The corticocerebellar network in speech adaptation: evidence from non-invasive brain stimulation and brain imaging

Qiming Yuan

St John's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity term, 2026

Supervised by Professor Kate E. Watkins and Professor Heidi Johansen-Berg

Acknowledgements

My deepest gratitude goes to my supervisor, Professor Kate Watkins. These have been four wonderful years of learning under your guidance. Your knowledge and experience have shaped me as a researcher in ways I will continue to draw on throughout my academic journey, and your kindness has shaped me just as profoundly as a person. I could not have wished for a better supervisor, and I will always treasure the memories we have shared.

I am also grateful to Professor Heidi Johansen-Berg. Thank you for your steady support throughout my DPhil. Our conversations were always a pleasure, and I will fondly remember the summer and Christmas gatherings at your house.

My sincere thanks go to Professor Kate Nation. Thank you for your warmth and support at St John's College. St John's is the finest college in Oxford to call home, and it would not be the same without you.

I am also grateful to Professor Matt Jones. Thank you for your continued support and mentoring on the British Neuroscience Association (BNA) Scholars Programme.

I am deeply grateful to Professor Guosheng Ding at Beijing Normal University. Thank you for your generous support of our collaborative research and for securing the funding that made it possible.

I would like to thank my two examiners, Professor Charlotte Stagg and Professor Christian Kell, for their insightful discussion and their suggestions on this thesis.

To Birtan and Nilgoun, my lab mates during my first two years — thank you for all your help and companionship. Thank you, Amy and Levi, for being wonderful and cheerful friends to us. Thank you, Dan, for your contributions to our research, and Izara and Hui, for your dedication to our brain stimulation projects.

To Mum and Dad — thank you for your unconditional love and unwavering support. I love you both.

To Grandpa — thank you for your love and belief in me. I am glad I made you proud. I miss you.

Most of all, thank you to my fiancée, Yiran. Life would have been unimaginably different without you, and I cannot picture how I would have made it through this DPhil with anyone else by my side. I love you.

Finally, I am deeply grateful to the Oxford Clarendon Fund and St John's College for the scholarships that supported my DPhil studies, to the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University for the Open Project Grant that supported my research, and to the British Neuroscience Association for supporting me through the BNA Scholars Programme.

Abstract

Auditory feedback is crucial for accurate speech production. During speaking, the brain continuously compares predicted auditory outcomes with incoming feedback, and mismatches trigger corrective adjustments that support speech motor control. This thesis investigates speech adaptation using the altered auditory feedback paradigm, in which speakers hear their own speech in near real time with a perturbation to vowel formant frequency. Across studies, adaptation is measured as systematic changes in vowel production that offset the induced auditory error.

In Study 1, I tested the effects of transcranial direct-current stimulation over left speech motor cortex and right cerebellum to speech adaptation using a preregistered, double-blind, sham-controlled design. Participants produced words while the frequency of the first vowel formant was increased. Ninety participants received 2-mA anodal tDCS over the left speech motor cortex, or right cerebellum, or sham stimulation. All groups showed reliable adaptation, but tDCS did not modulate adaptation contrary to the previous report. This result highlights a lack of robustness in the effects of tDCS as well as the sensitivity of speech adaptation outcomes to individual differences and the need for adequately powered samples and replication when evaluating neuromodulation effects.

Study 2 adapted the altered auditory feedback paradigm for Mandarin Chinese

speakers and refined task parameters for implementation in a subsequent MRI experiment in Study 3. In 76 participants, both up-shifts and down-shifts in the frequency of the first vowel formant produced clear adaptation in speech production in the direction opposite to the perturbation. These effects were limited to changes in the first vowel formant unlike typical findings in English, Mandarin speakers did not show additional systematic changes in the unshifted second vowel formant. This finding is consistent with language-specific constraints on vowel production.

In Study 3, I used functional MRI to record brain activity in 48 native Mandarin speakers producing syllables either under normal or altered feedback (increased frequency of first vowel formant). Adaptation of speech production during altered feedback was accompanied by increased activity in left posterior cerebellum and decreased activity in sensorimotor cortex, consistent with cerebellar forward modelling and refinement of feedforward motor commands, respectively. Resting-state scans acquired before and after speech adaptation further revealed strengthened functional connectivity within networks involving superior temporal cortex bilaterally and within a dorsomedial sensorimotor cortex network, reflecting the consolidation of a new sensorimotor mapping.

In summary, these studies advance our understanding of the neural correlates of speech adaptation to formant perturbations. They highlight: (i) the importance of understanding individual difference in speech adaptation for neuromodulation studies; (ii) the robustness of this formant perturbation cross-linguistically; and (iii) the neural network changes associated with speech motor learning. Together, the results contribute to a clearer understanding of how the human brain maintains flexible and stable speech production in the face of changing sensory input.

Statement of use of Gen AI

I declare that any Gen AI use complies with the University's guidance on Safe and responsible use of generative AI tools, and the divisional guidance on the use of Gen AI in PGR summative assessment. The name of the Gen AI used is ChatGPT 5, published by OpenAI and under the ChatGPT Edu licence provided by the University of Oxford. The Gen AI was used to search for background research, refine code, gain general subject understanding, correct grammar, generate experimental setup figures, and improve language use.

Ethics Statement

The study of Chapter 3 was approved by the Central University Research Ethics Committee (CUREC, reference R85441/RE001) of the University of Oxford in March 2023. All participants provided informed consent before taking part in this study.

The studies of Chapters 4 and 5 were approved by the Human Participants Research Ethics Committee of State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University (ICBIR_D_0084_001) in April 2024. All participants provided informed consent before taking part.

Open Science Statement

For the brain stimulation study presented in Chapter 2:

Data: Some raw and processed data supporting this research are publicly available, while some are subject to restrictions: https://osf.io/yku6h/

Code: All analysis code supporting this research is publicly available: https://osf.io/3s984/ MATERIALS: All study materials supporting this research are publicly available: https://osf.io/u9jxa/

Design: This article reports, for all studies, how the author(s) determined all sample sizes, all data exclusions, all data inclusion and exclusion criteria, and whether inclusion and exclusion criteria were established prior to data analysis.

Pre-registration: The study procedures were pre-registered in a time-stamped, institutional registry prior to the research being conducted: https://osf.io/9bswe The study analysis plans were pre-registered in a time-stamped, institutional registry prior to the research being conducted: https://osf.io/9bswe

For the behavioural study presented in Chapter 3:

Data and code will be publicly available upon completion and publication.

For the brain imaging study presented in Chapters 3 & 4:

Data and code will be publicly available upon completion and publication.

Declaration

This thesis comprises work conducted by Qiming Yuan in fulfilment of the degree of Doctor of Philosophy in Experimental Psychology at the University of Oxford. The contributions of Qiming Yuan and collaborators are as follows:

For the brain stimulation study presented in Chapter 3, conceptualisation was co-led by Qiming Yuan and Prof. Kate E. Watkins. Dr Daniel R. Lametti provided the necessary methodology of the study. Qiming Yuan led the data collection, formal analysis, visualisation, and writing. Izara Williams and Hui Zhu assisted with the data collection. Prof. Kate E. Watkins led the project administration and funding acquisition.

For the behavioural study and MRI study presented in Chapters 4 and 5, conceptualisation was co-led by Qiming Yuan and Prof. Kate E. Watkins. Qiming Yuan led the data collection. Haijun Yao assisted with the data collection. Qiming Yuan led the formal analysis, visualisation, and writing. Prof. Guosheng Ding and Prof. Kate E. Watkins led the project administration and funding acquisition. The project was supported by the Open Research Fund of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University (CNLZD2302).

The brain stimulation study in Chapter 3 was presented at the British Neuroscience Association (BNA) 2025 International Festival of Neuroscience at Liverpool, UK, in April 2025 as a flash talk and poster named *Failure to replicate tDCS effects on speech adaptation.* It was collected in: The International BNA 2025 Festival of Neuroscience. *Brain and Neuroscience Advances.* 2025;*9.* https://doi.org/10.1177/23982128251339668

It was published in *Cortex* in September 2025 as:

Yuan, Q., Lametti, D. R., Williams, I., Zhu, H., & Watkins, K. E. (2025). Failure to replicate enhancement of speech adaptation using tDCS over motor cortex and cerebellum. *Cortex*, *192*, 152–164. https://doi.org/10.1016/j.cortex.2025.09.007

Results from the behavioural and MRI studies in Chapters 4 and 5 were presented at the Society for the Neurobiology of Language Annual Meeting (SNL 2025) at Gallaudet University, Washington DC, USA, as a poster named *Neural mechanisms involved in speech adaptation during altered auditory feedback.* The behavioural study along with the MRI study will be submitted for publication upon completion.

Aims and Structure of Thesis

This thesis is structured into an introduction to speech motor control, an introduction to speech adaptation paradigm used in the three studies presented next, and a general discussion.

Chapter 1 introduces human speech production and how speech motor control plays an important part in maintaining the delivery of speech. I introduce the history and models of speech motor control and discuss the function of auditory feedback, somatosensory feedback, and feedforward control systems and their neural correlates mainly based on the DIVA model of speech motor control.

Chapter 2 introduces the speech adaptation paradigm which uses altered auditory feedback paradigm. I introduce the adaptive response in speech production during altered auditory feedback tasks and focus on tasks in which the frequency of vowel formants is shifted. I also review previous findings of studies examing the neural correlates of speech adaptation using brain imaging and brain stimulation.

Chapter 3 (Study 1) is a pre-registered study investigating the role of the corticocerebellar network in speech adaptation using transcranial direct current stimulation (tDCS). This study aimed to replicate a previous study by Lametti et al. (2018), which found different effects on speech adaptation following stimulation over the left speech motor cortex and right cerebellum. The results of the

current study in a larger sample (90 participants in total) failed to replicate the original findings. We found that speech adaptation was unaffected by anodal tDCS over either left speech motor cortex or right cerebellum. This work demonstrates the inconsistent effects of tDCS on behavioural tasks. Our findings indicate that even larger samples are needed to avoid spurious results arising from individual differences in task performance.

Chapter 4 (Study 2 and 3) includes a behavioural investigation of speech adaptation in Mandarin Chinese speakers. We adapted the altered auditory feedback task for the production of Chinese syllables. The feedback perturbation involved shifting the frequency of the first vowel formant (F1) up or down by 110 Mel. Data from 76 participants were collected. Both F1 up-shift and down-shift resulted in adaptation in the direction opposite to the shift, as seen in English speakers. In contrast with effects seen in English speakers, however, Mandarin speakers did not alter their production of the unshifted second formant (F2). This different pattern of results may be due to the different vowel space observed in Mandarin vowel production compared with English. The results of this behavioural study helped us refine our plans for the subsequent MRI study using this paradigm.

MRI study was used to investigate the task-evoked brain activity during altered auditory feedback. We scanned 48 native Mandarin Chinese speakers while they read the syllable "dē" out loud. Speech was recorded and fed back to them either normally (baseline) or with a 110-Mel increase in the frequency of the first vowel formant (adaptation). Speech production and feedback occurred during short silent intervals between echo-planar volumes. As participants adapted to the altered auditory feedback by reducing the frequency of the first formant in their production,

activity increased in the left posterior cerebellum and decreased in sensorimotor cortex. We interpret these changes to reflect the role of the cerebellum in forward modelling and the establishment of refined feedforward motor commands in the sensorimotor cortex.

Chapter 5 (Study 3) presents the findings from the second part of Study 3. Two 10-minute resting-state (task-free) MRI scans were collected before and after the speech adaptation. Participants were instructed to fixate on a cross at the centre of the screen. We then compared the resting-state brain activity between these two sessions to measure changes in functional connectivity due to acquisition of the adapted state. Resting-state data obtained before and after the task, revealed that the adapted state was characterised by strengthened functional connectivity within the superior temporal (auditory) cortex bilaterally and dorsomedial sensorimotor cortex, consistent with consolidation of a new sensorimotor mapping.

Chapter 6 provides general discussion of the studies and highlights the functional roles of auditory and sensorimotor areas in speech adaptation during altered auditory feedback.

Contents

List of Figures

List of Tables

1

Introduction to speech motor control

1.1 Speech production

Speech is created when air is pushed out from the lungs and passes through the larynx. Inside the larynx are the vocal folds, which can vibrate very quickly as air moves through them. This vibration produces voiced sounds, giving speech its basic sound source. The sound then travels through the throat, mouth, and sometimes the nose, where it is shaped into different speech sounds. By moving the tongue, lips, and jaw, speakers change the shape of these spaces, which changes the quality of the sound we hear. In this way, speech is not produced by a single organ, but by a coordinated system that shapes airflow and vibration into meaningful sounds.

Speech is one of the most complex motor skills performed by the human brain. A

fluent speaker can produce about two to three words per second and approximately 100 muscles in the respiratory, laryngeal and supralaryngeal vocal tract activate to achieve this goal (Duffy, 1995). Each of these muscles is composed of multiple motor units – a motor neuron together with the muscle fibres it innervates – with the number per muscle varying widely across the respiratory, laryngeal and lingual musculature, from a few dozen in the small intrinsic laryngeal muscles to several hundred in the larger respiratory muscles (Kent, 2004). Even on conservative assumptions, this means that producing a single utterance requires the coordinated, time-locked recruitment of many thousands of motor units. Communication using speech is unique to human primates even though the vocal tracts of other primates are considered "speech ready" (Fitch et al., 2016). The lack of speech must reflect differences in the neural control of articulators, which in the human brain involves integration of a complex matrix of auditory and somatosensory information when generating motor commands to the required muscles.

Much of our current understanding of the neural basis of speech production is based on the seminal work of Broca, Wernicke, and Lichtheim (Broca, 1861; Lichtheim, 1885; Wernicke, 1969). Informed by studies of patients with brain damage, they determined a network of left hemisphere brain regions necessary for fluent speech production. Geschwind's model of "speaking a heard word" included primary and secondary auditory areas in posterior temporal cortex, white matter connecting to posterior inferior frontal areas and output via premotor and motor cortex where the articulators are represented (Goodglass & Geschwind, 1976). More modern methods using functional imaging in intact brains confirm the involvement of new regions but extend the early models to include connected subcortical

areas such as the basal ganglia and cerebellum, and other cortical regions (e.g. the medial and dorsolateral frontal cortex and premotor cortex) (Fedorenko et al., 2024; Hickok et al., 2022; Price, 2012).

These early lesion based studies and more recent brain imaging ones focused on the brain regions involved in language, including linguistic processes such as word selection, phonological planning, the assembly of phonemes for final execution of speech production via the movement of articulators. The focus of the work in this thesis is on the final part of this sequence, in which speech movements are executed while monitoring the sensory consequences within auditory or somatosensory cortex. The study of speech motor control has been advanced by computational models derived from a parallel set of studies of more general motor control. Below, I focus on these models of speech motor control.

1.2 Speech motor control

As with other motor systems, speech motor control is a dynamic system consisting of both feedback and feedforward components. The speech motor control system can be divided into a feedback control system, which monitors the sensory output of speech in order to determine the motor command to send, and a feedforward control system, which generates motor commands for a syllable or phoneme from memory without monitoring the ongoing acoustic signal or the somatosensory state.

