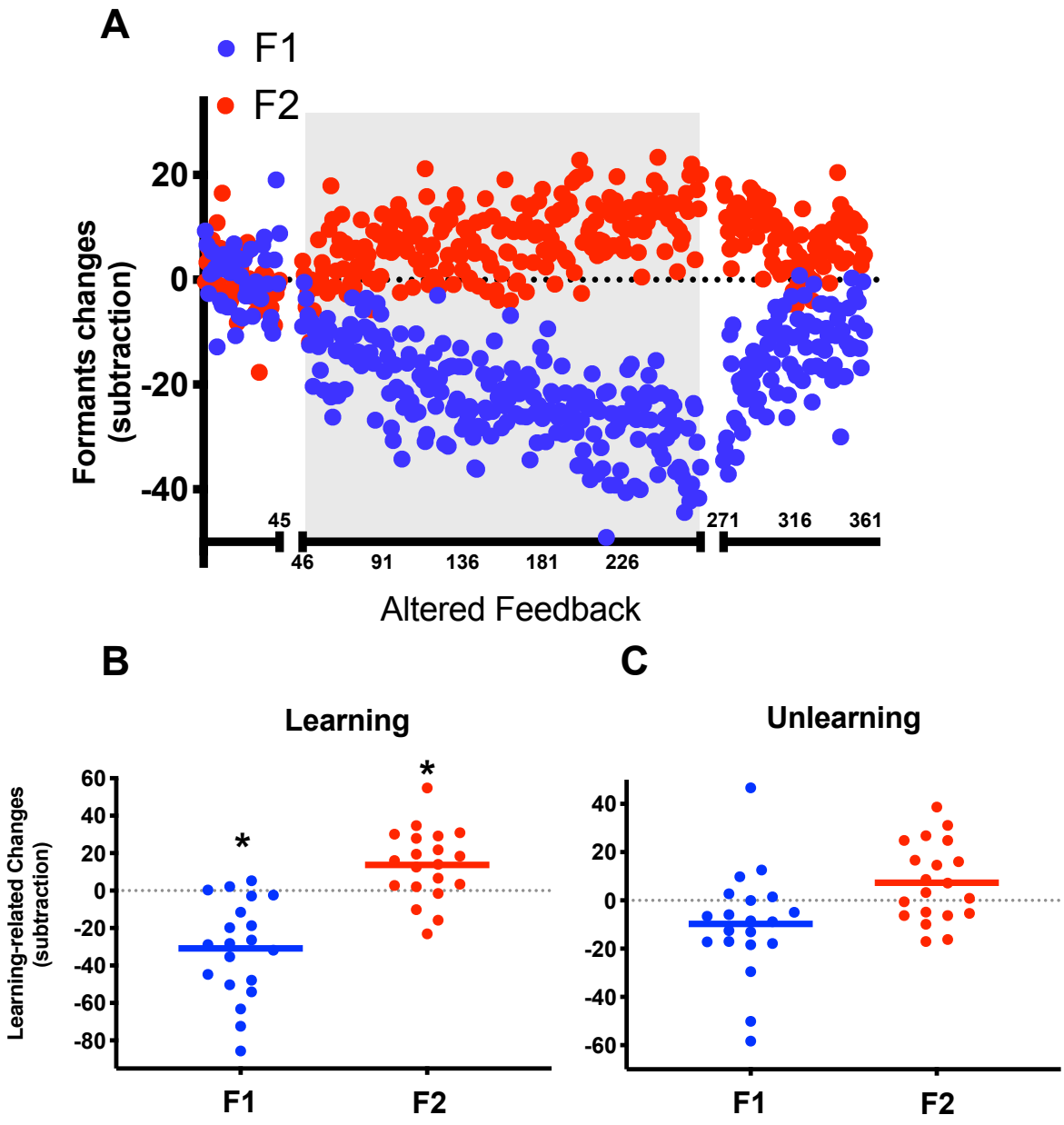


Figure S1 Experiment 1: F1 shifted upwards by 110 mel. Change in formant frequencies (mel) during speech production.

*Values for F1 are shown in blue and F2 in red. The dashed line reflects no change from baseline. The graph in the top panel (A) shows the group average changes across learning and unlearning phases of the experiment relative to baseline; solid circles indicate the group average for each trial; the shaded area represents the trials where the altered feedback was applied. The graphs in the bottom panel (B and C) show average changes in formant frequencies for individual participants for the Learning (B) and Unlearning (D) phases of the experiment relative to baseline; the solid lines represent the group mean; * indicates significant change from baseline.*

Experiment 2: F2 shift of -110 mels

Figure S2A shows how F1 and F2 produced by participants, normalized by subtraction, changed during the sensorimotor learning and unlearning phases in response to a 110 mel decrease in F2. By the end of the learning phase, participants had increased their F2 on average by 50 mel to compensate for the decrease in F2 and decreased F1 on average by 14 mel. Directional t-tests showed that both F1 and F2 frequency at the end of the learning phase (last 30 utterances of learning phase) were significantly different from zero (baseline average) (Fig. S2B, F2: t_{19} = 9.02, p < .001, d = 1.74; F1: t_{19} = 2.42, p = .013, d = 0.54). Analysis of the normalised values (subtraction) for F1 and F2 averaged across the whole learning phase similarly showed a significant difference from baseline (F1: t_{19}= 1.95, p = .033, d = 0.44; F2: t_{19} = 7.73, p < .001, d = 1.73).

By the end of the unlearning phase, F2 frequency returned to baseline (one-sample t-test against baseline average, Z = 0: t_{19} = 1.78, p = .092), while F1 frequency remained significantly reduced relative to baseline (t_{19} = 2.45, p = .012, d = 0.61) after these 90 trials with normal feedback (Fig. S2C).