1.2.1 Models of speech motor control

The motor control approach to understanding speech production originates within the question of how these motor speech organs are controlled to produce the target sound. Several models have been proposed to explain the mechanism of speech motor control (see Parrell et al. (2019a) for a review). Prominent examples include the Directions Into Velocities of Articulators (DIVA) model (Golfinopoulos et al., 2010; Guenther et al., 2006; Tourville & Guenther, 2011), State Feedback Control (SFC) model (Hickok et al., 2011; Houde & Nagarajan, 2011), the Task Dynamics (TD) framework (Saltzman & Kelso, 1987; Saltzman & Munhall, 1989), and the Feedback Control of Tasks in Speech (FACTS) model (Parrell et al., 2019b). Collectively, these approaches share the idea that speech is controlled through the interaction of predictive feedforward control and sensory feedback, but they differ in (i) what is treated as the primary task variable of speech motor control (e.g. acoustic pattern in DIVA vs. constrictions in vocal tract in TD), (ii) how the vocal tract is represented and controlled, and (iii) the computational role assigned to auditory and somatosensory feedback during ongoing production of speech.

The State Feedback Control (SFC) model formalises speech production as control of an internal estimate of the vocal tract state. Because sensory feedback is delayed and noisy, the controller combines "efference copy"-based predictions (see section 1.2.3) with incoming auditory and motor signals to infer the current state, and uses this state estimate to generate corrective motor commands (Hickok et al., 2011; Houde & Nagarajan, 2011).

The Task Dynamics (TD) framework focuses on how invariant linguistic targets

can generate articulatory movements. It treats speech as the coordinated control of articulatory gestures – goal-directed dynamical units defined in task space (e.g. constrictions formed by the lips, tongue tip, or tongue body) (Saltzman & Kelso, 1987; Saltzman & Munhall, 1989). Here, movement patterns arise from coupled dynamical systems that specify both the desired task (such as achieving a constriction degree at a vocal-tract location) and the temporal coordination among gestures. This framework has been influential in explaining coarticulation and the stability of speech movements despite variability at the level of individual articulators, since multiple articulatory configurations can realise the same task-level goal.

The Feedback-Aware Control of Tasks in Speech (FACTS) model can be understood as a synthesis that embeds Task Dynamics (TD) within a state feedback control (SFC) architecture (Kim et al., 2023b; Parrell et al., 2019b). In the FACTS model, the task-level goals of Task Dynamics are maintained, but the system also includes an explicit observer that estimates the current vocal tract state by integrating predictions with auditory and somatosensory feedback to generate a sensory error, enabling principled feedback-based corrections. This combination is particularly useful for linking task-level representations (gestures and constrictions) with control components (state estimation, feedback gains), and for accounting for both compensatory responses to perturbation and longer-term adaptation. The FACTS model is a relatively new model, which has been tested in studies where participants compensate for perturbations to auditory feedback (Parrell et al., 2019b). This work is reviewed here in detail in the next chapter.

1.2.2 The DIVA model

Among these models, the DIVA model is the only one with an explicit, region-by-region neuroanatomical implementation, assigning each computational component to specific cortical, subcortical, and cerebellar substrates and generating falsifiable predictions about where particular computations should occur. The DIVA model was developed by Frank Guenther and colleagues (Golfinopoulos et al., 2010; Guenther et al., 2006; Tourville & Guenther, 2011). DIVA refers to the Directions Into Velocities of Articulators — how the brain transforms desired movement directions in sensory space into the movement velocities of articulators. In the DIVA model, the three subsystems of speech motor control are the (i) auditory feedback control system; (ii) somatosensory feedback control system; and (iii) feedforward control system (Figure 1.1).

Specifically, the auditory feedback control system is responsible for detecting and correcting discrepancies between the desired auditory signal for a speech sound and the current auditory feedback. Likewise, the somatosensory control system detects and corrects differences between the desired somatosensory signal of a speech movement and the current somatosensory feedback. The feedforward control system generates previously learned motor commands for production of a speech sound. Before introducing these three control systems in greater detail, I will first introduce a process which is critically involved in both feedback and feedforward control referred to as "internal model".

1.2.3 Internal model

In brief, a forward model predicts the sensory consequences of an outgoing motor command from a copy of that command ("efference copy"), while an inverse model solves the opposite problem, mapping a desired sensory outcome back onto the motor command needed to produce it. Motor control researchers use the umbrella term "internal model" to refer to these transformations between motor commands and sensory signals, which encompass both forward and inverse models. To make this concrete in the context of speech: when you decide you want to sing a particular high-pitched note, an inverse model maps that desired sound onto the motor commands needed to produce it – how much to tense the vocal folds, how much subglottal pressure to generate, and how to shape the vocal tract – and a forward model predicts the pitch you are about to hear, before the sound has actually reached your ears. The actual source of that prediction varies among theories. One long-standing idea is that the prediction comes from an "efference copy" of motor commands. The term "efference copy" was first introduced by the German physiologists Erich von Holst and Horst Mittelstaedt, and developed by others (Miall & Wolpert, 1996; von Holst & Mittelstaedt, 1950). The term is sometimes used interchangeably with "corollary discharge" referring to a signal sent from the motor system to sensory areas that pre-announces the upcoming motor command (Paus et al., 1996; Sperry, 1950). Specifically, according to theory, when the motor command is generated and sent to articulators, a prediction of its sensory outcome is simultaneously sent to the sensory areas in the brain, in order to allow the upcoming comparison from actual sensory input to this sensory prediction. In the SFC and FACTS models,

sensory predictions generated from "efference copy" are integrated with the current sensory state in order to update the predictions of sensory consequences of speech movements. Although the recent version of the DIVA model proposed that the predictions come directly from the sensory goal of the intended articulation, predictions and adjustment using internal models and the update of internal models are clearly involved in both feedback and feedforward control systems.

The idea that the cerebellum computes internal models has a long history. Ito first proposed that the cerebellar cortex operates as an adaptive 'neuronal machine' capable of learning internal models of the body and environment, in which climbing-fibre input from the inferior olive carries an error, or teaching, signal that adjusts the strength of parallel-fibre-to-Purkinje-cell synapses, gradually refining the mapping performed by a given cerebellar microzone (Ito, 1970). Wolpert, Miall and Kawato later synthesised this idea within a general computational framework for motor control, proposing that discrete cerebellar modules implement both forward models and inverse models, and reviewed converging behavioural, lesion, and imaging evidence consistent with this proposal (Wolpert et al., 1998).

The cerebellum is organised into functionally distinct territories. The anterior lobe and adjoining lobule VIII form a 'motor cerebellum', receiving somatotopically organised input relevant to limb and body movement, whereas posterior lobule VI and the lateral hemispheres (Crus I and Crus II) form a largely non-motor 'cognitive cerebellum', preferentially connected with prefrontal, parietal, and language-related association cortex rather than primary motor cortex (Buckner et al., 2011; Stoodley & Schmahmann, 2009). Each territory receives cortical input indirectly via the pontine nuclei (the cortico-ponto-cerebellar pathway) and

returns output to the cerebral cortex via the deep cerebellar nuclei and thalamus (the cerebello-thalamo-cortical pathway). Because these projections cross the midline twice in series, the net corticocerebellar loop is functionally crossed, such that the left cerebral hemisphere is preferentially connected with the right cerebellar hemisphere (Buckner et al., 2011; Ramnani, 2006).

Below, I provide further detail on the three control systems: auditory feedback control, somatosensory feedback control, and feedforward control, and I describe these in the context of the DIVA model.

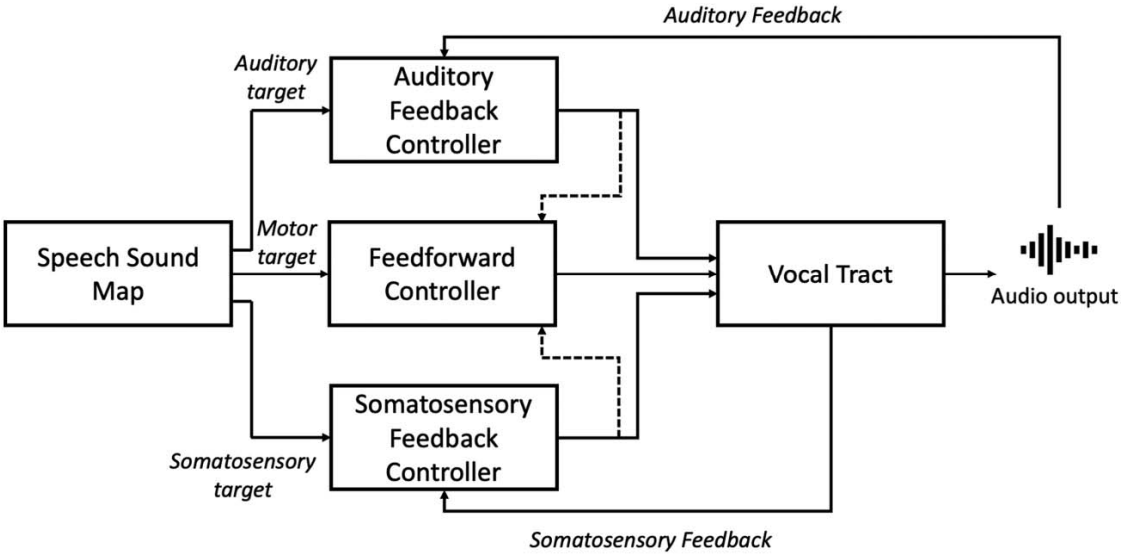

Figure 1.1: Simplified control scheme used by DIVA model for speech production.
Figure and caption are from Bradshaw et al. (2021) which was originally adapted from Guenther (2016). Motor commands sent to the vocal tract are the result of the summed output of the three control subsystems: the Feedforward controller, the Auditory Feedback controller, and the Somatosensory Feedback controller. The feedback controllers compare sensory feedback to their corresponding sensory targets for a speech production, and, if needed, issue online corrective motor commands based on detected errors. Dashed arrows indicate the use of these corrective commands to update the feedforward commands offline.

1.3 Auditory feedback control

The use of auditory feedback to refine motor commands for the speech articulators is crucial for achieving fluent speech. Auditory feedback control is especially important in childhood, when children are learning to produce the correct sounds of their language. When this control system is disrupted, speech can become inaccurate or dysfluent. In normal operation, the auditory feedback control system detects and corrects the difference between the desired auditory outcome and the actual feedback. Clearly, this system relies on a forward model that transforms motor commands into a predicted auditory outcome, which can then be compared to the incoming auditory feedback to drive the intended error correction.

1.3.1 Auditory feedback perturbation

The role of auditory feedback in speech can be investigated experimentally by perturbing feedback and observing consequent changes in speech production. Lombard (1911) introduced the idea that the auditory environment affects ongoing speech by demonstrating that speakers increase their loudness (sound level) in a noisy environment. The phenomenon is often referred to as the Lombard Effect - an unconscious and automatic process performed by the speakers, as they are usually unaware of changing the loudness of their speech. Lane and Tranel (1971) discovered that speakers normally adapt by increasing or decreasing their sound level by about half the amount of the change in noise level in order to maintain an acceptable signal-to-noise ratio for the speech sound.

A common perturbation explored previously in various experimental contexts is delayed auditory feedback. Delayed auditory feedback can severely impair speech fluency in fluent speakers. Little or no disruption occurs when the delay is shorter than 50 ms, but when the delay is around 200 ms (approximate duration of a syllable), speakers typically show a range of disruptions, such as sound repetitions, prolongations, and an overall slowing of speech (Stuart et al., 2002). Interestingly, delayed auditory feedback can temporarily enhance speech fluency in people who stutter, but this enhancement typically fades as they adapt to the perturbation.

In recent years, researchers have used more subtle perturbations of auditory feedback by spectrally altering auditory feedback of speech. These experiments involve recording the sound signal from a speaker via a microphone using a digital signal processing system to then perturb the spectral properties of the sound and feeding it back to the speaker through headphones. Most spectral perturbation studies have altered either the fundamental frequency (pitch) of the sound (Elman, 1981) or formant frequency of vowel sounds (Houde & Jordan, 1998). Perturbation of the sound signal is either expected (sustained altered auditory feedback across blocks of trials in an experiment), or unexpected (altered auditory feedback occurs only on randomly chosen trials). I will introduce more detail about these types of altered auditory feedback in Chapter 2.

1.4 Somatosensory feedback control

During speech, our brain receives a wealth of somatosensory information from the vocal tract. This tactile input originates from mechanoreceptors in the skin, which

send signals to the central nervous system whenever the skin is touched. The tongue, the most intricate speech articulator, transmits the most somatosensory information to the brain during speech. Its musculature is densely innervated with muscle spindles – on the order of several hundred in the human tongue (Kubota et al., 1975) – giving it a uniquely rich proprioceptive supply among the speech articulators. For example, when producing the alveolar consonant /s/, the tongue tip contacts the alveolar ridge, creating a vocal tract constriction. This area of constriction must be precisely controlled. Mechanoreceptors in the tongue and palate provide detailed information about its location and degree of constriction. Proprioception, the sense of body position and movement, originates from multiple sources beyond skin mechanoreceptors, notably muscle spindles which provide information about the length of the vocal tract and shortening speeds (Proske & Gandevia, 2012).

Somatosensory information is crucial for both specifying the speech target and activating the correct articulator movements to reach it. It is constantly monitored during speech and if a mismatch is detected, our brain adjusts the upcoming speech motor commands. This process, similar to auditory feedback control, involves fast forward modelling: transforming motor commands into somatosensory signals for comparison with the actual somatosensory signal perceived.

1.4.1 Somatosensory feedback perturbation

A number of experiments have demonstrated the critical role of somatosensory feedback in speech production.

Similar to auditory feedback perturbation, the mechanism of somatosensory feedback control is investigated by perturbing the feedback and observing the

changes in subsequent speech movement. Perturbation of somatosensory feedback is achieved by a block to impede mouth, tongue, or vocal tract movements. However, somatosensory perturbations do not disrupt speech completely as speakers adapt to this change quickly. For example, if you put a pen between your teeth, articulation is constrained and speech production is difficult. However, speakers adapt to this constraint quickly and produce perceptible speech. This phenomenon was first studied by Lindblom et al. (1979), who had participants produce Swedish vowels while using a bite block between their teeth. The findings indicated that even though the bite block resulted in physiologically unnatural jaw openings, all subjects were able to produce vowels which fell within the ranges of variation observed for a set of normally spoken vowels.

Somatosensory feedback control can also be tested by applying a short unexpected perturbation to the sensory input to articulators. These forms of short and unexpected perturbations avoid the confounding effects of feedforward motor learning. Folkins and Abbs (1975) applied a load resistance to the jaw during speech production. Loads were applied during the jaw closing movement associated with production of bilabial stops, creating a situation in which bilabial closure would be disrupted if motor control were independent of somatosensory feedback. Despite the unexpected load, participants produced rapid compensatory adjustments of the lips and jaw that largely preserved bilabial closure, with corrective movements emerging within the same syllable as the perturbation. This demonstrated that the speech motor system can exploit short-latency somatosensory feedback to re-achieve a task-level articulatory goal online, without relying on auditory information.

1.4.2 Relative contributions of auditory and somatosensory feedback

The response latency of somatosensory feedback control appears to be shorter than the response latency of auditory feedback control. Abbs and Gracco (1984) applied unanticipated loads to the lower lip during the generation of combined upper lip-lower lip speech gestures. The latencies of compensatory responses were in the 22 to 75 ms range. This is considerably shorter than the latencies observed for auditory feedback control (between 100 to 150 ms). It is worth noting that responses with latencies of lower brain stem-mediated reflexes (i.e. 10–18 ms) were not seen with inspection of individual records. This suggests that somatosensory feedback control involves a higher level of cortical processing. The fact that both somatosensory and auditory feedback control can only achieve partial compensation for the applied perturbations suggests that the compensation is not a simple reflexive response and indicates that both somatosensory and auditory speech targets are defined at higher cortical levels.

It is clear that somatosensory and auditory feedback control are necessary to ensure the ongoing delivery of speech gesture and speech sound respectively. It is interesting to consider whether speakers use auditory information to compensate for the somatosensory perturbations to the articulators. It seems obvious that when a speaker hears that the speech does not sound correct due to a somatosensory feedback perturbation they would use this auditory feedback to correct it. However, evidence indicates that this viewpoint is incorrect. Lindblom et al. (1979) reported that compensation to the somatosensory perturbation occurred before the speaker

uttered the first glottal pulse of the vowel, that is before any auditory information was available to the speaker. Nasir and Ostry (2008) extended this point by demonstrating that postlingually deafened cochlear implant users still compensated for small jaw-displacement perturbations when their implants were switched off, that is, in the complete absence of auditory feedback. These results demonstrate that the compensation to the somatosensory perturbation is at least in part driven by somatosensory feedback control.

But, do auditory and somatosensory information interact at the cortical level? Lametti et al. (2012) tested this idea by placing the auditory and somatosensory systems in competition during a speech perturbation task. They perturbed the somatosensory feedback by obstructing jaw movement and auditory feedback by changing the frequency of the vowel formant. Speakers compensated to either auditory (53%) or somatosensory (21%) perturbation alone or both (26%). The result of this study indicates that there are significant individual differences in the preference of speakers to use auditory or somatosensory feedback during speech.

1.5 Feedforward control

Fluent speakers produce speech sounds and movements quickly. However, conversational speech is often too fast to rely solely on feedback control. This is because neural processing delays in sensorimotor signals introduce time lags (Neilson & Neilson, 1987). As mentioned earlier, the delay in auditory feedback is about 100–150 milliseconds and in somatosensory feedback it is roughly 22–75 milliseconds. Articulatory movement patterns often need to be initiated and, in some cases,

largely executed before auditory or somatosensory feedback is received and used to adjust movements. This challenges the notion that speech motor control is solely driven by auditory or somatosensory feedback. Instead, the brain develops a central control and learning system that generates predicted preplanned movement patterns for speech. We refer to this mechanism as a feedforward control system.

Feedforward control of speech is learned throughout childhood and allows us to generate appropriate motor patterns for the phonemes and syllables of our native languages. Through repeated practice and exposure, the nervous system builds stable mappings between linguistic targets (e.g. intended sounds) and the articulatory commands required to achieve them, so that speech can be initiated quickly and produced smoothly without waiting for slow sensory feedback. During fluent production, these feedforward commands provide the primary drive for articulator movements, while auditory and somatosensory feedback serve a secondary role: they monitor the output, detect deviations from the intended target, and support gradual recalibration of the feedforward commands over time. In this way, feedback does not "run" speech moment-to-moment, but it remains essential for maintaining accuracy, adapting to changes in the vocal tract (e.g. growth, fatigue, perturbations), and refining the motor programmes that support stable speech production.

The DIVA model proposed a simple feedforward control scheme (Tourville & Guenther, 2011). The feedforward control system uses signals from auditory and somatosensory control systems to tune the motor signal. When producing a new speech sound, the feedforward controller's contribution to the overall motor command is imprecise because the motor target has not been tuned yet. Initially, the

model relies more on feedback controllers. Over time, the motor target and feedforward controllers improve and eventually eliminate the need for auditory and somatosensory feedback control unless these are perturbed. As speech articulators learn, corrective motor commands from auditory and somatosensory feedback eventually become part of the motor target, allowing the feedforward control to remain tuned.

1.6 Neural correlates of the DIVA model

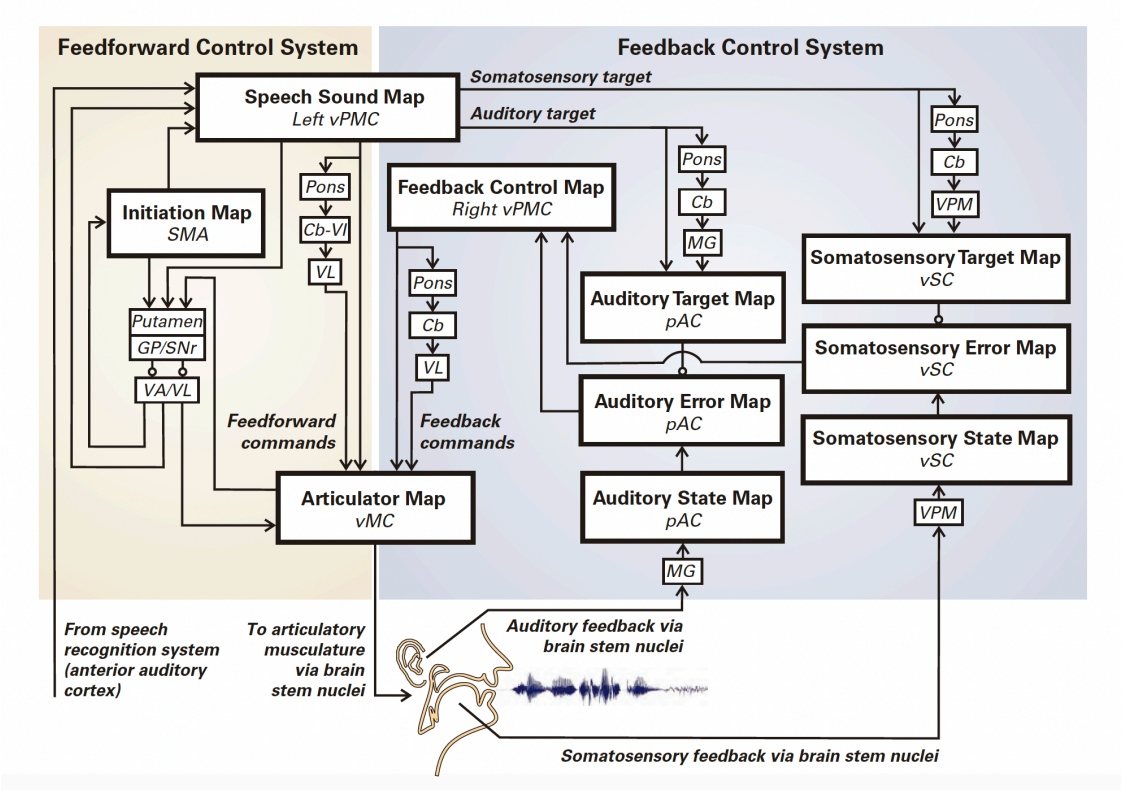

Figure 1.2: Neural correlates of the DIVA model.
Figure and caption are from Guenther (2016). Cb-V/VI, superior paravermal cerebellum (lobules V–VI); GP, globus pallidus; MG, medial geniculate nucleus of the thalamus; pAC, posterior auditory cortex; SMA, supplementary motor area; SNr, substantia nigra pars reticula; VA, ventral anterior nucleus of the thalamus; VL, ventral lateral nucleus of the thalamus; vMC, ventral motor cortex; VPM, ventral posterior medial nucleus of the thalamus; vPMC, ventral premotor cortex; vSC, ventral somatosensory cortex.

A key feature of the DIVA model is the set of neural correlates developed along-

side its cognitive schematics (Guenther et al., 2006; Tourville & Guenther, 2011). Specifically, different components of the feedback and feedforward control systems are linked to distinct neural substrates, as illustrated in Figure 1.2. Functions of different brain regions and loops were primarily derived from human clinical, neuroimaging, and microstimulation studies and non-human primate single cell recording and anterograde and retrograde tracing studies (Guenther et al., 2006).

In the DIVA model, producing a speech sound begins when the neural representation of that sound is activated in a "speech sound map" (Ghosh et al., 2008; Tourville et al., 2008). Each speech sound map unit is proposed to correspond to a population of neurons located mainly in left ventral premotor cortex (vPMC), spanning the rostral ventral precentral gyrus and adjacent areas of posterior inferior frontal gyrus and anterior insula. Once this representation becomes active, it drives motor commands that reach primary motor cortex through two interacting pathways: a feedforward control system and a feedback control system. The feedback control system is further divided into an auditory feedback controller and a somatosensory feedback controller (Tourville & Guenther, 2011).

The feedforward system is responsible for issuing well-learned motor programmes for speech sounds. It includes two main stages. First, the system must initiate the appropriate motor programme at the correct time. This initiation is attributed to a cortico–basal ganglia loop that includes an initiation map in the supplementary motor area (SMA) on the medial frontal wall (Bohland & Guenther, 2006; Ghosh et al., 2008). The loop is proposed to determine whether the current cognitive and sensorimotor context is appropriate for launching the next speech sound. For instance, during production of the word *enter*, initiating the motor programme for

ter depends on (i) a cognitive state reflecting the intention to say *enter* and (ii) a sensorimotor state indicating that articulation of *en* is nearing completion. In this account, these contextual cues are monitored by basal ganglia structures, with sensorimotor information associated with the putamen and cognitive information associated with caudate nucleus (the caudate nucleus is not depicted in the referenced figure and treated more fully in discussions of speech sequencing, see Bohland et al. (2010) for detail). Once the appropriate context is detected, the corresponding initiation unit is activated via basal ganglia output nuclei (globus pallidus and substantia nigra pars reticulata) and thalamic relays (Bohland & Guenther, 2006; Ghosh et al., 2008). Activation of the SMA initiation unit triggers the "readout" of the learned motor programme for the current sound.

In the second stage, the feedforward commands are encoded in projections from the speech sound map to an articulator map in ventral primary motor cortex (vMC) in both hemispheres (Bohland & Guenther, 2006; Ghosh et al., 2008). These connections from left vPMC to vMC are supplemented by a cerebellar contribution, described as a loop through the pons, superior paravermal cerebellum (lobules V–VI), and the ventral lateral thalamus, which supports the generation and refinement of the feedforward motor drive.

The auditory feedback controller is designed to detect and correct discrepancies between the intended auditory outcome of a speech sound and the auditory feedback actually received. In DIVA, the speech sound map units send projections—both directly and through a cortico–cerebellar pathway involving the pons, cerebellum, and thalamic auditory relays—to an auditory target map in secondary posterior auditory cortex (pAC), including regions such as planum temporale and posterior

superior temporal gyrus/sulcus. Activity in this auditory target map represents the expected sound of the utterance: it specifies what auditory feedback should be heard while producing the current speech sound. Crucially, these targets are not single-point values but time-varying "target regions" that define an acceptable range of acoustic variation across the syllable. This target-region concept is a core element of DIVA and is used to account for a broad set of phenomena, including motor equivalence, contextual variability, coarticulation (both anticipatory and carry-over), and speaking-rate effects (Tourville et al., 2008).

Incoming auditory information from the periphery is relayed to cortex via thalamic auditory nuclei and represented as an auditory state map (Buchsbaum et al., 2001). The model compares this auditory state with the auditory target region. If the current feedback falls outside the allowable region, units in an auditory error map (also hypothesised to lie within posterior auditory cortex) become active. The resulting error activity is then converted into corrective motor commands through projections from the auditory error map to a feedback control map located primarily in right vPMC. This feedback control map, in turn, projects to the articulator map in vMC—both directly and via a loop through pons, cerebellum, and ventral lateral thalamus—thereby implementing rapid motor adjustments that reduce the auditory error.

Alongside auditory feedback control, DIVA posits a somatosensory feedback controller that monitors and corrects tactile and proprioceptive aspects of articulation. The main somatosensory components are proposed to lie in ventral somatosensory cortex (vSC), including ventral postcentral gyrus and supramarginal gyrus (Ghosh et al., 2008; Guenther et al., 2004). Projections from the speech

sound map to a somatosensory target map—via both corticocortical routes and cortico–cerebellar loops involving thalamic relays such as ventral posterior medial (VPM) nucleus—encode the expected tactile and proprioceptive feedback associated with producing the current sound (arising from mechanoreceptors and muscle spindles in the vocal tract). Actual somatosensory input from the articulators is represented in a somatosensory state map, receiving signals that travel from cranial nerve nuclei in the brainstem through VPM. When the current somatosensory state deviates from the target region, nodes in a somatosensory error map become active. The output of this error map is then sent to the shared feedback control map, where somatosensory errors—like auditory errors—are transformed into corrective motor commands that adjust articulation and bring sensory feedback back within the intended target ranges.

This neural account of the DIVA model has also been tested directly by Golfinopoulos et al. (2010), in which they simulated functional MRI activation patterns from DIVA model predictions and compared them with observed activation during speech production. This provides an important form of model validation. The results showed a good correspondence between the simulated fMRI activation patterns and observed brain activation during speech production. However, it is important to note that the neural correlates of the DIVA model are derived from independent studies of speech production, and therefore specific experimental tasks are needed to test the feedforward and feedback control loops more directly and independently.

1.7 Summary

In this chapter, I outlined the historical development of research on speech motor control, before reviewing influential computational accounts of how speech movements are planned, executed, monitored, and corrected. I then focused on the DIVA model, describing how it conceptualises speech production as the interaction of three complementary control systems: an auditory feedback controller that detects and compensates for deviations between predicted and perceived acoustic outcomes, a somatosensory feedback controller that monitors and stabilises articulatory states using tactile and proprioceptive information, and a feedforward controller that supports rapid, fluent production by generating well-learned motor commands with minimal reliance on online sensory feedback. Finally, I summarised the proposed neural correlates of these components in the DIVA model.

In the next chapter, I turn to speech adaptation (also known as sensorimotor adaptation of speech) as a paradigm for investigating speech motor control. By examining how speakers adjust their productions when auditory feedback is systematically altered, speech adaptation experiments make it possible to characterise the contributions of each control system and, crucially, determine how they interact cortically. In this way, studying speech adaptation can help specify when feedback-based correction dominates, when feedforward commands are recalibrated, and how auditory and somatosensory signals jointly shape the learning processes that maintain accurate and stable production of speech.

2

Introduction to speech adaptation

This chapter reviews how experimental paradigms using altered auditory feedback have been used to study speech adaptation, and what this work reveals about the behavioural and neural mechanisms of speech motor control. The first half focuses on behavioural findings from pitch and formant perturbation paradigms, including their implications for feedforward and feedback control, perception, and individual differences. The second half turns to the neural correlates of speech adaptation, drawing on functional and structural neuroimaging, evidence from populations with developmental disorders of speech and language, populations with other neurological disorders, and noninvasive brain stimulation. The chapter closes by identifying knowledge gaps that motivate studies of the thesis.

2.1 Investigating speech adaptation using altered auditory feedback

As I introduced in Chapter 1, speech adaptation can be studied using altered auditory feedback. Many studies exploring speech adaptation focused on the role of sensory feedback and the mechanisms involved in speech motor control. These studies often investigated the use of internal models in speech motor control and the function of the auditory feedback system (MacDonald et al., 2010) in speech production.

Auditory feedback can be altered unexpectedly with an upward or downward shift applied randomly to utterances (Burnett et al., 1998) or sustained as the same type of shift is applied continuously for a number of utterances (Houde & Jordan, 1998). In both unexpected and sustained perturbation scenarios, speakers typically adjust their speech in the direction opposite to the perturbation. This phenomenon is often regarded as "compensation". Specifically, in sustained altered auditory feedback, speakers show increased magnitude of compensation and reach a plateau after a number of utterances (see Figure 2.3 B). This training-related learning effect is often regarded as "adaptation". Applying a sustained perturbation to sensory feedback during speech leads to a sustained modification of the motor commands used to produce speech sounds. This adaptability allows us to change the way we speak according to the surrounding environment (e.g. the Lombard effect) or to imitate another's speech sound. This adaptability also enables us to maintain intelligible speech as the geometry of our articulators changes naturally (e.g. lengthening of vocal tract) throughout life. By perturbing the sound of

speech in real time, researchers can investigate how auditory feedback control and feedforward control interact to support rapid and accurate speech production.