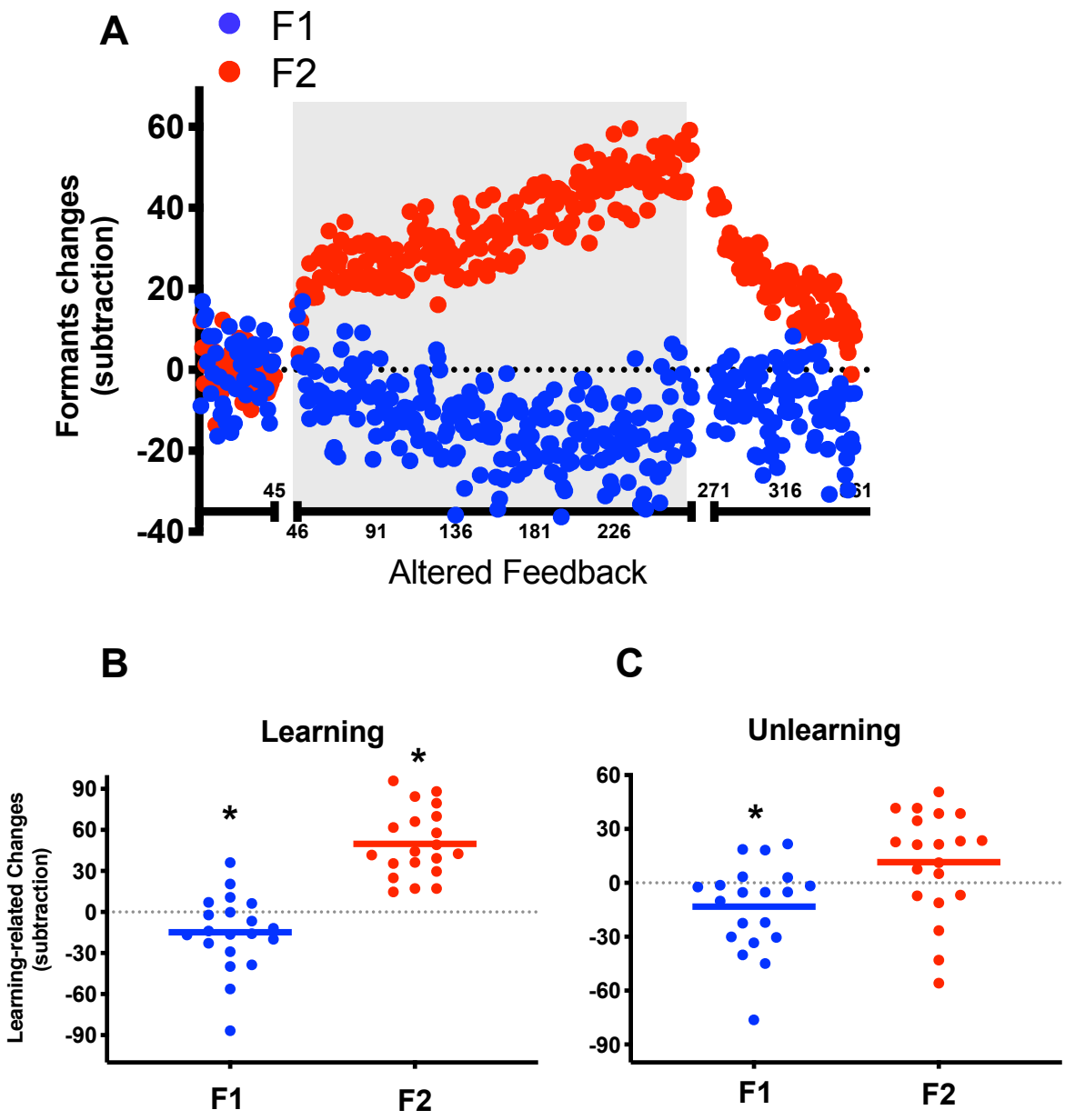


Figure S2 Experiment 2: F2 shifted downwards by 110 mel.

See legend to Fig. S1 for details.

Appendix 2: Study 2 results

Following the method used in Study 1 reported in Chapter 3, the F1 and F2 frequency data for utterances produced during the experiments in Study 2 were first normalised to Z-scores and then statistically analysed. Here, I present analysis of the same data normalised by subtracting the mean baseline frequencies from the frequencies in other phases of the experiment. The pattern of results is the same as that presented in Chapter 4.

Results

Figure S3 shows average F1 frequency produced for each utterance, normalized by subtraction of the mean value from the baseline phase, during the whole adaptation experiment for hand and tongue groups. During the learning phase, a compensatory decrease in F1 production was clearly observed in the hand group following the introduction of F1 upwards shift, while such compensatory change was disrupted by rTMS applied over the tongue representation, and this effect persisted for the remaining course of learning phase. An independent t-test was conducted to compare the learning-related changes across the whole learning phase between groups and showed there was significant difference between hand and tongue group (Figure S3b, t_{38} = 2.15, p= .038, d = 0.68). Critically, rTMS over the tongue representation before the learning phase, resulted in no F1 change from baseline productions, which means the subthreshold rTMS impaired speech motor learning without affecting movement execution (one-sample t-test against zero (baseline average); t_{19} = 1.42, p= .172). At the end of the unlearning phase, F1 frequency of both groups had returned to baseline

(Figure S3b, one-sample t-test against zero (baseline average); tongue group: t_{18} = 0.05, p= .961; hand group: t_{16} = 0.28, p= .784).

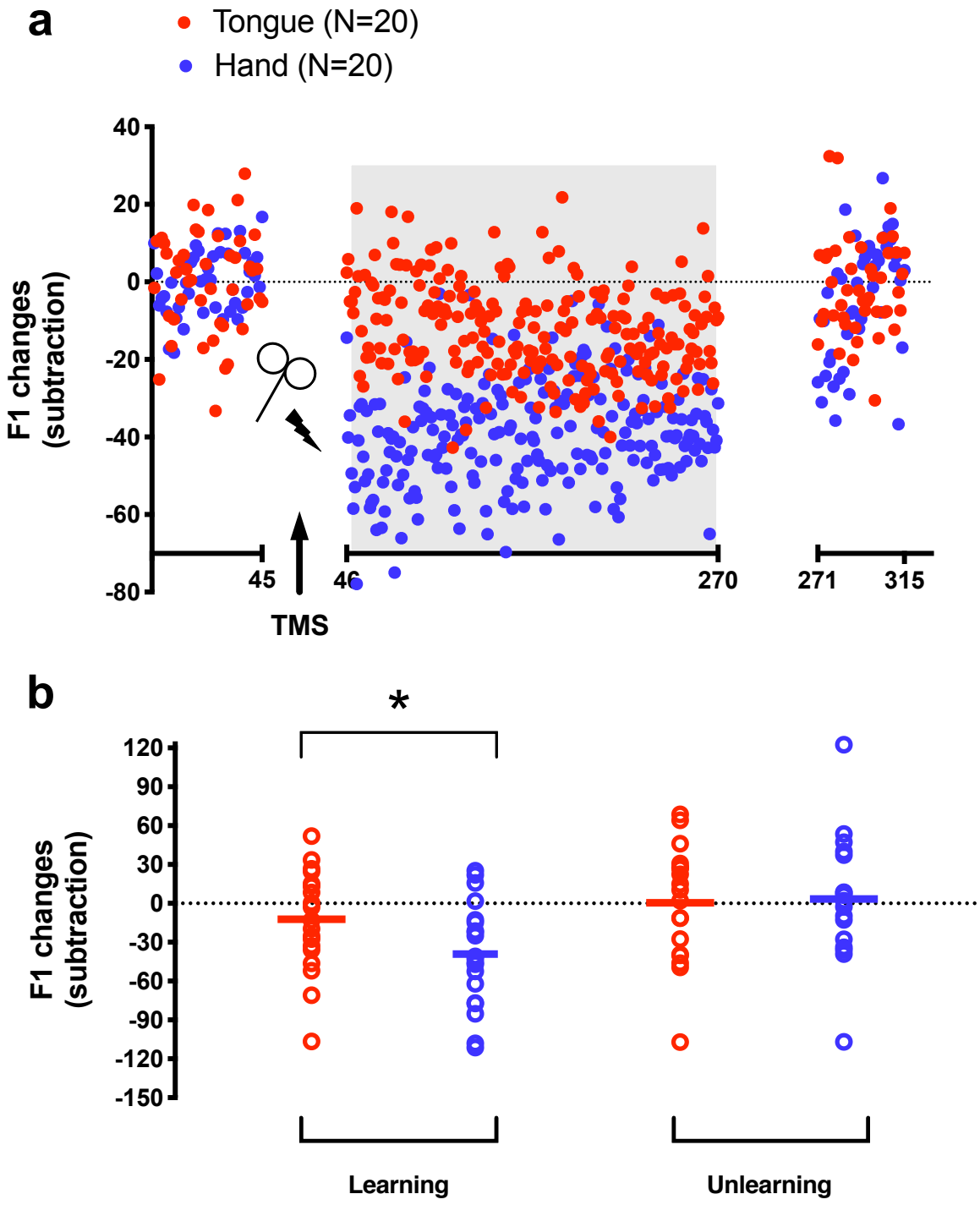


Figure S3 Changes in F1 production (subtraction) related to sensorimotor learning for the hand and tongue groups.

*(a) Changes in F1 production (subtraction) during baseline, learning, and unlearning phases of the experiment. Solid dots indicate the group average for each utterance. The shaded area represents the trials where the altered feedback was applied. The dotted line at zero is the baseline average. (b) Mean F1 changes for the whole learning phase (utterances 46-270) and the end of the unlearning phase (utterances 301-315). Open circles represent individual participants, the solid lines represent the group means Values for hand group are shown in blue and values for tongue group are shown in red. Asterisks (*P < 0.05) indicate reliable between group differences.*

Figure S4 shows average F2 frequency produced for each utterance, normalized by traditional way (subtraction), during the whole experiment for hand and tongue groups. In response to F1 shift up, the hand group showed expected response in F2 by increasing the frequency of their F2 productions (baseline production, t_{19} = 3.22, p= .005, d = 0.72). In contrast, the tongue group exhibited a slight tendency to shift F2 downwards, although F2 changes in tongue group did not significantly differ from zero (baseline production, t_{19} = 1.77, p= .093). Again, we compared F2 changes across the whole learning phase between hand and tongue group. A significant difference between these two groups was found in F2 learning (Figure S4b, t_{38} = 3.31, p = .002, d = 1.59). These results confirm that as above, the rTMS over the tongue representation interferes with adaptive responses but does not impair or alter speech production more generally. By the end of the unlearning phase, F2 frequency of production in both groups was at baseline values (one-sample t-test against zero (baseline average); tongue group: t_{16} = 1.60, p= .126; hand group: t_{18} = 1.12, p= .278).