Auditory feedback has most commonly been altered along two acoustic dimensions: fundamental frequency, perceived primarily as pitch, and formant frequency, which contributes to vowel identity. Although the present thesis focuses on formant perturbation rather than pitch perturbation, pitch perturbation studies are reviewed first for three reasons. First, pitch perturbation was introduced earlier in the auditory feedback literature and has provided much of the foundational evidence for speech adaptation. Second, the basic experimental logic of pitch perturbation studies closely parallels that of formant perturbation studies. Third, because Mandarin is a tonal language, pitch plays a linguistically contrastive role, and a comparatively large body of work has examined how speakers of tonal languages respond to pitch-shifted feedback. Reviewing this literature therefore provides a useful conceptual bridge to the subsequent discussion of formant perturbation.

2.1.1 Pitch perturbation

Elman (1981) first demonstrated the compensatory responses to real-time auditory perturbation on fundamental frequency. Fundamental frequencies of voice were perturbed by using a device called Lexicon Varispeech II Speech Time Compressor/Expander. Most subjects compensated for this perturbation by changing their pitch in the opposite direction to the perturbation applied. This compensatory response to pitch perturbation occurred without speakers being consciously aware of the shift. Since then, a number of studies have been conducted to carefully address the properties of compensatory responses to pitch perturbation. For example, Burnett

et al. (1998) found that the magnitude of the compensatory response was not systematically affected by the intensity of the feedback and the averaged magnitude of the compensatory response was smaller than the perturbation. The authors also noticed that not all trials showed compensatory responses (response in the opposite direction of perturbation). A few trials showed following responses (response in the direction of perturbation). In addition, a short period of sustained pitch perturbation led to after-effects in pitch production, i.e. the speaker still compensated when the feedback was returned to normal (Jones & Munhall, 2000).

Mandarin is often chosen as a representative tonal language because pitch patterns (lexical tones) are contrastive: relatively small changes in fundamental frequency can shift a syllable from one tone category to another and therefore change syllable meaning. This makes Mandarin a particularly informative test case for models of auditory-motor control that were originally developed largely from English and other non-tonal languages, where pitch is not directly tied to segmental lexical contrasts.

Jones and Munhall (2002) investigated pitch control in Mandarin using a pitch perturbation paradigm. In the experiment, Mandarin speakers exhibited a clear compensatory response: when feedback pitch was increased or decreased, speakers adjusted their produced pitch in the opposite direction, consistent with rapid error correction based on auditory feedback. Crucially, speakers also showed an after-effect once the perturbation was removed and feedback returned to normal. That is, pitch production did not immediately revert to baseline; instead, it remained biased for a time in a direction consistent with having adapted to the

earlier perturbation. Together, results showed that pitch control in Mandarin involves both immediate feedback-driven correction (compensation) and short-term sensorimotor learning (adaptation).

Subsequent work has reproduced these findings, reinforcing the idea that Mandarin speakers can adapt quickly to pitch perturbations (Xu et al., 2004). In these studies, speakers typically produced Mandarin vowels, syllables, or words while the fundamental frequency of their auditory feedback was shifted upward or downward. The resulting mismatch between produced and heard pitch allowed researchers to examine whether speakers corrected for the perturbation, how quickly they responded, and whether repeated exposure led to longer-term adaptation. Jones and Munhall (2002) reported that Mandarin speakers responded to pitch perturbation with an average latency of 211 ms, considerably slower than the average latency of about 100–120 ms reported in English speakers. This difference suggests that pitch-feedback control in Mandarin may involve additional linguistic constraints, because pitch is not only an acoustic property of the voice but also a cue to lexical tone.

Importantly, compensation and adaptation in a tonal language are therefore not arbitrary changes in voice pitch. They are shaped by the functional demand to maintain tone categories and preserve lexical contrasts. For example, if a perturbation shifts the perceived pitch contour toward a neighbouring tone category, the speaker's corrective response may be constrained by the need to keep the intended tone intelligible. Corrective movements can therefore interact with the tonal target being produced and with the surrounding tonal space relevant for intelligibility (Feng et al., 2018). Consistent with this interpretation, several studies suggest that

the magnitude and neural correlates of pitch-perturbation responses vary with language experience and proficiency, implying that sensorimotor mappings for pitch are modulated by long-term experience with tonal categories and their production demands (Chen et al., 2012; Liu et al., 2010; Ning et al., 2014). Taken together, this body of evidence indicates that compensatory responses and after-effects are not specific to English or to non-tonal languages. Instead, they reflect more general mechanisms of speech motor control that operate across languages, while also being tuned by the linguistic role of pitch in the language.

For a broader discussion of how pitch perturbation paradigms have been used with tonal-language speakers and what they reveal about tone production, auditory feedback control, and sensorimotor learning, see the review by Tang (2024).

2.1.2 Vowel formants

Having reviewed pitch perturbation studies, the following section turns to formant perturbation. Whereas pitch perturbation alters fundamental frequency, formant perturbation alters the spectral cues that contribute to vowel identity. This distinction is important because vowel production depends strongly on the configuration of the vocal tract.

Formants are the resonance frequencies of the vocal tract, and the frequencies of the first formant (F1) and second formant (F2) are major acoustic cues that determine which vowel is perceived by the listener (Ladefoged & Broadbent, 1957; Ladefoged & Johnson, 2014). The frequencies of F1 and F2 are related to the tongue's position in the vocal tract, with F1 being affected by the height of the tongue and F2 reflecting its horizontal position when raised known as the degree

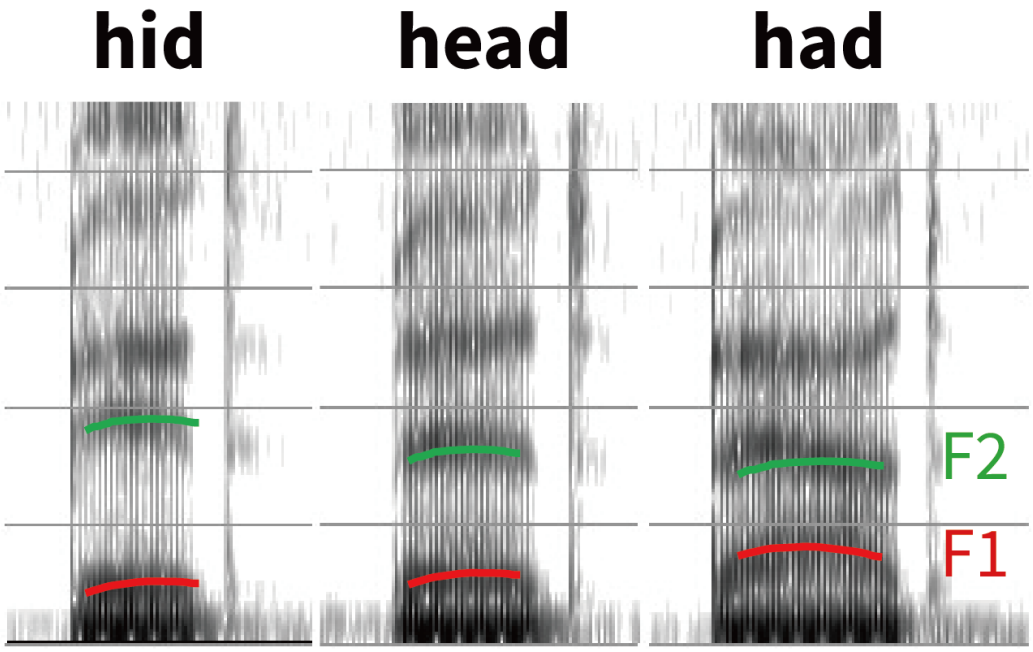

Figure 2.1: Vowel formants.
Spectrogram showing the first and second formant frequencies. Red indicates F1 and green indicates F2.

of "backness" of a vowel or "openness" of the vocal tract (Ladefoged & Johnson, 2014). See Figure 2.1 for a spectrogram showing the first and second formant frequencies. See Figure 2.2 A for an illustration of first and second frequencies of vowel formants in English and see Figure 2.2 B for a illustration of first and second frequencies of vowel formants in Mandarin.

2.1.3 Formant perturbation

The experimental paradigm involving altered auditory feedback through formant perturbation was first introduced by Houde and Jordan (1998). In their experiment, participants whispered consonant-vowel-consonant (CVC) words while their auditory feedback was transformed in real time. The training words were bilabial CVC words containing the vowel /ɛ/, namely "pep", "peb", "bep", and "beb". Dur-

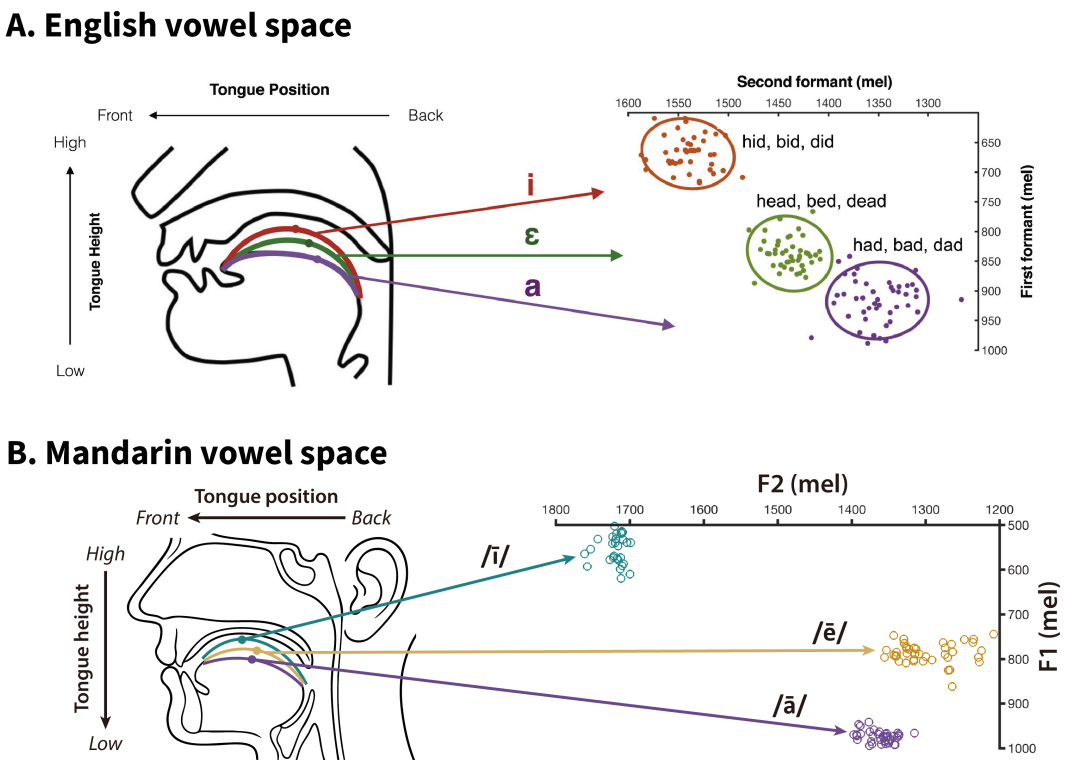

Figure 2.2: Vowel space in formant frequencies.
A. English vowel space: Figure and caption are from Tang et al. (2021). Schematic showing the relationship between tongue height and position in the vocal tract (left) and the corresponding frequencies of the first and second formants of vowels (/i/, /ɛ/, and /æ/) in vowel space from a single participant in the 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 the same colour.
B. Mandarin vowel space: 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 Mandarin vowels (right) from a single participant in the study presented in Chapter 4. Right figure shows vowel production. F1 and F2 frequencies of each trial utterance are plotted in yellow for the vowel /ē/, green for the vowel /ī/ and purple for the vowel /ā/.

ing productions of these training words, the formant frequencies of participants' feedback were shifted along an individually defined vowel path in F1–F2 space. The perturbation was designed so that, when a participant produced /ɛ/, they heard a vowel closer to either /i/ or /a/, depending on the direction of the shift. Participants compensated by changing their vowel productions in the direction op-

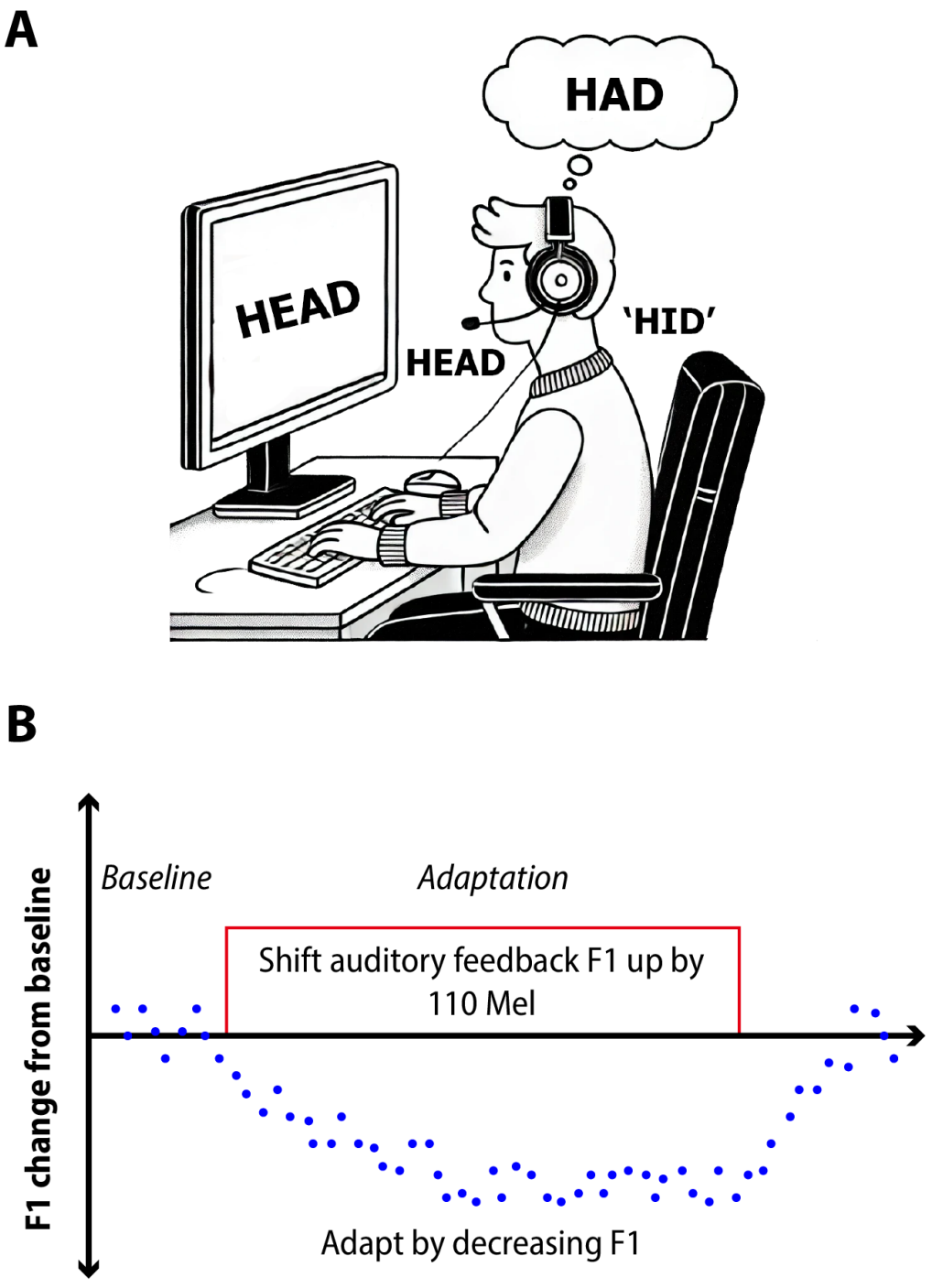

Figure 2.3: Formant perturbation task.
A. Experimental setup. This image was generated with the assistance of ChatGPT. **B.** Schematic illustration of the predicted change in produced F1 from baseline when F1 feedback is increased during the adaptation block. Blue dots indicate F1 frequency on individual trials.

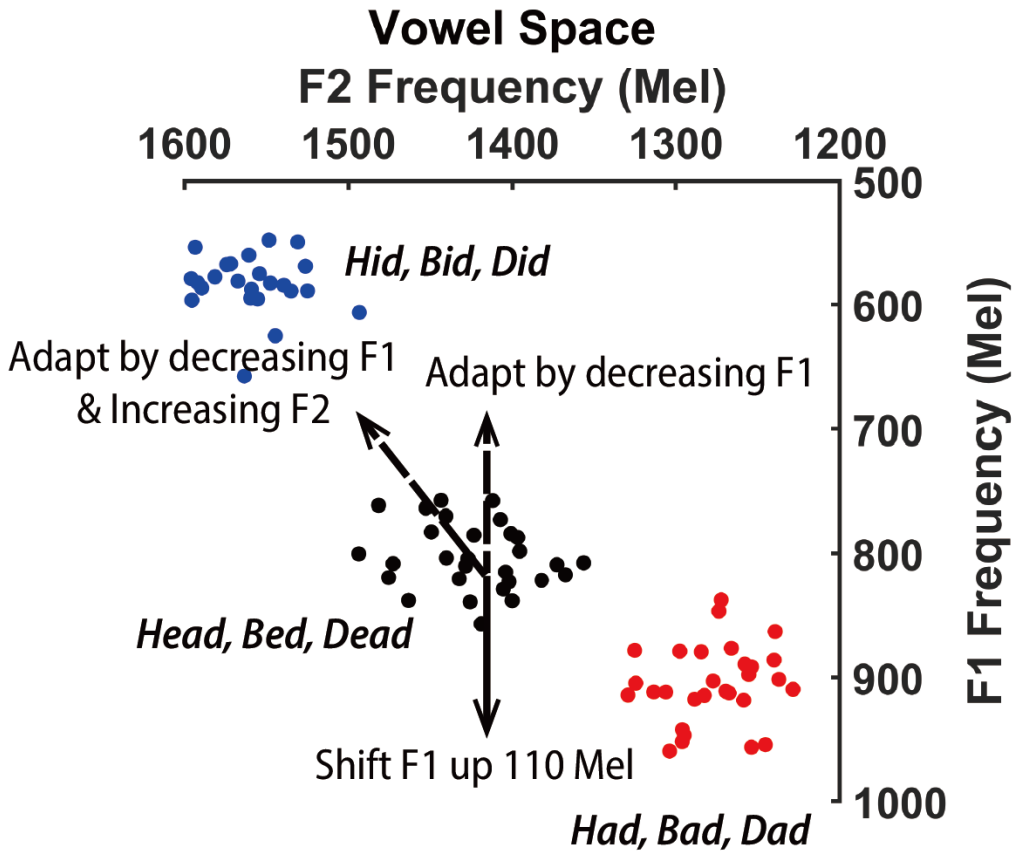

Figure 2.4: Response to formant perturbation.
Vowel space showing vowel production data from a single participant. F1 and F2 frequencies of each utterance are plotted in black for the vowel in "head", blue for the vowel in "hid", and red for the vowel in "had". The downward black arrow indicates the direction of the shift applied to F1 feedback during the adaptation block. The two dashed arrows illustrate two possible compensatory responses to the upward shift in F1 feedback: (i) adapt by decreasing F1 only; and (ii) adapt by decreasing F1 and increasing F2 that move production toward the vowel in "hid".

posite to the altered feedback, and part of this change was retained when feedback was masked by noise, indicating adaptation. Houde and Jordan also showed that adaptation transferred to /ɛ/ in untrained consonant contexts (i.e. "gep" and "teg") and generalised to other vowels (i.e. "pip" and "pap"). Together, these findings provided early evidence that auditory feedback can recalibrate the motor commands involved in vowel production.

Many later formant-perturbation studies adopted spoken words, especially "head",

"hid", and "had". This stimulus set is useful because the vowels form neighbouring regions in F1–F2 space: the vowel /ε/ in "head" lies between the higher vowel /i/ (low in F1, high in F2) in "hid" and the lower vowel /a/ (high in F1, low in F2) in "had". As a result, perturbing the formants of "head" allows researchers to examine not only whether speakers compensate for the altered formant, but also whether their compensation is affected by nearby vowel categories. For example, shifting the F1 of auditory feedback upward makes the vowel in "head" sound acoustically closer to "had". Speakers may therefore respond by decreasing their F1 production to offset the perturbation. Alternatively, they may adapt by decreasing F1 and increasing F2 that move production toward another familiar vowel category, i.e. "hid". This logic is illustrated in Figure 2.4.

Figure 2.3 summarises the formant-perturbation paradigm. Panel A shows the experimental setup, in which speakers produce prompted words while their auditory feedback is perturbed. Panel B shows the expected change in produced F1 across the task when F1 feedback is increased: after the baseline phase (feedback unperturbed), speakers are expected to reduce their production in F1 in response to the perturbation, and this compensatory change may persist briefly when normal feedback is restored. Figure 2.4 places this response in vowel space. The vowel in "head" is the perturbed target, while "hid" and "had" provide neighbouring vowel categories. The solid arrow shows the direction of the imposed upward F1 shift in auditory feedback, and the dashed arrows show two possible compensatory responses: a direct reduction in F1, or a combined change in F1 and F2 that moves production toward "hid".

Subsequent formant-perturbation studies have extended this basic observation in several important ways. In particular, they have examined whether compensatory changes are consistently opposite to the perturbation, whether compensation is complete or partial, whether adaptation depends on conscious awareness of the altered feedback, and whether adaptation persists after the perturbation is removed. These questions are important because they move the interpretation of formant perturbation beyond a simple online correction response and instead frame it as evidence of sensorimotor adaptation of speech.

First, compensation is generally in the direction opposite to the auditory perturbation. This pattern was examined in more detail by Houde and Jordan (2002), who showed that changes in F1 and F2 were largely compensatory rather than assimilative. That is, speakers modified their productions in a direction that reduced the mismatch introduced by the altered feedback. This finding has since been replicated in a range of formant-perturbation studies (Munhall et al., 2009; Villacorta et al., 2007), supporting the view that speakers use auditory feedback to monitor and correct vowel production.

Second, compensation is typically partial rather than complete. Speakers usually do not fully compensate for the perturbation by the end of training. This issue was systematically tested by MacDonald et al. (2010). In one experiment, F1 was increased by 50, 100, and 200 Hz while F2 was simultaneously decreased by 75, 125, and 250 Hz. In a second experiment, F1 and F2 shifts were incrementally changed on each utterance by +4 and −5 Hz, respectively, up to maxima of +350 and −450 Hz. In both experiments, speakers compensated in the direction

opposite to the perturbation. However, compensation was only approximately 25–30% of the shift magnitude for perturbations up to 200 Hz in F1 and 250 Hz in F2; for larger shifts, compensation reached a plateau and then declined. These findings suggest that the speech motor system does not simply mirror the size of the auditory error. Instead, compensation appears to be constrained by factors such as the reliability of the auditory feedback, the plausibility of the heard vowel, and the stability of existing vowel targets.

Third, Houde and Jordan (2002) noted that compensatory responses did not appear to be related to conscious awareness of the auditory perturbation. Most speakers reported little or no awareness of either the feedback perturbation or changes in their own speech. This observation was tested more directly by Munhall et al. (2009), who compared three instruction conditions. In the control condition, participants were not informed about the perturbation. In the "ignore headphones" condition, they were told that their voice might sound different and were instructed to ignore the auditory feedback. In the "avoid compensation" condition, they were explicitly informed about the perturbation and instructed not to compensate. Despite these different instructions, participants showed robust compensation in all three conditions, with no reliable differences in either the magnitude of compensation or the trajectory of adaptation during sustained perturbation. These results suggest that speech adaptation is not simply a voluntary strategy, but reflects an implicit speech motor learning.

Finally, speech adaptation can persist after altered feedback is removed. Houde and Jordan (2002) reported that when participants returned approximately one

month later for a control session with unaltered feedback, modified formant production was still evident. The authors interpreted this long-term retention as evidence for implicit learning or stable changes in speech motor control mechanisms. Although this specific long-term retention effect has not been systematically examined in later work, many altered-auditory-feedback paradigms included an "after-effect" or "wash-out" block in which feedback was returned to normal following a sustained perturbation. In one study, Purcell and Munhall (2006) found that compensation persisted for at least 115 trials after feedback was restored to normal. However, the return to baseline values for F1 and F2 production shows various patterns across different studies (MacDonald et al., 2010; Parrell & Niziolek, 2021; Purcell & Munhall, 2006). These differences raise the question of how adaptation is maintained and updated over time, which leads to the relative contributions of feedback and feedforward control.

2.1.4 Involvement of feedforward and feedback control

Many studies using formant perturbation have been designed to characterise the contributions of feedforward and feedback control to speech adaptation. In the DIVA framework, online auditory feedback is used to detect errors during speaking, and the resulting corrective commands are incorporated into later feedforward motor plans (Guenther, 2016). By contrast, the FACTS model (see Chapter 1, Section 1.2.1) proposes that adaptation is driven by sensory prediction error: auditory reafference is compared with expected sensory consequences generated by a task- or planning-level controller, rather than by a motor "efference copy" of

corrective commands (Kim et al., 2023b). Related accounts in the broader motor-learning literature make a similar claim, namely that learning is guided primarily by sensory error signals rather than by the direct transfer of online corrections into future motor commands.

A useful way to test these ideas is to ask whether speech adaptation can occur even when no overt speech movement is produced. Parrell et al. (2026) addressed this question using "speech cancellation" trials. In this paradigm, speakers were cued to withhold an upcoming utterance, but they still experienced a sensory mismatch because a formant perturbation was applied to a recording of their own speech from a previous trial at the moment they would otherwise have spoken. Speakers adapted not only in standard speaking trials but also in these cancellation trials, showing compensatory change even in the absence of overt articulation. Parrell and colleagues therefore argued that sensory prediction errors, rather than corrective motor commands alone, are sufficient to drive speech adaptation. This finding strengthens the broader view that formant perturbation reveals not just immediate feedback correction, but also the mechanisms by which sensory error is integrated to support feedforward learning.

2.1.5 Perception and sensorimotor integration

The previous section considered how altered auditory feedback engages feedback and feedforward control in speech adaptation. A related issue is how the sensory information provided by that feedback is interpreted and incorporated into the speech system. Formant perturbation is therefore also useful for studying perception and sensorimotor integration, because it directly manipulates what speakers

hear while they speak and allows researchers to examine how perceptual targets, sensory feedback, and motor output are coordinated during speech adaptation.

One line of work asks whether speech adaptation depends on perceptual precision. A common interpretation is that more precise perceptual representations provide more stable production targets, which lead to larger compensatory responses. Positive correlations between auditory perceptual acuity and amplitude of speech adaptation have been reported in previous studies (Martin et al., 2018; Villacorta et al., 2007). However, adaptation was not clearly related to either auditory acuity or somatosensory acuity, measured through jaw-position sensitivity (Feng et al., 2011). This suggests that the link between perception and adaptation is more complex than a simple one-to-one mapping between perceptual precision and speech adaptation.

One factor that may contribute to this complexity is the relative weight speakers assign to auditory versus somatosensory feedback when updating their motor commands. Two perturbations were combined in one study: a somatosensory perturbation produced by a robotic jaw pull and an auditory perturbation produced by an F1 shift. Speakers mainly compensated for the auditory shift (Feng et al., 2011). A similar dual-perturbation design showed that speakers typically adapted more to one modality; speakers who compensated more for the somatosensory perturbation compensated less for the auditory perturbation and vice versa (Lametti et al., 2012). These findings suggest that individuals differ in the extent to which they rely on auditory versus somatosensory feedback when updating speech motor commands. They also indicate that adaptation is not driven by auditory feedback alone, but emerges from the interaction between auditory and somatosensory feedback control.

As noted earlier, adaptation to auditory perturbation is often incomplete (Houde & Jordan, 2002). The proportion of compensation tends to decrease as perturbation size increases, and may even decline for very large shifts (Katseff et al., 2012; MacDonald et al., 2010). This pattern has been interpreted as a trade-off between auditory and somatosensory control: when the auditory error becomes large, auditory feedback may be treated as less reliable, and greater weight may be placed on somatosensory feedback to preserve speech stability (Katseff et al., 2012). Therefore, incomplete compensation may reflect a balance between competing sensory constraints.

2.1.6 Phonological and perceptual categories

In English and Mandarin, vowel identity is determined mainly by the frequencies of first two formants, F1 and F2. Because vowel categories depend on the relationship between these cues, speakers do not always respond to perturbations by changing only one formant. For example, increasing F1 in "head" can make the vowel sound more like "had". Some speakers compensate by changing both F1 and F2, for instance by lowering F1 and raising F2, which can move production toward "hid" (see Figure 2.4). This suggests that some speakers aim for vowel categories rather than isolated acoustic targets.

This finding could relate to the fact that speech is perceived categorically. Continuous acoustic variation is mapped onto a limited set of phonemic categories, with greatest sensitivity near category boundaries and reduced sensitivity within categories. As a result, changes in production and perception often reflect movement across category boundaries rather than simple shifts in raw acoustic values.

Perceptual boundaries can shape adaptation to auditory perturbations. In one study, the "head"–"had" boundary was shifted through perceptual reinforcement before participants were exposed to an auditory perturbation that increased F1 in "head" toward "had" (Shiller et al., 2013). During training, participants repeatedly judged the same acoustic continuum and received trial-by-trial feedback, but the feedback boundary was shifted relative to each participant's original perceptual boundary. This biased the mapping between acoustics and category labels and allowed the effect of perceptual learning on later compensation to be tested. Greater adaptation followed training that made ambiguous tokens more likely to be classified as "had" than training in the opposite direction (Shiller et al., 2013). Similar findings were reported in children after boundary training in the relevant direction, compared with training on an unrelated contrast (Shiller & Rochon, 2014). In adults, adaptation magnitude has also been found to correlate with the trained boundary position (Lametti et al., 2014b).

Perceptual boundaries can also be shifted without explicit training. The acoustics of a carrier phrase were manipulated by changing pitch and formants, and exposure to a higher-frequency carrier phrase shifted the "bit"–"bet" boundary toward "bet" (Bourguignon et al., 2016). Speakers exposed to that carrier phrase later adapted more to an auditory perturbation that pushed /ε/ toward /i/ than speakers exposed to a lower-frequency carrier phrase (Bourguignon et al., 2016). This suggests that context-dependent changes in perception can transfer to production.