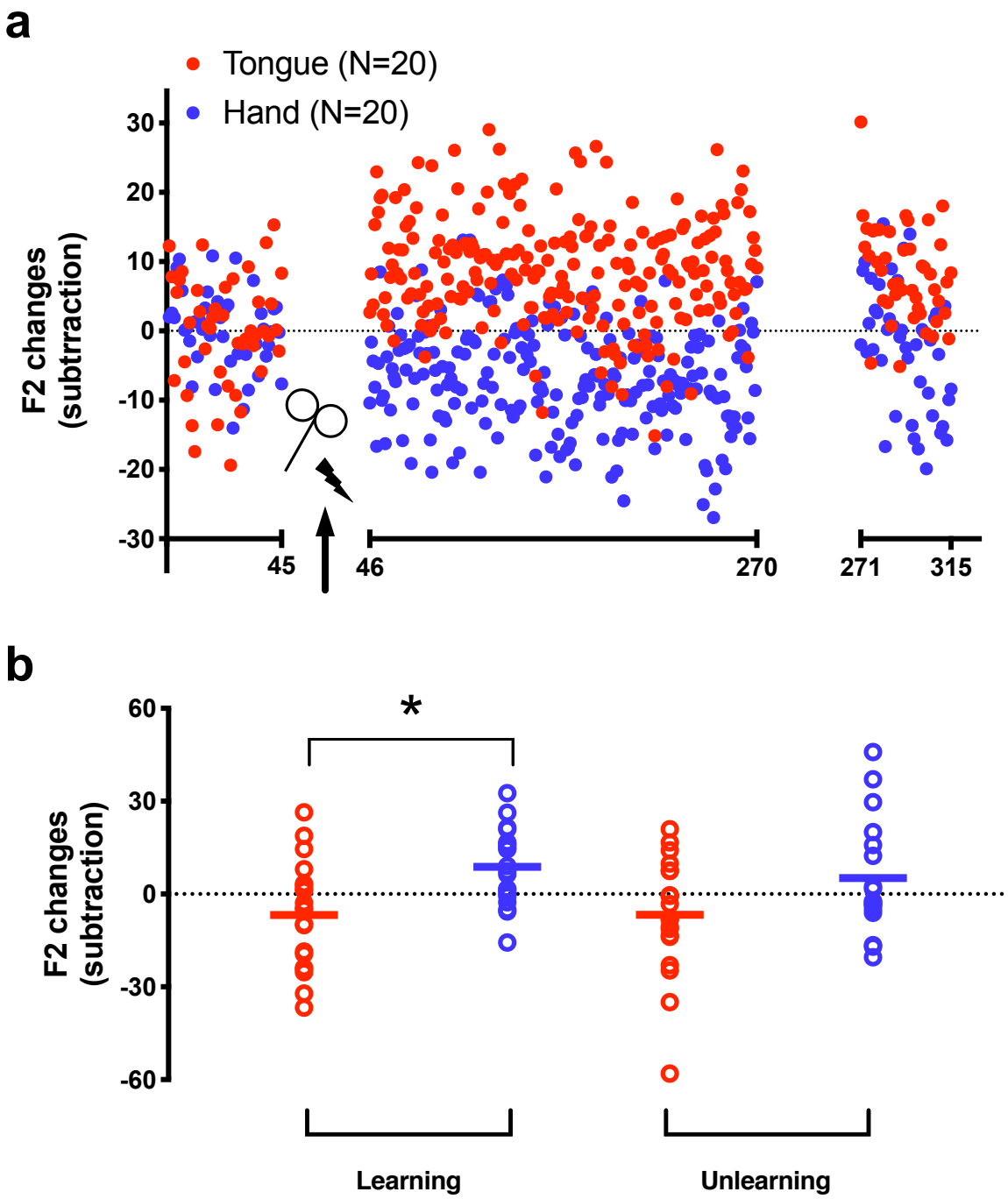


Figure S4 Changes in F2 production (subtraction) related to sensorimotor learning for hand and tongue groups.

(a) Changes in F2 production (subtraction) during baseline, learning, and unlearning phases of the experiment. (b) Mean F2 for the whole learning phase (utterances 46-270) and the end of the unlearning phase (utterances 301-315). See legend to Fig. S3 for details.

References

Ackermann, H., Mathiak, K., & Riecker, A. (2007). The contribution of the cerebellum to speech production and speech perception: Clinical and functional imaging data. *Cerebellum*. https://doi.org/10.1080/14734220701266742

Adank, P., Kennedy-Higgins, D., Maegherman, G., Hannah, R., & Nuttall, H. E. (2018). Effects of coil orientation on motor evoked potentials from orbicularis oris. *Frontiers in Neuroscience*. https://doi.org/10.3389/fnins.2018.00683

Ahmed, M., Mällo, T., Leppänen, P. H. T., Hämäläinen, J., Äyräväinen, L., Ruusuvirta, T., & Astikainen, P. (2011). Mismatch brain response to speech sound changes in rats. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2011.00283

Albouy, P., Benjamin, L., Morillon, B., & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*, *367*(6481), 1043–1047. https://doi.org/10.1126/science.aaz3468

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNM) elicited by sound changes. *Ear and Hearing*. https://doi.org/10.1097/00003446-199502000-00004

Arai, N., Okabe, S., Furubayashi, T., Terao, Y., Yuasa, K., & Ugawa, Y. (2005). Comparison between short train, monophasic and biphasic repetitive transcranial magnetic stimulation (rTMS) of the human motor cortex. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2004.09.020

Assaneo, M. F., Rimmele, J. M., Orpella, J., Ripollés, P., de Diego-Balaguer, R., & Poeppel, D. (2019). The Lateralization of Speech-Brain Coupling Is Differentially Modulated by Intrinsic Auditory and Top-Down Mechanisms. *Frontiers in Integrative Neuroscience*, *13*. https://doi.org/10.3389/fnint.2019.00028

Badin, P., Boë, L. J., Sawallis, T. R., & Schwartz, J. L. (2014). Keep the lips to free the larynx: Comments on de Boer's articulatory model (2010). *Journal of Phonetics*. https://doi.org/10.1016/j.wocn.2014.07.002

Barker, A. T., Jalinous, R., & Freeston, I. L. (1985). NON-INVASIVE MAGNETIC STIMULATION OF HUMAN MOTOR CORTEX. *The Lancet*. https://doi.org/10.1016/S0140-6736(85)92413-4

Bates, J. F., & Goldman-Rakic, P. S. (1993). Prefrontal connections of medial motor areas in the rhesus monkey. *Journal of Comparative Neurology*. https://doi.org/10.1002/cne.903360205

Bäumer, T., Lange, R., Liepert, J., Weiller, C., Siebner, H. R., Rothwell, J. C., & Münchau, A. (2003). Repeated premotor rTMS leads to cumulative plastic changes of motor cortex excitability in humans. *NeuroImage*. https://doi.org/10.1016/S1053-8119(03)00310-0

Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*. https://doi.org/10.1186/1471-2202-12-54

Belyk, M., & Brown, S. (2017). The origins of the vocal brain in humans. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2017.03.014

Bestmann, S., Baudewig, J., Siebner, H. R., Rothwell, J. C., & Frahm, J. (2004). Functional MRI of the immediate impact of transcranial magnetic stimulation on cortical and subcortical motor circuits. *European Journal of Neuroscience*. https://doi.org/10.1111/j.1460-9568.2004.03277.x

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhp055

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.17-01-00353.1997

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*(3), 389–395. https://doi.org/10.1038/nn1409