Lexical status can also influence adaptation. This lexical bias is often described in terms of the "Ganong effect", in which ambiguous speech sounds are more likely to be perceived as real words than nonwords (Ganong, 1980). Larger adaptation

has been observed when a perturbation transforms a nonword into a real word than when it transforms a real word into a nonword (Bourguignon et al., 2014). Lexical knowledge therefore affects how altered feedback is categorised, which in turn shapes speech adaptation.

Cross-linguistic studies also show that adaptation depends on language-specific phoneme systems and perceptual boundaries. Similar upward and downward F1 shifts were applied to English speakers producing "head", Japanese speakers producing /he/, and Japanese learners of English producing "head"; adaptation to the downward shift was similar across groups, but Japanese speakers showed smaller adaptation to the upward shift than English speakers (Mitsuya et al., 2011). Differences have also been reported between French and English speakers in response to similar perturbations, with perceptual testing suggesting that the effects were mediated by language-specific perceptual boundaries and contrast sensitivity (Mitsuya et al., 2013).

Speech adaptation can also influence speech perception. Perceptual boundaries were measured before and after speech adaptation and again after a "washout" phase, and adaptation shifted the boundary in the region of the speaker's produced response rather than in the region of the heard perturbation (Lametti et al., 2014b). For example, speakers who produced "head" while hearing feedback shifted toward "had" compensated toward "hid", and their perceptual boundary between "head" and "hid" shifted accordingly, without a comparable shift between "head" and "had".

Not all studies report overt behavioural boundary shifts. A similar approach did not yield a reliable perceptual boundary shift, although EEG responses in N1

and P2 changed for ambiguous stimuli and those changes correlated with speech adaptation magnitude (Schuerman et al., 2017). Perceptual change may also depend on what speakers do during adaptation. In an F2 perturbation study shifting /*i*/ toward /*u*/, the "see"–"she" boundary shifted differently depending on whether speakers followed the perturbation or compensated for it (Schuerman et al., 2017).

Overall, adaptation to altered auditory feedback is shaped by perceptual category structure, lexical knowledge, training history, and language background. In some cases, adaptation also affects perception, but the size and direction of this effect vary across tasks and populations. Together, these findings suggest that perceptual change following speech adaptation is not always evident at the behavioural level, and that its emergence may depend on neural processing of ambiguous stimuli and on the speaker's adaptive response.

2.1.7 Speakers' ability and speech development

Beyond the factors discussed earlier, some studies have explored the link between speakers' abilities and the extent of their speech adaptation. This research aims to address the individual variability often observed in speech adaptation tasks. The evidence is mixed. General executive control did not predict adaptation, whereas auditory acuity did, and speakers informed about the perturbation showed compensation similar to naive speakers (Martin et al., 2018; Munhall et al., 2009). Social factors have also been implicated: speakers who scored lower on a self-report measure of personal empowerment – encompassing self-esteem, perceived control, and optimism – showed larger compensatory responses than those who

scored higher (Dimov et al., 2012). Together, these findings suggest that individual variability in speech adaptation tasks cannot be explained by ability or knowledge alone, and may instead reflect differences in auditory sensitivity, as well as social and motivational factors.

Speech adaptation has also been examined in children. Four-year-old children adapted to altered auditory feedback, whereas two-year-old toddlers did not, suggesting that very young children may either rely less on auditory feedback or have immature feedforward control (MacDonald et al., 2012). More recent work showed that auditory-perceptual training can improve speech adaptation in 5–7-year-old children, indicating that perceptual learning and motor learning are already linked during early childhood development (Shiller & Rochon, 2014). This line of work showed that speech adaptation ability is not fixed in childhood, but can be strengthened when perceptual representations are sharpened.

2.2 Neural correlates involved in speech adaptation during altered auditory feedback

As I introduced in the previous chapter, speech motor control involves the neural networks responsible for the auditory, somatosensory feedback control systems and the feedforward control system. In the human brain, the sensorimotor cortex, cerebellum, and connections between them are believed to be involved in sensorimotor adaptation. According to the DIVA model, speech motor cortex receives motor commands from the feedforward control system and receives the adjusted motor commands from the feedback control; the cerebellum supplements the projection

of speech motor commands from the ventral premotor cortex to ventral motor cortex for articulation (Tourville & Guenther, 2011). In addition, the cerebellum is also thought to receive direct sensory input about the current state of the effectors, which, when integrated with an "efference copy" of intended movements from the motor system, can be used to predict the sensory consequences of movement (Hickok et al., 2011). This process would integrate both feedback and feedforward control and can happen rapidly without interaction with the cerebral cortex, making it feasible for the cerebellum to achieve rapid correction of ongoing movements. Here, I will summarise studies investigating the neural correlates of speech adaptation during altered auditory feedback. These studies used techniques including functional and structural brain imaging methods, populations with developmental disorders of speech and language, populations with other neurological disorders, and brain stimulation methods.

2.2.1 Functional brain imaging studies

Some studies have noted that functional MRI may be less suitable for tracking rapid changes during sustained altered auditory feedback because the haemodynamic signal is relatively slow and may be difficult to separate from low-frequency fluctuations in speech tasks (Zheng et al., 2013). Even so, fMRI has been used to detect brain responses to unexpected perturbations. In one early study, speaking under unexpected altered feedback produced increased activity in posterior auditory cortex bilaterally, including the superior temporal gyrus and planum temporale, suggesting detection of an auditory mismatch (Tourville et al., 2008). Increased activity was also observed in right ventral motor and premotor cortex and in the

anterior medial cerebellum bilaterally. The results suggest that feedback control involves more right hemisphere activity whereas the feedforward control involves more left hemisphere activity.

More fine-grained multivariate analysis identified distinct brain systems engaged during speech adaptation (Zheng et al., 2013). The experimental procedure consisted of production of monosyllabic words with normal feedback (baseline), altered feedback where the frequency of the first vowel formant (F1) was increased, and feedback with a masking noise. Using fMRI with multivoxel pattern analysis, three functional networks were identified: an error-signal network involving the right angular gyrus, right supplementary motor area, and cerebellum; a passive-listening network involving left and right anterolateral superior temporal gyrus and left precentral gyrus; and a third network involving the left and right inferior frontal gyrus captured signals from the previous two networks. These results suggest that speech adaptation recruits separable systems for error detection, speech perception, and sensorimotor integration.

Perceptual category structure also affects neural responses to altered auditory feedback. When perturbations were designed to move speech across a vowel-category boundary, greater activity was observed in right posterior superior temporal gyrus for near-boundary shifts than for far-boundary shifts (Niziolek & Guenther, 2013). Significant correlation was found between the amplitude of speech adaptation and brain activity: as participants adapted more to the altered feedback, increased brain activity was found in a network involving superior temporal and inferior frontal brain regions and decreased brain activity was found in a network involving ventral motor and somatosensory cortex bilaterally.

More recently, Floegel et al. (2020) assessed the neural correlates of both spectral and temporal perturbations to the auditory feedback using fMRI. Participants received sustained altered auditory feedback that was either delayed (temporal) or involved a formant frequency shift (spectral) and adjusted their speech production accordingly. Early auditory processing was shown to be lateralised, with right lateralised processing of spectral features and left lateralised processing of temporal features. During adaptation to the vowel formant shifts, activity increased in both the inferior frontal gyrus and superior temporal sulcus, predominantly on the right. During adaptation to the delayed feedback, activity increased in both the superior temporal gyrus and superior temporal sulcus, predominantly on the left. Increased brain activity was also observed in the inferior frontal gyrus on the left for temporally altered feedback and on the right for spectrally altered feedback. The DIVA model proposes a feedback control system in the superior temporal gyrus bilaterally that compares auditory feedback with the sensory predictions from a left-lateralised feedforward system in the frontal cortex. Here, Floegel et al. (2020) proposed a right-hemisphere preference for spectrally tuned and a left-hemisphere preference for temporally tuned processes in the superior temporal gyrus involved in detecting mismatches between intended and produced speech sounds. Potential mismatches in either domain are thought to be transmitted via temporo-frontal connections and translated into corrective motor commands for the timing of articulator movement in the left inferior frontal gyrus and the position of articulator movement in the right inferior frontal gyrus. This finding differs from the general view of a right lateralised processing of feedback control and a left lateralised processing of feedforward control as in Tourville et al. (2008).

Floegel et al. (2020) also compared speech adaptation-induced changes in resting-state brain connectivity. Results showed that the more speakers adapted to formant perturbation, the more they increased fronto-temporal connectivity in the right hemisphere and temporo-parietal connectivity in the left.

EEG and phase coherence analysis provides evidence about timing and coordination across brain regions. During altered auditory feedback, theta-gamma phase coherence changed over the course of adaptation, increasing over central regions and decreasing over fronto-temporal regions (Sengupta & Nasir, 2015). These changes were correlated with adaptation magnitude in the late phase of speech adaptation. Coherence was positively correlated with the magnitude of speech adaptation in the central area but negatively correlated with the magnitude of speech adaptation in the frontal area. This redistribution of coherence was interpreted as the establishment of a new sensorimotor mapping. Related work showed that theta-, beta-, and gamma-band activity in the late phase of adaptation was linked to whether speakers adapted to the perturbation, and that these rhythms interacted with one another rather than operating independently (Sengupta & Nasir, 2016). ERP evidence has also linked adaptation to auditory prediction: larger compensation for an F1 shift was associated with greater speaking-induced suppression for P2 (Sato & Shiller, 2018). This was interpreted as an auditory prediction during speech adaptation.

More recently, MEG has been used to examine sensory prediction error directly. Speaking-induced suppression was measured as an index of auditory prediction error (Kim et al., 2025). Researchers found that speaking-induced suppression in auditory cortex changed across the course of speech adaptation, becoming reduced (indicating larger prediction error) during early adaptation relative to the initial

unaltered-feedback phase. This reduction was positively correlated with behavioural speech adaptation, indicating that larger prediction errors were associated with greater speech adaptation.

In summary, neuroimaging and electrophysiological studies show that speech adaptation relies on a distributed sensorimotor network. Auditory cortex detects mismatch, frontal and motor regions support correction, and changes in connectivity and oscillatory activity track the formation of new sensorimotor mappings.

2.2.2 Disrupted speech adaptation in people with developmental disorders of speech and language

Speech adaptation has been tested in developmental disorders of speech, especially developmental stuttering and childhood apraxia of speech. Stuttering is widely thought to involve atypical sensorimotor integration in speech motor control, which is why it has been a central focus of auditory perturbation research. Using unexpected formant shifts, weaker compensation was found in people who stutter than in fluent speakers, despite comparable response latency across groups (Cai et al., 2012). They interpreted the reduced magnitude of compensation as evidence for a weak inverse model, i.e. a reduced ability to translate auditory error signals into appropriate updates of speech motor commands. Smaller adaptation effects were also reported together with abnormal EEG phase coherence in adults who stutter, suggesting a weakened coordination between sensory and motor regions, consistent with impaired sensorimotor integration (Sengupta et al., 2016). Reduced adaptation in adults who stutter, but not in children who stutter, has also been reported,

suggesting that the deficit may reflect an age-dependent compensatory change in speech motor control rather than a fixed early impairment (Daliri et al., 2017).

A related line of evidence comes from developmental speech disorders that are also characterised by impaired speech motor planning and integration. Computational simulations of childhood apraxia of speech (CAS) using the DIVA model have shown that weakened feedforward control causes the system to over-rely on auditory feedback, producing speech output consistent with the deficits observed clinically (Terband et al., 2009). Consistent with this prediction, empirical work using a paradigm similar to that employed with adults and children who stutter (Daliri et al., 2017) found that children with speech sound disorders tended to follow an auditory perturbation rather than compensate for it, in contrast to age-matched controls (Terband et al., 2014). This pattern suggests that many children with speech disorders can detect the mismatch between expected and heard feedback, but have difficulty generating the compensatory update needed to oppose the perturbation. Taken together, the simulation and behavioural findings indicate that atypical responses to altered auditory feedback in these populations may reflect an increased reliance on feedback control as a result of weaker feedforward control.

A further related line of work comes from developmental dyslexia. Adults with dyslexia showed increased responses to altered auditory feedback, suggesting a weaker sensorimotor integration and less typical coupling between phonological categories and speech-motor control (van den Bunt et al., 2017). In children with dyslexia, individual differences in reading and reading-related skills were associated with the response to altered feedback (van den Bunt et al., 2018). These findings indicate that atypical speech adaptation in dyslexia is not only a reading-related

problem, but also reflects differences in how auditory speech feedback is mapped onto phonological representations.

The behavioural patterns observed in children and adults with speech and language disorders are consistent with the broader neurobiological view that speech motor control depends on distributed corticocerebellar networks. The cerebellum has been proposed to support speech by helping coordinate prediction, timing, and signal transfer within these networks (Bernard, 2022; Schmahmann et al., 2019). Abnormal white matter microstructure of the cerebellar peduncles has been reported in people who stutter (Connally et al., 2014), and meta-analytic work has proposed that disrupted cortico-subcortical connectivity may contribute to abnormal cerebellar activity in stuttering (Belyk et al., 2014; Brown et al., 2005). In this context, differences in speech adaptation during altered auditory feedback may reflect not only atypical local computations within sensory or motor regions, but also reduced efficiency in the corticocerebellar interactions required to integrate auditory error signals with ongoing speech motor commands.

2.2.3 Pathology and structural brain imaging studies

Given the roles of the cerebellum and basal ganglia in speech motor control, the speech adaptation paradigm has also been used to test in populations with atypical cerebellar and basal ganglia function. In individuals with cerebellar degeneration, adaptation to sustained pitch or formant perturbations is reduced, whereas compensation to unexpected perturbations is often enhanced (Houde et al., 2019; Li et al., 2019; Parrell et al., 2017). This dissociation suggests that cerebellar dysfunction affects the formation and updating of predictive motor commands in feedforward

control, limiting learning from stable sensory errors. When the perturbation is unexpected, however, individuals with cerebellar degeneration may rely more heavily on feedback control to stabilise speech output, which can produce larger online corrections (Houde et al., 2019). Overall, the pattern points to a shift in control strategy: reduced predictive adaptation together with increased dependence on moment-to-moment sensory feedback.

The basal ganglia also contribute to motor planning and execution across human behaviour, including speech. In Parkinson's disease (PD), basal ganglia dysfunction has been linked to impaired sensorimotor integration. Reduced formant adaptation was reported in individuals with PD during sustained altered auditory feedback, compared with age-matched controls (Mollaei et al., 2013). This suggests that basal ganglia dysfunction may hinder the formation of new sensorimotor mappings needed for speech adaptation.

Adaptation effects in PD also appear to differ across speech parameters. Larger compensatory responses to pitch perturbations, but smaller compensation to formant perturbations, were reported in individuals with PD relative to controls (Mollaei et al., 2016). This dissociation has been interpreted as a shift in sensory weighting: pitch control may depend more on auditory feedback when somatosensory information from the larynx is less reliable, whereas formant control may be limited by impaired activation and coordination of the oral articulators.

Further studies of patients with Parkinson's disease indicate that evidence is mixed once medication status is considered. No group differences in formant adaptation were found when individuals with PD were tested on their usual dopaminergic

treatment schedule (Abur et al., 2021). This suggests that auditory–motor learning for formant control can be preserved under dopaminergic medication. Similar evidence was reported for pitch perturbation, where larger compensations were observed overall in patients with PD relative to controls, but these compensatory responses were reduced when an external auditory cue preceded each trial (Huang et al., 2019).

Deep brain stimulation (DBS) of the subthalamic nucleus (STN) has recently been tested as a factor influencing speech adaptation in PD. One study reported reduced compensation to a downward pitch shift with DBS-on relative to DBS-off (Behroozmand et al., 2019), whereas another found no effect of DBS on compensatory responses to pitch perturbation (Bahners et al., 2020). These mixed findings suggest that the influence of DBS on speech adaptation may depend on stimulation parameters, task design, or the specific acoustic dimension being perturbed.

Structural differences in the brain have also been linked to individual variation in speech adaptation. Greater adaptation to pitch perturbation was associated with larger grey matter volume in the left inferior parietal lobe, as well as greater cortical thickness and surface area in the left temporal lobe (Chen et al., 2021). These regions are often implicated in integrating auditory information with speech motor control, suggesting that stronger error-based updating may depend on more developed parietal–temporal structure. By contrast, greater cortical thickness in the right inferior frontal gyrus and superior parietal lobe, and larger surface area in the left precuneus and cuneus, were associated with smaller adaptation magnitudes (Chen et al., 2021). Taken together, these findings suggest that speech adaptation

is supported by a broader brain network, and that variation in that network is reflected in behavioural differences during altered auditory feedback.

2.2.4 Brain stimulation studies

Another important approach to test the role of different brain areas in speech adaptation is to use noninvasive brain stimulation to modify activity and observe the responses on speech adaptation during altered auditory feedback. Specifically, repetitive transcranial magnetic stimulation (rTMS) has been used to negatively interfere with processing in the targeted brain area. In one study, low-frequency rTMS was used to temporarily inhibit the function in the left primary motor cortex (M1) (Tang et al., 2021). Compared with inhibitory rTMS over the hand representation in M1, rTMS over the tongue area completely abolished F1 and F2 adaptation effects in response to an increase in F1 feedback. Reduced adaptation to formant perturbation was also seen when rTMS was used to interfere with processing in the left supramarginal motor gyrus (SMG) (Shum et al., 2011). Continuous theta burst stimulation (c-TBS) over the left supplementary motor area (SMA), combined with EEG recording during pitch perturbation, led to a decrease in compensation and a decrease in N1/P2 response (Dai et al., 2022). These results contradict the finding that larger adaptation magnitude correlated with smaller N1/P2 amplitude (Sato & Shiller, 2018), but support the idea that the left SMA receives auditory feedback and potentially contributes to auditory prediction during speech adaptation. Most recently, a study using c-TBS to disrupt auditory cortex or somatosensory cortex found that retention of speech adaptation effect was impaired when speakers were tested on the following day, whereas disrupting M1 did not impair retention (Rao

et al., 2026). This suggests that retention of the speech adaptation is more likely to be stored in sensory cortex rather than in motor cortex.

Some studies have used transcranial direct current stimulation (tDCS) to modulate neural activity in a brain area and examine the effects on adaptation. Specifically, anodal tDCS facilitates neural activity by increasing the resting membrane potential (membrane depolarisation, (Medeiros et al., 2012)) and typically enhances behavioural performance, whereas cathodal tDCS is thought to down-regulate neural activity by decreasing the resting membrane potential (membrane hyperpolarisation) and has mixed effects on performance (Jacobson et al., 2011). In speech adaptation studies, anodal tDCS over the left inferior parietal lobe increased adaptation to the F1 perturbation relative to cathodal and sham stimulation (Deroche et al., 2017). High-definition anodal tDCS over the left ventral sensorimotor cortex also increased adaptation in F1 during perturbed auditory feedback (Scott et al., 2020). Anodal tDCS over the right cerebellum led to larger pitch compensation for pitch errors in vowel production (Peng et al., 2021).

Among these tDCS studies, Lametti et al. (2018a) investigated the roles of the left speech motor cortex (M1) and the right cerebellum in speech adaptation. The study revealed that anodal tDCS of the left M1 drove speech adaptation towards a previously learned vowel category in the speaker's speech repertoire (adaptation to both F1 and F2), whereas anodal tDCS of the right cerebellum drove correction of the perceived error by offsetting the change in only the shifted formant (adaptation to F1 only). An alternative explanation is that anodal cerebellar tDCS abolished the adaptation in F2, which was seen in the groups who received sham and M1 stimulation. Anodal cerebellar tDCS may impair speech processing. For instance,

interference in the timing of perceptual decisions in speech was observed following anodal tDCS over the right cerebellum (Lametti et al., 2016).

Among studies showing mixed effects of cathodal tDCS, cathodal cerebellar tDCS has been shown to facilitate verb generation (Marangolo et al., 2018) and auditory working memory (Pope & Miall, 2012). In the latter study, the facilitatory effect of cathodal tDCS was attributed to a reduction in the inhibitory influence of the cerebellum on the prefrontal cortex (Pope & Miall, 2012).

Transcranial alternating current stimulation (tACS) is similar to tDCS in that it uses weak electric currents over the scalp at a certain frequency aimed at entraining intrinsic neural oscillations. Demirel et al. (2025) conducted a tACS study, where researchers delivered 75 Hz gamma coupled to the peak of a 6 Hz theta envelope (TGP) over left M1 at an intensity of 2 mA peak-to-peak to enhance speech adaptation during sustained altered auditory feedback. However, no effect of TGP-tACS was observed on the magnitude of speech adaptation compared with sham stimulation. The group receiving active stimulation showed greater retention of the adapted state during "wash-out", suggesting further studies exploring accumulative effects of repeated TGP-tACS would be informative.

Overall, noninvasive stimulation studies show that speech adaptation depends on a distributed network including sensorimotor cortex, SMA, and cerebellum. The effects of stimulation are variable, however, and differ across regions, polarity, and stimulation type.

2.3 Summary

In this chapter, I focused specifically on studies using the paradigm involving speech adaptation during altered auditory feedback. Since the first study of formant perturbation (Houde & Jordan, 1998), a growing body of work has used this paradigm to investigate the behavioural and neural mechanisms underlying speech adaptation. The altered auditory feedback paradigm using formant perturbation has proven particularly informative for understanding speech adaptation, speech motor control, and the relationship between speech production and perception. To date, these findings have demonstrated the importance of auditory feedback in maintaining stable speech production and supporting speech development.

Studies using brain imaging and brain stimulation have identified neural correlates of speech adaptation. Collectively, this work suggests that activity in portions of auditory and sensorimotor cortex, as well as the cerebellum, contribute to adaptive responses during altered feedback. In parallel, research on clinical and developmental populations, including patients with cerebellar degeneration, people who stutter, and individuals with Parkinson's disease, has advanced our understanding of both atypical and typical speech motor control.

However, this review highlights gaps in our knowledge, which remain to be investigated. Firstly, further clarification is needed regarding the impact of brain stimulation on speech adaptation since previous studies have produced mixed effects. Secondly, relatively few studies have used functional imaging to examine the neural correlates of speech adaptation during sustained altered auditory feedback. So far, existing findings are inconsistent regarding the roles of the left and right

hemispheres in speech adaptation. Also, changes in intrinsic brain activity due to speech adaptation have yet to be fully investigated. Comprehensive whole-brain functional brain mapping would deepen our understanding of the neural correlates of speech adaptation, strengthen models of speech motor control, and inform the design of future brain stimulation studies.

Here, these knowledge gaps were addressed in separate studies using brain stimulation and brain imaging.

First, in a pre-registered study, I aimed to replicate the effects of anodal tDCS on speech adaptation over the left speech motor cortex and the right cerebellum described by Lametti et al. (2018a). The results of a large study in ninety participants failed to replicate the original findings. These results are described in Chapter 3 and published in Yuan et al. (2025). Second, since nearly all previous work on the speech adaptation paradigm has involved English or French speaking participants, I developed a Mandarin version of the task. I first tested this task behaviourally in seventy-six participants and then adapted it for the use in task functional MRI (forty-eight participants). These data are presented in Chapter 4. I acquired resting-state fMRI before and after the speech adaptation task to investigate long-term changes in intrinsic brain networks. The results of the analysis of the resting-state data are presented in Chapter 5.

3

Enhancing speech adaptation using transcranial direct current stimulation

3.1 Abstract

Transcranial direct current stimulation (tDCS) modulates cortical excitability and when applied in combination with a cognitive task has potential to enhance performance. In the speech domain, previous work indicated that anodal tDCS over left motor cortex and right cerebellum increased the magnitude of sensorimotor control of speech (will also be called "speech adaptation" later in this chapter). Here, we aimed to replicate these findings in a pre-registered, double-blind, randomised, sham-controlled study of a large sample (three groups of N = 30). Participants

read words out loud. Speech was recorded and fed back to them either normally or with a 110-Mel increase in the frequency of the first vowel formant. Participants responded to altered feedback by changing their speech production (adaptation). Participants were randomly allocated to receive 2-mA anodal tDCS over either left speech motor cortex, or right cerebellum, or sham stimulation. We tested for differences in speech adaptation among the three groups using one-way analyses of variance. We also explored the relationship between speech adaptation and measures of speech perception. All groups showed significant adaptation while receiving altered auditory feedback. Contrary to the previous study, we found no impact of anodal tDCS on the magnitude of the speech adaptation. In conclusion, speech adaptation was unaffected by anodal tDCS over speech motor cortex or cerebellum. This study is another example of the inconsistent effects of tDCS on task performance particularly when participants are young and healthy. Even larger samples may be needed to detect small effects and to avoid spurious results arising from individual differences in task performance.

3.2 Introduction

Transcranial direct current stimulation (tDCS) shows promise as an adjunct to therapeutic interventions. In the speech domain, tDCS can improve the outcomes of speech therapy in patients with aphasia due to stroke or neuro-degenerative disease (Coemans et al., 2021; Elsner et al., 2019) and in people who stutter (Chesters et al., 2018; Moein et al., 2022).

TDCS can also enhance performance and improve learning across a range of tasks in neurologically intact participants (e.g. Reis et al. (2009)). Such studies are necessary to optimise and refine stimulation protocols and could inform future clinical applications. By modulating cortical excitability during a speech adaptation task, tDCS was used to test the contributions of different brain areas to sensorimotor learning in speech. Specifically, anodal tDCS, which raises the resting membrane potential of neurons in the cortex (Medeiros et al., 2012; Nitsche & Paulus, 2000) enhanced speech adaptation when applied over left inferior parietal cortex (Deroche et al., 2017), left ventral sensorimotor cortex (Lametti et al., 2018a; Scott et al., 2020), and right cerebellum (Lametti et al., 2018a; Peng et al., 2021).

Here, we sought to replicate the findings of one of these studies demonstrating different roles for the left motor cortex and the right cerebellum in speech adaptation (Lametti et al., 2018a). Speech adaptation is studied using a altered auditory feedback paradigm (Houde & Jordan, 1998). Participants produce syllables, which are recorded and fed back to them in near real-time. Speech adaptation is evaluated by altering the frequencies of the vowel formants and measuring the effects of such perturbation on subsequent speech production. The frequencies of the first (F1) and second (F2) vowel formants determine vowel identity (Ladefoged & Johnson, 2014). In an experiment, increasing F1 when "head" is produced, results in a word that sounds closer to "had" and causes participants to adapt their speech production to achieve the target. The magnitude of change in speech production during a period of altered feedback and the persistence of the change once normal feedback is restored provide measures of sensorimotor learning.

In the Lametti et al. (2018a) study, anodal tDCS over left motor cortex enhanced sensorimotor adaptation in both F1 and F2 driving adaptation towards the formant frequencies associated with an existing vowel category. In contrast, anodal tDCS over the right cerebellum enhanced adaptation in F1 only; participants corrected the increase in the F1 frequency by reducing the frequency of this formant in their productions, thereby offsetting the sensory error. Anodal cerebellar tDCS appeared to abolish adaptation in F2, which accompanied F1 adaptation in both sham and motor cortex stimulation groups.

In the current study, we aimed to replicate the findings in the Lametti et al. (2018a) study and address some of its limitations. We used a randomised, double-blind, sham-controlled design; methods and analysis were pre-registered (https://osf.io/9bswe). The original study used a cerebellar montage that placed the cathode on the right cheek and secured the electrodes with bandages around the jaw, potentially affecting speech production. To avoid this issue, we used a different cerebellar tDCS montage validated in previous studies (Ferrucci et al., 2013; Weightman et al., 2023): the anode was positioned over the right cerebellum and the cathode was placed on the right deltoid muscle (upper arm). We also increased power by increasing the sample size from 20 to 30 participants per group. We replaced the acoustical effects processor and analogue filters with a MATLAB MEX-based program called 'Audapter' (Cai et al., 2008; Tourville et al., 2013) to record speech, alter it and feed it back to participants with minimal delay. This software allowed us to introduce an additional measure of speech adaptation by masking auditory feedback with noise scaled to speech volume; measures

of F1 and F2 production during blocks of noise-masked feedback were acquired before and after adaptation.

Previous studies showed that individual differences in vowel perception affect the amount of sensorimotor adaptation during speech production (Villacorta et al., 2007), and that those who show larger adaptation effects also show larger changes in perception (Lametti et al., 2014a). To explore whether anodal tDCS over left motor cortex or right cerebellum would further enhance these effects, we included measures of categorical perception of vowels on the "head" to "had" continuum before and after adaptation.

3.3 Materials and Methods

3.3.1 Participants and study design

We recruited 90 native English speakers aged between 18 and 47 years; 45 female, 45 male, 80 right-handers. Most were students at the University of Oxford and were given either class credits or a small monetary reimbursement for participation. All were naïve to the purpose of the experiment. Exclusion criteria included a history of or current neurological disorder, speech or hearing impairments, and tDCS contraindications.

The experimental setup is shown in Figure 3.1 A. Participants wore over-ear headphones and a head-mounted microphone (SHURE WH20) and sat in front of a computer screen where words were displayed for reading aloud. Speech was recorded and feedback altered using a MATLAB-MEX based interface named "Audapter" (Cai et al., 2008; Tourville et al., 2013). Brain stimulation was applied over

either the left motor cortex or right cerebellum using a direct current stimulator (NeuroConn DC Stimulator Plus; Rogue Resolutions, Cardiff, UK) running in remote mode to ensure the double-blind sham-controlled design. A researcher who was not involved in data collection or analysis allocated participants randomly into the three groups (anodal-M1, anodal-cerebellum, sham). This was done separately for males and females to ensure all groups were balanced for sex. Participants received a unique five-digit code that delivered either active or sham stimulation when entered into the stimulator. The researchers and participants were unaware of the stimulation group assignment.

The brain stimulation started at the beginning of the baseline block (normal feedback), lasted 16 minutes, and ended before the start of the post-adaptation perception task. Measures of speech perception were acquired before baseline and after the adaptation block of the speech production task.

3.3.2 Sample size justification

The primary objective of the study was to replicate the findings from a previous study by Lametti et al. (2018a). In that study, one-way analyses of variance (ANOVA) were used to compare changes in F1 and F2 in three groups of 20 participants. A significant effect of stimulation group was found for F1 during the last 30 trials of the adaptation block (f = 0.395, personal communication) and for F2 during the first 15 trials of the after-effect block (when normal feedback resumed; f = 0.424). Using the smaller effect size, we determined that a sample size of 87 participants would give us > 90% power to detect differences among three groups with an alpha of 0.05 (G*Power software, Version 3.1, Faul et al., 2009).