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2006.04.173

Bohning, D. E., Shastri, A., McConnell, K. A., Nahas, Z., Lorberbaum, J. P., Roberts, D. R., … George, M. S. (1999). A combined TMS/fMRI study of intensity-dependent TMS over motor cortex. *Biological Psychiatry*. https://doi.org/10.1016/S0006-3223(98)00368-0

Bohning, D. E., Shastri, A., Nahas, Z., Lorberbaum, J. P., Andersen, S. W., Dannels, W. R., … George, M. S. (1998). Echoplanar BOLD fMRI of brain activation induced by concurrent transcranial magnetic stimulation. *Investigative Radiology*. https://doi.org/10.1097/00004424-199806000-00004

Bourguignon, N. J., Baum, S. R., & Shiller, D. M. (2014). Lexical-perceptual integration influences sensorimotor adaptation in speech. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2014.00208

Brasil-Neto, J. P., Cohen, L. G., Panizza, M., Nilsson, J., Roth, B. J., & Hallett, M. (1992). Optimal focal transcranial magnetic activation of the human motor cortex: Effects of coil orientation, shape of the induced current pulse, and stimulus intensity. *Journal of Clinical Neurophysiology*. https://doi.org/10.1097/00004691-199201000-00014

Browman, C. P., & Goldstein, L. (1990a). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*. https://doi.org/10.1016/s0095-4470(19)30376-6

Browman, C. P., & Goldstein, L. (1990b). Representation and reality: physical systems and phonological structure. *Journal of Phonetics*. https://doi.org/10.1016/s0095-4470(19)30382-1

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*. https://doi.org/10.1016/S0364-0213(01)00048-9

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*. https://doi.org/10.1016/j.neuron.2005.09.029

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.423073

Bushara, K. O., Wheat, J. M., Khan, A., Mock, B. J., Turski, P. A., Sorenson, J., & Brooks, B. R. (2001). Multiple tactile maps in the human cerebellum. *NeuroReport*. https://doi.org/10.1097/00001756-200108080-00039

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2008). A system for online dynamic perturbation of formant trajectories and results from perturbations of the mandarin triphthong /iau. *Proceedings of ISSP 2008 - 8th International Seminar on Speech Production*.

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.3479539

Caramazza, A. (1997). How Many Levels of Processing Are There in Lexical Access? *Cognitive Neuropsychology*. https://doi.org/10.1080/026432997381664

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1216827110

Chen, C. M. A., Mathalon, D. H., Roach, B. J., Cavus, I., Spencer, D. D., & Ford, J. M. (2011). The corollary discharge in humans is related to synchronous neural oscillations. *Journal of*

Cognitive Neuroscience. https://doi.org/10.1162/jocn.2010.21589

Chen, R., Classen, J., Gerloff, C., Celnik, P., Wassermann, E. M., Hallett, M., & Cohen, L. G. (1997). Depression of motor cortex excitability by low-frequency transcranial magnetic stimulation. *Neurology*, *48*(5), 1398–1403. https://doi.org/10.1212/WNL.48.5.1398

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*. https://doi.org/10.1038/1561

Cheour, M. (2002). Sleep sounds learned by sleeping newborns. *Nature*.

Cheour, Marie, Korpilahti, P., Martynova, O., & Lang, A.-H. (2001). Mismatch Negativity and Late Discriminative Negativity in Investigating Speech Perception and Learning in Children and Infants. *Audiology and Neuro-Otology*, *6*(1), 2–11. https://doi.org/10.1159/000046804

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*. https://doi.org/10.3758/BF03211886

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*. https://doi.org/10.1016/j.cortex.2011.05.017

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The Motor Somatotopy of Speech Perception. *Current Biology*, *19*(5), 381–385. https://doi.org/10.1016/j.cub.2009.01.017

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S.-E. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science*, *21*(2), e12521. https://doi.org/10.1111/desc.12521

Davare, M., Lemon, R., & Olivier, E. (2008). Selective modulation of interactions between ventral premotor cortex and primary motor cortex during precision grasping in humans. *Journal of Physiology*. https://doi.org/10.1113/jphysiol.2008.152603

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1–2), 132–147. https://doi.org/10.1016/j.heares.2007.01.014

de Beukelaar, T. T., Alaerts, K., Swinnen, S. P., & Wenderoth, N. (2016). Motor facilitation during action observation: The role of M1 and PMv in grasp predictions. *Cortex*. https://doi.org/10.1016/j.cortex.2015.11.009

Della-Maggiore, V. (2004). Stimulation of the Posterior Parietal Cortex Interferes with Arm Trajectory Adjustments during the Learning of New Dynamics. *Journal of Neuroscience*, *24*(44), 9971–9976. https://doi.org/10.1523/JNEUROSCI.2833-04.2004

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Deng, Z. De, Lisanby, S. H., & Peterchev, A. V. (2013). Electric field depth-focality tradeoff in transcranial magnetic stimulation: Simulation comparison of 50 coil designs. *Brain Stimulation*. https://doi.org/10.1016/j.brs.2012.02.005

Devlin, J. T., & Aydelott, J. (2009). Speech Perception: Motoric Contributions versus the Motor Theory. *Current Biology*. https://doi.org/10.1016/j.cub.2009.01.005

Di Lazzaro, V., Restuccia, D., Oliviero, A., Profice, P., Ferrara, L., Insola, A., … Rothwell, J. C. (1998). Effects of voluntary contraction on descending volleys evoked by transcranial stimulation in conscious humans. *Journal of Physiology*. https://doi.org/10.1111/j.1469-7793.1998.625bq.x

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2017.02.011

Dubach, P., Guggisberg, A. G., Rösler, K. M., Hess, C. W., & Mathis, J. (2004). Significance of coil orientation for motor evoked potentials from nasalis muscle elicited by transcranial magnetic stimulation. *Clinical Neurophysiology*.

https://doi.org/10.1016/j.clinph.2003.11.033

Eichert, N., Papp, D., Mars, R. B., & Watkins, K. E. (2020). Mapping Human Laryngeal Motor Cortex during Vocalization. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhaa182

Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.386580

Eng, N., Obler, L. K., Harris, K. S., & Abramson, A. S. (1996). Tone perception deficits in Chinese-speaking Broca's aphasics. *Aphasiology*, *10*(6), 649–656. https://doi.org/10.1080/02687039608248441

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., … Wildgruber, D. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *NeuroImage*, *30*(2), 580–587. https://doi.org/10.1016/j.neuroimage.2005.09.059

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1995.73.6.2608

Fadiga, Luciano, Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*. https://doi.org/10.1046/j.0953-816x.2001.01874.x

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., & Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clinical Neurophysiology*. https://doi.org/10.1016/S1388-2457(99)00131-5

Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., & Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nature Human Behaviour*, *3*(4), 393–405. https://doi.org/10.1038/s41562-019-0548-z

Floegel, M., Fuchs, S., & Kell, C. A. (2020). Differential contributions of the two cerebral hemispheres to temporal and spectral speech feedback control. *Nature Communications*, *11*(1), 2839. https://doi.org/10.1038/s41467-020-16743-2

Fowler, C. A., & Dekle, D. J. (1991). Listening With Eye and Hand: Cross-Modal Contributions to Speech Perception. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.17.3.816

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03193857

Galea, J. M., Vazquez, A., Pasricha, N., Orban de Xivry, J.-J., & Celnik, P. (2011). Dissociating the Roles of the Cerebellum and Motor Cortex during Adaptive Learning: The Motor Cortex Retains What the Cerebellum Learns. *Cerebral Cortex*, *21*(8), 1761–1770. https://doi.org/10.1093/cercor/bhq246

Gandour, Jack, Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., … Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language*, *84*(3), 318–336. https://doi.org/10.1016/S0093-934X(02)00505-9

Gandour, Jack, Ponglorpisit, S., & Dardarananda, R. (1992). Tonal disturbances in Thai after brain damage. *Journal of Neurolinguistics*, *7*(1–2), 133–145. https://doi.org/10.1016/0911-6044(92)90015-O

Gandour, Jack, Wong, D., Hsieh, L., Weinzapfel, B., Lancker, D. Van, & Hutchins, G. D. (2000). A Crosslinguistic PET Study of Tone Perception. *Journal of Cognitive Neuroscience*, *12*(1), 207–222. https://doi.org/10.1162/089892900561841

Gandour, Jack, Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *NeuroReport*. https://doi.org/10.1097/00001756-199806220-00038

Gandour, Jackson, Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., … Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2004.06.004

Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D., & Cutler, A. (2011). Slips of the

tongue in the London-Lund corpus of spontaneous conversation. In *Slips of the Tongue and Language Production*. https://doi.org/10.1515/9783110828306.251

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2008.11.029

Giard, M. -H, Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain Generators Implicated in the Processing of Auditory Stimulus Deviance: A Topographic Event-Related Potential Study. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.1990.tb03184.x

Giroud, J., Trébuchon, A., Schön, D., Marquis, P., Liegeois-Chauvel, C., Poeppel, D., & Morillon, B. (2020). Asymmetric sampling in human auditory cortex reveals spectral processing hierarchy. *PLOS Biology*, *18*(3), e3000207. https://doi.org/10.1371/journal.pbio.3000207

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage*, *55*(3), 1324–1338. https://doi.org/10.1016/j.neuroimage.2010.12.065

Grabski, K., Lamalle, L., Vilain, C., Schwartz, J.-L., Vallée, N., Tropres, I., … Sato, M. (2012). Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Human Brain Mapping*, *33*(10), 2306–2321. https://doi.org/10.1002/hbm.21363

Grodd, W., Hülsmann, E., & Ackermann, H. (2005). Functional MRI localizing in the cerebellum. *Neurosurgery Clinics of North America*. https://doi.org/10.1016/j.nec.2004.07.008

Groppa, S., Oliviero, A., Eisen, A., Quartarone, A., Cohen, L. G., Mall, V., … Siebner, H. R. (2012). A practical guide to diagnostic transcranial magnetic stimulation: Report of an IFCN committee. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2012.01.010

Guthrie, D., & Buchwald, J. S. (1991). Significance Testing of Difference Potentials. *Psychophysiology*, *28*(2), 240–244. https://doi.org/10.1111/j.1469-8986.1991.tb00417.x

Hadipour-Niktarash, A., Lee, C. K., Desmond, J. E., & Shadmehr, R. (2007). Impairment of Retention But Not Acquisition of a Visuomotor Skill Through Time-Dependent Disruption of Primary Motor Cortex. *Journal of Neuroscience*, *27*(49), 13413–13419. https://doi.org/10.1523/JNEUROSCI.2570-07.2007

Hari, R., Hämäläinen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, K., Salminen, J., … Sams, M. (1984). Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: Neuromagnetic recordings in man. *Neuroscience Letters*. https://doi.org/10.1016/0304-3940(84)90474-9

Hashimoto, Y., & Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human Brain Mapping*. https://doi.org/10.1002/hbm.10119

Hertrich, I., Dietrich, S., & Ackermann, H. (2016). The role of the supplementary motor area for speech and language processing. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2016.06.030

Hickok, G. (2009). The functional neuroanatomy of language. *Physics of Life Reviews*. https://doi.org/10.1016/j.plrev.2009.06.001

Hickok, G. (2010). The role of mirror neurons in speech perception and action word semantics. *Language and Cognitive Processes*, *25*(6), 749–776. https://doi.org/10.1080/01690961003595572

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*. https://doi.org/10.1162/089892903322307393

HICKOK, G., OKADA, K., BARR, W., PA, J., ROGALSKY, C., DONNELLY, K., … GRANT, A. (2008). Bilateral capacity for speech sound processing in auditory comprehension: Evidence from Wada procedures. *Brain and Language*, *107*(3), 179–184. https://doi.org/10.1016/j.bandl.2008.09.006

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(00)01463-7

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*. https://doi.org/10.1016/j.cognition.2003.10.011

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Higgins, C. M., & Hodge, M. M. (2002). Vowel area and intelligibility in children with and without dysarthria. *Journal of Medical Speech-Language Pathology*.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*. https://doi.org/10.1126/science.279.5354.1213

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*. https://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*. https://doi.org/10.3389/fnhum.2011.00082

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/089892902760807140

Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional Heterogeneity of Inferior Frontal Gyrus Is Shaped by Linguistic Experience. *Brain and Language*, *76*(3), 227–252. https://doi.org/10.1006/brln.2000.2382

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. *NeuroReport*. https://doi.org/10.1097/00001756-200106130-00046

Humphries, Colin, Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.2006.18.4.665

Humphries, Colin, Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human Brain Mapping*. https://doi.org/10.1002/hbm.20148

Hunter, T., Sacco, P., Nitsche, M. A., & Turner, D. L. (2009). Modulation of internal model formation during force field-induced motor learning by anodal transcranial direct current stimulation of primary motor cortex. *The Journal of Physiology*, *587*(12), 2949–2961. https://doi.org/10.1113/jphysiol.2009.169284

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., … Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0303760101

Jacobsen, T., & Schröger, E. (2003). Measuring duration mismatch negativity. *Clinical Neurophysiology*, *114*(6), 1133–1143. https://doi.org/10.1016/S1388-2457(03)00043-9

JACOBSEN, T., & SCHRÖGER, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology*, *38*(4), S0048577201000993. https://doi.org/10.1017/S0048577201000993

Jamison, H. L., Watkins, K. E., Bishop, D. V. M., & Matthews, P. M. (2006). Hemispheric Specialization for Processing Auditory Nonspeech Stimuli. *Cerebral Cortex*, *16*(9), 1266–1275. https://doi.org/10.1093/cercor/bhj068

Jemel, B., Achenbach, C., Müller, B. W., Röpcke, B., & Oades, R. D. (2002). Mismatch negativity results from bilateral asymmetric dipole sources in the frontal and temporal lobes. *Brain Topography*. https://doi.org/10.1023/A:1019944805499

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*.

https://doi.org/10.1093/brain/123.1.155

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feeback during phonation: Studies of Mandarin tone production. *Journal of Phonetics*. https://doi.org/10.1006/jpho.2001.0160

Jung, J. Y., Bungert, A., Bowtell, R., & Jackson, S. R. (2020). Modulating Brain Networks With Transcranial Magnetic Stimulation Over the Primary Motor Cortex: A Concurrent TMS/fMRI Study. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2020.00031

Kaas, J. H., Hackett, T. A., & Tramo, M. J. (1999). Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology*. https://doi.org/10.1016/S0959-4388(99)80022-1

Kasai, T., Kawai, S., Kawanishi, M., & Yahagi, S. (1997). Evidence for facilitation of motor evoked potentials (MEPs) induced by motor imagery. *Brain Research*. https://doi.org/10.1016/S0006-8993(96)01101-8

Kiers, L., Fernando, B., & Tomkins, D. (1997). Facilitatory effect of thinking about movement on magnetic motor-evoked potentials. *Electroencephalography and Clinical Neurophysiology - Electromyography and Motor Control*. https://doi.org/10.1016/S0921-884X(97)00027-1

Kim, H., Hasegawa-Johnson, M., & Perlman, A. (2011). Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica*. https://doi.org/10.1159/000318881

Kim, K. S., Daliri, A., Flanagan, J. R., & Max, L. (2020). Dissociated development of speech and limb sensorimotor learning in stuttering: speech auditory-motor learning is impaired in both children and adults who stutter. *BioRxiv*.

Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A Cross-Linguistic PET Study of Tone Perception in Mandarin Chinese and English Speakers. *NeuroImage*, *13*(4), 646–653. https://doi.org/10.1006/nimg.2000.0738

Krakauer, J. W., Pine, Z. M., Ghilardi, M. F., & Ghez, C. (2000). Learning of visuomotor transformations for vectorial planning of reaching trajectories. *Journal of Neuroscience*.

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, *74*(1), 1–19. https://doi.org/10.1016/j.biopsycho.2006.06.001

Lambon Ralph, M. A., Lowe, C., & Rogers, T. T. (2007). Neural basis of category-specific semantic deficits for living things: Evidence from semantic dementia, HSVE and a neural network model. *Brain*. https://doi.org/10.1093/brain/awm025

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. https://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. https://doi.org/10.1523/JNEUROSCI.0108-14.2014

Lametti, D. R., Smith, H. J., Freidin, P. F., & Watkins, K. E. (2018). Cortico-cerebellar Networks Drive Sensorimotor Learning in Speech. *Journal of Cognitive Neuroscience*, *30*(4), 540–551. https://doi.org/10.1162/jocn_a_01216

Lametti, D. R., Smith, H. J., Watkins, K. E., & Shiller, D. M. (2018). Robust Sensorimotor Learning during Variable Sentence-Level Speech. *Current Biology*, *28*(19), 3106-3113.e2. https://doi.org/10.1016/j.cub.2018.07.030

Lane, H. (1965). The motor theory of speech perception: A critical review. *Psychological Review*. https://doi.org/10.1037/h0021986

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(99)01319-4

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X99001776

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*. https://doi.org/10.1037/h0020279

Liberman, Alvin M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of

consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied*. https://doi.org/10.1037/h0093673

Liberman, Alvin 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*. https://doi.org/10.1037/h0044417

Liberman, Alvin M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*. https://doi.org/10.1016/0010-0277(85)90021-6

Liberman, Alvin M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187–196. https://doi.org/10.1016/S1364-6613(00)01471-6

Liebenthal, E., Ellingson, M. L., Spanaki, M. V., Prieto, T. E., Ropella, K. M., & Binder, J. R. (2003). Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *NeuroImage*. https://doi.org/10.1016/S1053-8119(03)00228-3

Liebenthal, E., & Möttönen, R. (2018). An interactive model of auditory-motor speech perception. *Brain and Language*, *187*, 33–40. https://doi.org/10.1016/j.bandl.2017.12.004

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1898623

Lu, M. -T, Preston, J. B., & Strick, P. L. (1994). Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. *Journal of Comparative Neurology*. https://doi.org/10.1002/cne.903410308

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of Motor Learning in Treatment of Motor Speech Disorders. *American Journal of Speech-Language Pathology*, *17*(3), 277–298. https://doi.org/10.1044/1058-0360(2008/025)

MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, *127*(2), 1059–1068. https://doi.org/10.1121/1.3278606

MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). Probing the independence of formant control using altered auditory feedback. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.3531932

MacNeilage, P. F., Rootes, T. P., & Chase, R. A. (1967). Speech production and perception in a patient with severe impairment of somesthetic perception and motor control. *Journal of Speech and Hearing Research*. https://doi.org/10.1044/jshr.1003.449

Maeda, F., Keenan, J. P., Tormos, J. M., Topka, H., & Pascual-Leone, A. (2000). Interindividual variability of the modulatory effects of repetitive transcranial magnetic stimulation on cortical excitability. *Experimental Brain Research*. https://doi.org/10.1007/s002210000432

Manni, E., & Petrosini, L. (2004). A century of cerebellar somatotopy: A debated representation. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn1347

Marco-Pallarés, J., Grau, C., & Ruffini, G. (2005). Combined ICA-LORETA analysis of mismatch negativity. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2004.11.028

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (2004). Unstable or Insufficiently Activated Internal Models and Feedback-Biased Motor Control as Sources of Dysfluency: A Theoretical Model of Stuttering. *Contemporary Issues in Communication Science and Disorders*. https://doi.org/10.1044/cicsd_31_s_105

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., … Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.1993.5.4.467

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The Essential Role of Premotor Cortex in Speech Perception. *Current Biology*, *17*(19), 1692–1696. https://doi.org/10.1016/j.cub.2007.08.064

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain*

Mapping, *17*(2), 73–88. https://doi.org/10.1002/hbm.10042

Mills, K. R., Boniface, S. J., & Schubert, M. (1992). Magnetic brain stimulation with a double coil: the importance of coil orientation. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials*. https://doi.org/10.1016/0168-5597(92)90096-T

Mitsuya, T., MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). A cross-language study of compensation in response to real-time formant perturbation. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.3643826

Mitsuya, T., Munhall, K. G., & Purcell, D. W. (2017). Modulation of auditory-motor learning in response to formant perturbation as a function of delayed auditory feedback. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4981139

Mitsuya, T., & Purcell, D. W. (2016). Occlusion effect on compensatory formant production and voice amplitude in response to real-time perturbation. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4968539

Mitsuya, T., Samson, F., Ménard, L., & Munhall, K. G. (2013). Language dependent vowel representation in speech production. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4795786

Molholm, S., Martinez, A., Ritter, W., Javitt, D. C., & Foxe, J. J. (2005). The neural circuitry of pre-attentive auditory change-detection: An fMRI study of pitch and duration mismatch negativity generators. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhh155

Mollaei, F., Shiller, D. M., & Gracco, V. L. (2013). Sensorimotor adaptation of speech in Parkinson's disease. *Movement Disorders*. https://doi.org/10.1002/mds.25588

Morton, S. M., & Bastian, A. J. (2006). Cerebellar contributions to locomotor adaptations during splitbelt treadmill walking. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.2622-06.2006

Mottolese, C., Richard, N., Harquel, S., Szathmari, A., Sirigu, A., & Desmurget, M. (2013). Mapping motor representations in the human cerebellum. *Brain*. https://doi.org/10.1093/brain/aws186

Möttönen, R., Dutton, R., & Watkins, K. E. (2013). Auditory-motor processing of speech sounds. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhs110

Möttönen, R., Rogers, J., & Watkins, K. E. (2014). Stimulating the Lip Motor Cortex with Transcranial Magnetic Stimulation. *Journal of Visualized Experiments*, (88). https://doi.org/10.3791/51665

Mottonen, R., van de Ven, G. M., & Watkins, K. E. (2014). Attention Fine-Tunes Auditory-Motor Processing of Speech Sounds. *Journal of Neuroscience*, *34*(11), 4064–4069. https://doi.org/10.1523/JNEUROSCI.2214-13.2014

Mottonen, R., & Watkins, K. E. (2009). Motor Representations of Articulators Contribute to Categorical Perception of Speech Sounds. *Journal of Neuroscience*, *29*(31), 9819–9825. https://doi.org/10.1523/JNEUROSCI.6018-08.2009

Muellbacher, W., Ziemann, U., Boroojerdi, B., & Hallett, M. (2000). Effects of low-frequency transcranial magnetic stimulation on motor excitability and basic motor behavior. *Clinical Neurophysiology*. https://doi.org/10.1016/S1388-2457(00)00284-4

Munhall, K. G., MacDonald, E. N., Byrne, S. K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.3035829

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, Risto. (2008). Mismatch negativity (MMN) as an index of central auditory system plasticity. *International Journal of Audiology*, *47*(sup2), S16–S20. https://doi.org/10.1080/14992020802340116

Näätänen, Risto, Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., … Alho,

K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*(6615), 432–434. https://doi.org/10.1038/385432a0

Näätänen, Risto, & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Näätänen, Risto, Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences*, *24*(5), 283–288. https://doi.org/10.1016/S0166-2236(00)01790-2

Näätänen, Risto, & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.125.6.826

Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*. https://doi.org/10.1038/nn.2193

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? evidence from speech production. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.2137-13.2013

Noppeney, U., Patterson, K., Tyler, L. K., Moss, H., Stamatakis, E. A., Bright, P., … Price, C. J. (2007). Temporal lobe lesions and semantic impairment: A comparison of herpes simplex virus encephalitis and semantic dementia. *Brain*. https://doi.org/10.1093/brain/awl344

Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1290-08.2008

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*. https://doi.org/10.1016/j.bandl.2006.04.006

Orban de Xivry, J.-J., Criscimagna-Hemminger, S. E., & Shadmehr, R. (2011). Contributions of the Motor Cortex to Adaptive Control of Reaching Depend on the Perturbation Schedule. *Cerebral Cortex*, *21*(7), 1475–1484. https://doi.org/10.1093/cercor/bhq192

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*. https://doi.org/10.1016/j.neuropsychologia.2007.06.024

Packman, A., & Attanasio, J. S. (2017). Theoretical issues in stuttering. In *Theoretical Issues in Stuttering: Second Edition*. https://doi.org/10.4324/9781315636559

Panouillères, M. T. N., & Möttönen, R. (2018). Decline of auditory-motor speech processing in older adults with hearing loss. *Neurobiology of Aging*, *72*, 89–97. https://doi.org/10.1016/j.neurobiolaging.2018.07.013

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Current Biology*. https://doi.org/10.1016/j.cub.2015.04.049

Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R. B. (2017). Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3363-16.2017

Pascual-Leone, A., Grafman, J., & Hallett, M. (1994). Modulation of cortical motor output maps during development of implicit and explicit knowledge. *Science*. https://doi.org/10.1126/science.8122113

Pascual-leone, A., Valls-solé, J., Wassermann, E. M., & Hallett, M. (1994). Responses to rapid-rate transcranial magnetic stimulation of the human motor cortex. *Brain*. https://doi.org/10.1093/brain/117.4.847

PATTON, H. D., & AMASSIAN, V. E. (1954). Single and multiple-unit analysis of cortical stage of pyramidal tract activation. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1954.17.4.345

Paus, T., Petrides, M., Evans, A. C., & Meyer, E. (1993). Role of the human anterior cingulate

cortex in the control of oculomotor, manual, and speech responses: A positron emission tomography study. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1993.70.2.453

Paus, T., Sipila, P. K., & Strafella, A. P. (2001). Synchronization of neuronal activity in the human primary motor cortex by transcranial magnetic stimulation: An EEG study. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.2001.86.4.1983

Paus, Tomáš, Jech, R., Thompson, C. J., Comeau, R., Peters, T., & Evans, A. C. (1997). Transcranial Magnetic Stimulation during Positron Emission Tomography: A New Method for Studying Connectivity of the Human Cerebral Cortex. *The Journal of Neuroscience*, *17*(9), 3178–3184. https://doi.org/10.1523/JNEUROSCI.17-09-03178.1997

Paz, R. (2005). Emerging Patterns of Neuronal Responses in Supplementary and Primary Motor Areas during Sensorimotor Adaptation. *Journal of Neuroscience*, *25*(47), 10941–10951. https://doi.org/10.1523/JNEUROSCI.0164-05.2005

Paz, R., Boraud, T., Natan, C., Bergman, H., & Vaadia, E. (2003). Preparatory activity in motor cortex reflects learning of local visuomotor skills. *Nature Neuroscience*, *6*(8), 882–890. https://doi.org/10.1038/nn1097

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, *25*(5), 382–407. https://doi.org/10.1016/j.jneuroling.2010.02.011

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*. https://doi.org/10.1038/331585a0

Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1993). Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.90.3.878

Picard, N., & Strick, P. L. (1996). Motor areas of the medial wall: A review of their location and functional activation. *Cerebral Cortex*. https://doi.org/10.1093/cercor/6.3.342

Pierce, L. J., Klein, D., Chen, J.-K., Delcenserie, A., & Genesee, F. (2014). Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences*, *111*(48), 17314–17319. https://doi.org/10.1073/pnas.1409411111

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time.” *Speech Communication*. https://doi.org/10.1016/S0167-6393(02)00107-3

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360. https://doi.org/10.1038/nrn2811

Pulvermüller, F., Huss, M., Kherif, F., Del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0509989103

Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966–977. https://doi.org/10.1121/1.2217714

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.2173514

Rauschecker, J. P. (1998). Parallel processing in the auditory cortex of primates. *Audiology and Neuro-Otology*. https://doi.org/10.1159/000013784

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*. https://doi.org/10.1038/nn.2331

Richardson, A. G., Overduin, S. A., Valero-Cabre, A., Padoa-Schioppa, C., Pascual-Leone, A., Bizzi, E., & Press, D. Z. (2006). Disruption of Primary Motor Cortex before Learning

Impairs Memory of Movement Dynamics. *Journal of Neuroscience*, *26*(48), 12466–12470. https://doi.org/10.1523/JNEUROSCI.1139-06.2006

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive Sensing of Periodic and Aperiodic Auditory Patterns. *Trends in Cognitive Sciences*, *22*(10), 870–882. https://doi.org/10.1016/j.tics.2018.08.003

Rinne, T., Alho, K., Ilmoniemi, R. J., Virtanen, J., & Näätänen, R. (2000). Separate time behaviors of the temporal and frontal mismatch negativity sources. *NeuroImage*. https://doi.org/10.1006/nimg.2000.0591

Rinne, Teemu, Degerman, A., & Alho, K. (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: An fMRI study. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2005.01.017

Rochet-Capellan, A., & Ostry, D. J. (2011). Simultaneous acquisition of multiple auditory-motor transformations in speech. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.6020-10.2011

Rochet-Capellan, A., Richer, L., & Ostry, D. J. (2012). Nonhomogeneous transfer reveals specificity in speech motor learning. *Journal of Neurophysiology*, *107*(6), 1711–1717. https://doi.org/10.1152/jn.00773.2011

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*. https://doi.org/10.1038/16056

Romero, J. R., Anschel, D., Sparing, R., Gangitano, M., & Pascual-Leone, A. (2002). Subthreshold low frequency repetitive transcranial magnetic stimulation selectively decreases facilitation in the motor cortex. *Clinical Neurophysiology*. https://doi.org/10.1016/S1388-2457(01)00693-9

Rösler, K. M., Hess, C. W., Heckmann, R., & Ludin, H. P. (1989). Significance of shape and size of the stimulating coil in magnetic stimulation of the human motor cortex. *Neuroscience Letters*. https://doi.org/10.1016/0304-3940(89)90711-8

Sallinen, M., Kaartinen, J., & Lyytinen, H. (1994). Is the appearance of mismatch negativity during stage 2 sleep related to the elicitation of K-complex? *Electroencephalography and Clinical Neurophysiology*. https://doi.org/10.1016/0013-4694(94)90035-3

Saltzman, E. L., & Munhall, K. G. (1989). A Dynamical Approach to Gestural Patterning in Speech Production. *Ecological Psychology*. https://doi.org/10.1207/s15326969eco0104_2

Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and Ventral Pathways for Prosody. *Current Biology*, *25*(23), 3079–3085. https://doi.org/10.1016/j.cub.2015.10.009

Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*. https://doi.org/10.1016/j.bandl.2009.03.002

Saur, D., Kreher, B. W., Schnell, S., Kümmerera, D., Kellmeyera, P., Vrya, M. S., … Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0805234105

Schirmer, A., Tang, S. L., Penney, T. B., Gunter, T. C., & Chen, H. C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/0898929052880057

Schoch, B., Dimitrova, A., Gizewski, E. R., & Timmann, D. (2006). Functional localization in the human cerebellum based on voxelwise statistical analysis: A study of 90 patients. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2005.09.018

Schönwiesner, M., Rübsamen, R., & Von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience*, *22*(6), 1521–1528. https://doi.org/10.1111/j.1460-9568.2005.04315.x

Scott, T. L., Haenchen, L., Daliri, A., Chartove, J., Guenther, F. H., & Perrachione, T. K. (2020). Noninvasive neurostimulation of left ventral motor cortex enhances sensorimotor adaptation in speech production. *Brain and Language*, *209*, 104840.

https://doi.org/10.1016/j.bandl.2020.104840

Shum, M., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2011). Sensorimotor integration for speech motor learning involves the inferior parietal cortex. *European Journal of Neuroscience*, *34*(11), 1817–1822. https://doi.org/10.1111/j.1460-9568.2011.07889.x

Siebner, H. R., Tormos, J. M., Ceballos-Baumann, A. O., Auer, C., Catala, M. D., Conrad, B., & Pascual-Leone, A. (1999). Low-frequency repetitive transcranial magnetic stimulation of the motor cortex in writer's cramp. *Neurology*. https://doi.org/10.1212/wnl.52.3.529

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue : Review of the role of the motor system in speech perception. *Brain and Language*, *164*, 77–105. https://doi.org/10.1016/j.bandl.2016.10.004

Skodda, S., Visser, W., & Schlegel, U. (2011). Vowel articulation in parkinson's disease. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2010.01.009

Smalle, E. H. M., Rogers, J., & Möttönen, R. (2015a). Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhu218

Smalle, E. H. M., Rogers, J., & Möttönen, R. (2015b). Dissociating Contributions of the Motor Cortex to Speech Perception and Response Bias by Using Transcranial Magnetic Stimulation. *Cerebral Cortex*, *25*(10), 3690–3698. https://doi.org/10.1093/cercor/bhu218

Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting Adaptive Processes with Different Timescales Underlie Short-Term Motor Learning. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.0040179

Smith, M. A., & Shadmehr, R. (2005). Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00943.2004

Snider, R., & Eldred, E. (1951). Electro-anatomical studies on cerebro-cerebellar connections in the cat. *Journal of Comparative Neurology*. https://doi.org/10.1002/cne.900950102

Sommer, M., Lang, N., Tergau, F., & Paulus, W. (2002). Neuronal tissue polarization induced by repetitive transcranial magnetic stimulation? *NeuroReport*. https://doi.org/10.1097/00001756-200205070-00015

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A Scale for the Measurement of the Psychological Magnitude Pitch. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1915893

Stokes, R. C., Venezia, J. H., & Hickok, G. (2019). The motor system's [modest] contribution to speech perception. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-019-01580-2

Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2008.08.039

Strafella, A. P., & Paus, T. (2001). Cerebral blood-flow changes induced by paired-pulse transcranial magnetic stimulation of the primary motor cortex. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.2001.85.6.2624

Strens, L. H. A., Oliviero, A., Bloem, B. R., Gerschlager, W., Rothwell, J. C., & Brown, P. (2002). The effects of subthreshold 1 Hz repetitive TMS on cortico-cortical and interhemispheric coherence. *Clinical Neurophysiology*. https://doi.org/10.1016/S1388-2457(02)00151-7

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1907309

Svirsky, M. A., Teoh, S. W., & Neuburger, H. (2004). Development of language and speech perception in congenitally, profoundly deaf children as a function of age at cochlear implantation. *Audiology and Neuro-Otology*. https://doi.org/10.1159/000078392

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., … Wartenburger, I. (2009). Sensitivity of newborn auditory cortex to the temporal structure of

sounds. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1246-09.2009

Terao, Y., & Ugawa, Y. (2002). Basic mechanisms of TMS. *Journal of Clinical Neurophysiology*. https://doi.org/10.1097/00004691-200208000-00006

Tiitinen, H., Salminen, N. H., Palomäki, K. J., Mäkinen, V. T., Alku, P., & May, P. J. C. (2006). Neuromagnetic recordings reveal the temporal dynamics of auditory spatial processing in the human cortex. *Neuroscience Letters*. https://doi.org/10.1016/j.neulet.2005.11.018

Tong, Y., Francis, A., & Gandour, J. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*. https://doi.org/10.1080/01690960701728261

Tong, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y., … Lowe, M. (2005). Neural circuitry underlying sentence-level linguistic prosody. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2005.06.002

Tourville, J. A., Cai, S., & Guenther, F. (2013). Exploring auditory-motor interactions in normal and disordered speech. *Proceedings of Meetings on Acoustics*. https://doi.org/10.1121/1.4800684

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. https://doi.org/10.1080/01690960903498424

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. https://doi.org/10.1016/j.neuroimage.2007.09.054

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*. https://doi.org/10.1038/nature01710

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*. https://doi.org/10.1111/j.1467-8624.2004.00726.x

Tseng, Y. W., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00266.2007

Van Lancker, D., & Fromkin, V. A. (1973). Hemispheric specialization for pitch and “tone”: Evidence from Thai. *Journal of Phonetics*, *1*(2), 101–109. https://doi.org/10.1016/S0095-4470(19)31414-7

van Vugt, F. T., Near, J., Hennessy, T. J., Doyon, J., & Ostry, D. J. (2020). Early stages of sensorimotor map acquisition: neurochemical signature in primary motor cortex and its relation to functional connectivity. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00285.2020

Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/08989290260045800

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. https://doi.org/10.1121/1.2773966

Villacorta, V., Perkell, J., & Guenther, F. (2005). Relations between speech sensorimotor adaptation and perceptual acuity. *The Journal of the Acoustical Society of America*, *117*(4), 2618–2619. https://doi.org/10.1121/1.4778207

Waldstein, R. S. (1990). Effects of postlingual deafness on speech production: Implications for the role of auditory feedback. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.400107

Wang, Yan, Isoda, M., Matsuzaka, Y., Shima, K., & Tanji, J. (2005). Prefrontal cortical cells projecting to the supplementary eye field and presupplementary motor area in the monkey. *Neuroscience Research*. https://doi.org/10.1016/j.neures.2005.05.005

Wang, Yue, Behne, D. M., Jongman, A., & Sereno, J. A. (2004). The role of linguistic

experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716404001213

Wang, Yue, Jongman, A., & Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language*. https://doi.org/10.1006/brln.2001.2474

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*. https://doi.org/10.1016/S0028-3932(02)00316-0

Wiener, S., & Turnbull, R. (2016). Constraints of Tones, Vowels and Consonants on Lexical Selection in Mandarin Chinese. *Language and Speech*. https://doi.org/10.1177/0023830915578000

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of affective and linguistic prosody: {fMRI} studies. *Progress in Brain Research, Vol. 156*. https://doi.org/10.1016/S0079-6123(06)56013-3

Wilson, S. M. (2009). Speech perception when the motor system is compromised. *Trends in Cognitive Sciences*, *13*(8), 329–330. https://doi.org/10.1016/j.tics.2009.06.001

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*. https://doi.org/10.1038/nn1263

Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*. https://doi.org/10.1016/S0006-8993(96)01008-6

WINKLER, I., KUJALA, T., TIITINEN, H., SIVONEN, P., ALKU, P., LEHTOKOSKI, A., … NÄÄTÄNEN, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*(5), S0048577299981908. https://doi.org/10.1017/S0048577299981908

Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E., & Frackowiak, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain*. https://doi.org/10.1093/brain/114.4.1803

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7–8), 1317–1329. https://doi.org/10.1016/S0893-6080(98)00066-5

Wong, P. C. M. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin*. https://doi.org/10.1016/S0361-9230(02)00860-2

Wu, Z. M., Chen, M. L., Wu, X. H., & Li, L. (2014). Interaction between auditory and motor systems in speech perception. *Neuroscience Bulletin*. https://doi.org/10.1007/s12264-013-1428-6

Xu, Y., Larson, C. R., Bauer, J. J., & Hain, T. C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1763952

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*. https://doi.org/10.1080/016909699386202

Yi, X., Fisher, K. M., Lai, M., Mansoor, K., Bicker, R., & Baker, S. N. (2014). Differences between Han Chinese and Caucasians in transcranial magnetic stimulation parameters. *Experimental Brain Research*. https://doi.org/10.1007/s00221-013-3763-2

Yip, M., (2002). Introduction. Tone.Cambridge University Press, Cambridge, England

1–16. Yost, W.A., 2009. Pitch perception. *Atten. Percept. Psychophys*. 71, 1701–1715.

Yiu, E. M.-L., & Fok, A. Y.-Y. (1995). Lexical tone disruption in Cantonese aphasic speakers. *Clinical Linguistics & Phonetics*, *9*(1), 79–92. https://doi.org/10.3109/02699209508985326

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe functiona. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.396834

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex.

Cerebral Cortex. https://doi.org/10.1093/cercor/11.10.946

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(00)01816-7

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.14-04-01908.1994

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*. https://doi.org/10.1126/science.1589767

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1087–1104. https://doi.org/10.1098/rstb.2007.2161

ZATORRE, R. J., & SAMSON, S. (1991). ROLE OF THE RIGHT TEMPORAL NEOCORTEX IN RETENTION OF PITCH IN AUDITORY SHORT-TERM MEMORY. *Brain*, *114*(6), 2403–2417. https://doi.org/10.1093/brain/114.6.2403

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2011.02.054