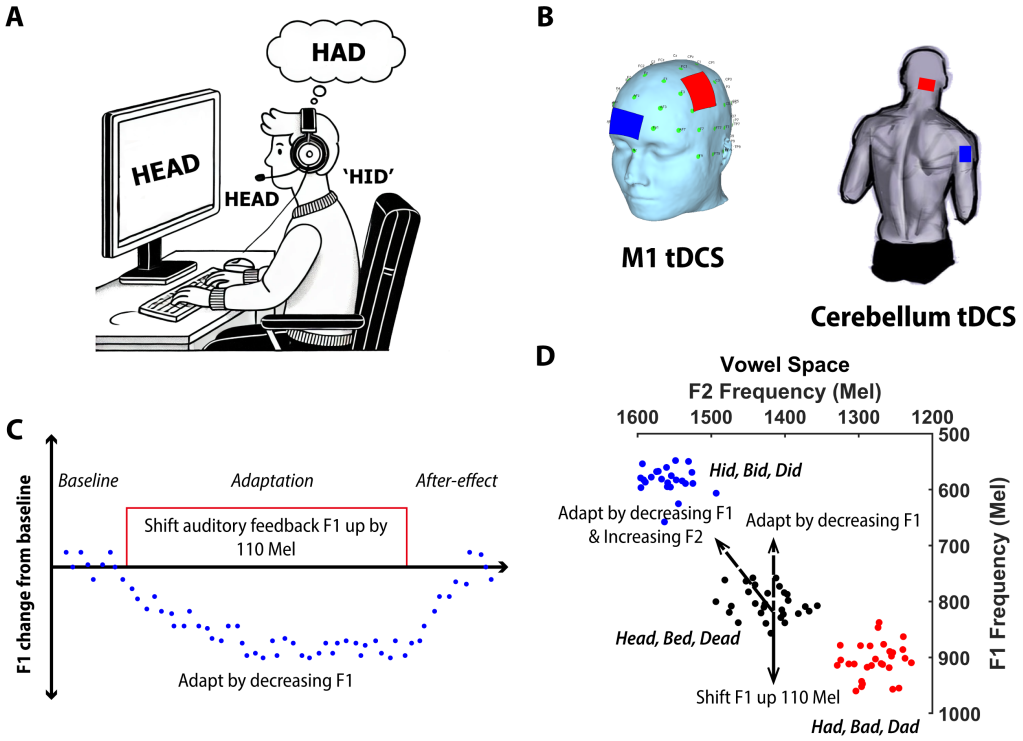

Figure 3.1: Speech Production Task.
A. Experimental setup (the figure was generated with the help of ChatGPT). **B.** Electrode montages for anodal-M1 and anodal-cerebellum stimulation. Red – anode, blue – cathode. **C.** Schematic of predicted change in F1 from baseline during an increase in F1 feedback in the adaptation block. Blue dots – F1 frequency on individual trials. **D.** Vowel space showing vowel production data in a single participant. F1 and F2 frequencies of each utterance are plotted in black for the vowel in "head", blue for the vowel in "hid" and red for the vowel in "had". The downwards black arrow indicates the direction of the shift applied to F1 feedback during the adaptation block. The two dashed arrows illustrate two alternative patterns of change participants were expected to make in response to the upward shift in F1 feedback: (i) offset upward shift by decreasing F1 production with no change in F2; (ii) coupled changes in F1 and F2 production to move towards the vowel in "hid".

3.3.3 Speech production task

The speech production task consisted of seven blocks: vowel exploration, baseline, adaptation, after-effect, and three noise-masked blocks. In each, participants read words presented for 1500-ms with 750-ms intervals. Recordings of these utterances were made via a head-mounted microphone and fed back via over-ear headphones (see Figure 3.1 A). In the vowel exploration block, participants produced nine words

("dead", "bed", "head", "dad", "bad", "had", "did", "bid", and "hid") presented in a random order, 10 times each and heard unaltered feedback. During baseline, adaptation and after-effect blocks, "head", "bed", and "dead" were each produced 15, 75 and 15 times, respectively. Breaks of 30 seconds occurred every 45 utterances. The formant alterations were applied using the Mel scale-a perceptually normalised scale widely used to represent frequency changes that listeners perceive as equal steps in pitch (Stevens et al., 1937). The magnitude of the shift applied in the original study was 21% (in Hz; Lametti et al. (2018a)), which converts to approximately 110-Mel. The conversion from Hz to Mel used the following formula:

$$\text{Mel} = 2595 \cdot \log_{10}\left(1 + \frac{\text{Hz}}{700}\right)$$

Feedback was altered during the adaptation block only. F1 was isolated from the utterance and increased by 110 Mel, combined with the rest of the unaltered speech signal and 60 dB masking pink noise, and played back through headphones with an imperceptible delay. In response to the altered auditory feedback, participants typically adjust their forthcoming utterances. For example, participants utter "head" but with the increase in F1, the word fed back over headphones sounds closer to the frequency of the vowel in "had". Participants typically adapt by reducing F1 and increasing F2 in their production (see Figure 3.1 C & D). The noise-masked blocks involved producing "head", "bed", and "dead" five times each with noise feedback scaled with speech volume (the feedback was 0 dB). One block was used for familiarisation immediately after the vowel exploration. The second and third noise-masked blocks took place before and after the adaptation block.

3.3.4 Analysis of speech production data

For all utterances, formant frequencies were extracted and analysed using MATLAB (Tang et al., 2021). Specifically, the average F1 and F2 frequencies were extracted across a 25-ms time frame at the centre of each vowel using Praat (Boersma & Weenink, 2022) with linear predictive coding. F1 and F2 values exceeding three standard deviations from the mean in each block were excluded (F1: 1.69% and F2: 1.73% of the data). Formant frequency changes were calculated for all trials in the baseline, adaptation, and after-effect blocks by subtracting the average value of the baseline for each formant in each participant. The change in formant frequency for the post-adaptation noise-masked block was calculated by subtracting the average value from the pre-adaptation noise-masked block.

3.3.5 Transcranial direct-current stimulation (tDCS)

Participants received tDCS for 16 minutes over either the left ventral sensorimotor cortex (M1) or right cerebellum during the baseline, pre-adaptation noise, adaptation, and post-adaptation noise blocks of the speech production task. A 2-mA direct current with 30-second ramp-up and ramp-down periods was delivered using 5 x 7 cm saline-soaked electrodes. For M1 stimulation, the anode was placed in portrait (short side horizontal) with the top posterior corner of the electrode located on a point one-third of the length along a line from the vertex to the tragus. This point lies between the C1 and C3 electroencephalography electrode locations in the 10-20 system and corresponds to the M1-hand representation (C3h in the 10-5 system; Kim et al., 2023a). Previous mapping of the speech motor system using MRI and found the larynx, lip, and tongue cortical representations lie between 19-

and 37-mm ventral and anterior to the hand representation (Eichert et al., 2021). Thus a 5 x 7 cm electrode oriented in portrait and placed on the scalp ventral and anterior to the hand location should cover this speech motor cortex. The cathode was placed in landscape (long side horizontal) above the right supraorbital ridge, not crossing the midline. For cerebellar stimulation, the anode was centred on a point 3-cm lateral to the inion on the right in landscape orientation, and the cathode was placed on the right upper arm over the deltoid muscle in portrait configuration (see Figure 3.1 B). The sham stimulation group received 90 seconds of stimulation including 30-second ramp-up and ramp-down periods. Electrodes were placed at the same positions and orientations as for the anodal stimulation groups. Half of the sham group received M1 sham stimulation, and the other half received cerebellar sham stimulation.

3.3.6 Speech perception tasks

For the speech perception tasks, an individually tailored vowel formant continuum ("head" to "had") was generated by recording the participant's production of "head" and "had" 10 times each. The average F1 and F2 frequencies were extracted as described above. We selected the two utterances (one for 'head' and one for 'had') closest in Euclidean distance to the average of each in F1-F2 vowel space and used a Praat script (http://www.mattwinn.com/praat.html) to generate an 8-step "head" to "had" continuum. These stimuli were then used in identification and discrimination tests of categorical speech perception, which were completed before baseline and after adaptation block of the speech production task.

In the identification task, eight stimuli, one from each step of the continuum, were presented 12 times in a random order with a 1500-ms stimulus onset asynchrony (SOA). Participants pressed a button to identify which word they heard ("head" or "had"). In the discrimination task, participants heard identical pairs (one pair for each of the 8 steps) or different pairs (pairs 1-3, 2-4, 3-5, 4-6, 5-7, 6-8) separated by two steps. The identical pairs were presented six times, and the different pairs were presented 12 times in a counterbalanced and randomised sequence. Words in a pair had a 500-ms SOA, and trials had a 2500-ms interval. Participants indicated with a button press whether pairs were the same or different.

3.3.7 Analysis of speech perception data

For the identification task, we first removed anticipatory responses (i.e. reaction time shorter than 200 ms) and calculated the proportion of "head" responses for each of the 8 steps of the continuum. Further data was lost due to technical issues and participants pressing the wrong buttons (percentage of total data loss in the identification task is about 6%). A logistic curve was fitted to the identification data for each participant in each session (before and after the adaptation) using the following formula:

$$E(Y_t) = \left(1 + \beta_0\, \beta_1^t\right)^{-1}$$

The position of the category boundary was defined as the point along the eight-step continuum where E(0.5). The logarithm of β_1 was used to estimate its slope.

For the discrimination task, we first removed data from trials where responses

were considered anticipatory (i.e. reaction time shorter than 200-ms after the onset of the second sound of the pair). There was further data loss due to technical issues and participant error (percentage of data loss in the discrimination task is about 18%). The proportion of 'different' responses was calculated for all pairs of stimuli. Responses in the identification task were used to determine whether the 'different' pairs in the discrimination task crossed the category boundary (i.e. between-category pairs). If the difference in the proportion of "head" responses to each stimulus in a pair was more than 0.5, the pair was classified as a between-category pair.

For each participant in each session, we calculated the hit rate (proportion of 'different' responses to different pairs; both between and within-category pairs) and false alarm rate (proportion of 'different' responses to identical pairs). Sensitivity (d') and response bias (c) were calculated using signal detection theory. Since our design involved a roving stimulus level and only a few repetitions for each pair of stimuli, we calculated each individual's sensitivity (d') and response bias (c) using the difference rule and their pooled proportions of the same and different responses for their between-category pairs instead of estimating for each pair separately and averaging across (see Macmillan et al. (2021), chapter 13, p. 316, for details).

3.3.8 Statistical analyses

For the speech production tasks, following Lametti et al. (2018a), the magnitude of adaptation was quantified as the average F1 and F2 changes from baseline in the last 30 trials of the adaptation block, and the first 15 trials in the after-effect block. Any data points outside 1.5 times the interquartile range were removed from the analysis (one data point from F2 change in adaptation block and one

from F1 change in noise block, four data points from F2 change in noise block). To replicate the analyses of the previous study, one-way ANOVAs were performed to test the effect of stimulation group (anodal-M1, anodal-cerebellum, sham) on F1 and F2 changes at the end of adaptation (last 30 trials) and in the after-effect block (first 15 trials). In addition, we compared F1 and F2 changes from the pre- to the post-adaptation noise block among the three groups using one-way ANOVA.

For the speech perception data, a mixed ANOVA with one within-subject factor (pre vs. post-adaptation) and one between-subject factor (group: anodal-M1, anodal-cerebellum, sham) was used to test for significant changes in the slopes and positions of category boundaries, hit and false alarm rates, sensitivity (d'), and response bias (c) after the speech production task and the effects of stimulation on these changes. To test for the relationship between speech perception abilities and adaptation, correlation analyses were conducted between baseline perceptual measures (the slopes and positions of category boundaries, hit rates, false alarm rates, d', and c) and the magnitude of speech adaptation (average F1 and F2 changes from baseline for last 30 trials). Correlation analysis was also conducted to explore the relationship between the magnitude of speech adaptation and the change in perceptual measures from before to after adaptation.

For the calculation of effect sizes, we reported Cohen's d to quantify the standardised mean difference in one-sample t-tests ($d_z = \frac{t}{\sqrt{N}}$). For one-way ANOVAs, we reported partial eta squared (η_p^2), which is an estimate of the variance explained by the factor.

3.4 Results

The main aim of this double-blind sham-controlled study was to replicate findings from a previous study that tDCS over the left motor cortex and right cerebellum enhances sensorimotor learning for speech (Lametti et al., 2018a). Ninety participants completed a speech adaptation task while receiving either anodal or sham stimulation over the left motor cortex or right cerebellum. Participants were randomly allocated to three groups of thirty participants maintaining an even sex balance in each: anodal-M1, anodal-cerebellum and sham groups. Groups did not differ in age ($F(2,87) = 1.33$, $p = 0.270$) or in terms of baseline F1 ($F(2,87) = 0.655$, $p = 0.522$) and F2 ($F(2,87) = 0.016$, $p = 0.984$) frequencies during speech production (see Table 3.1).

Table 3.1: Participant characteristics and changes in F1 production during the adaptation task.

Values are shown for the M1, cerebellum, sham, and full sample groups. Age and baseline formant values are shown as mean ± SD.

Measure	**M1**	**Cerebellum**	**Sham**	**All**
Participant characteristics				
N	30	30	30	90
Sex	15M, 15F	15M, 15F	15M, 15F	45M, 45F
Age (years)	23.5 ± 8.17	21.5 ± 4.89	21.4 ± 3.25	22.1 ± 5.83
Baseline formant frequencies				
Baseline F1 (Mel)	784 ± 75.2	768 ± 59.2	786 ± 67.6	779 ± 67.4
Baseline F2 (Mel)	1459 ± 60.4	1456 ± 47.9	1458 ± 67.4	1458 ± 58.5
Change in F1 production (% of F1 shift)				
End of adaptation	34.3%	37.6%	37.1%	36.3%
Start of after-effect	16.5%	18.3%	27.0%	19.0%
Post-adaptation noise block	22.6%	32.6%	22.0%	27.4%

3.4.1 The effect of tDCS on speech adaptation

Changes from baseline in F1 production during the adaptation and after-effect blocks of the speech production task are shown for each group in Figure 3.2. All three groups adapted F1 during this task. F1 significantly decreased relative to baseline by the end of the adaptation block, and at the start of the after-effect block (one-sample t-test against zero in each group: ps < 0.05; see Table 3.1). The magnitude of these F1 changes did not differ among the three groups, however (one-way ANOVA at the end of adaptation: $F(2,87) = 0.149$, $p = 0.862$, $\eta_p^2 = 0.00341$; start of after-effect block: $F(2,87) = 0.530$, $p = 0.591$, $\eta_p^2 = 0.0120$). These results contrast with those from the original study, which found significantly greater change in F1 from baseline in the two stimulation groups relative to sham. Also in the current study, all three groups retained significant F1 adaptation during the post-adaptation noise-masked block (ps < 0.05, Figure 3.4) but the magnitude of this change from baseline did not differ among the three groups ($F(2,86) = 0.984$, $p = 0.378$, $\eta_p^2 = 0.0224$).

Changes from baseline in F2 production are shown in Figure 3.3. F2 significantly increased relative to baseline at the end of the adaptation block and at the start of the after-effect block in all three groups (one-sample t-test against zero, all ps < 0.05). As seen for F1, the magnitude of the F2 changes did not differ among the three groups (one-way ANOVA at the end of adaptation: $F(2,86) = 0.290$, $p = 0.749$, $\eta_p^2 = 0.00670$; start of after-effect block: $F(2,87) = 0.334$, $p = 0.717$, $\eta_p^2 = 0.00762$). As above, these results contrast with those from the original study, which found no significant adaptation in F2 in the anodal-cerebellar group at the end of

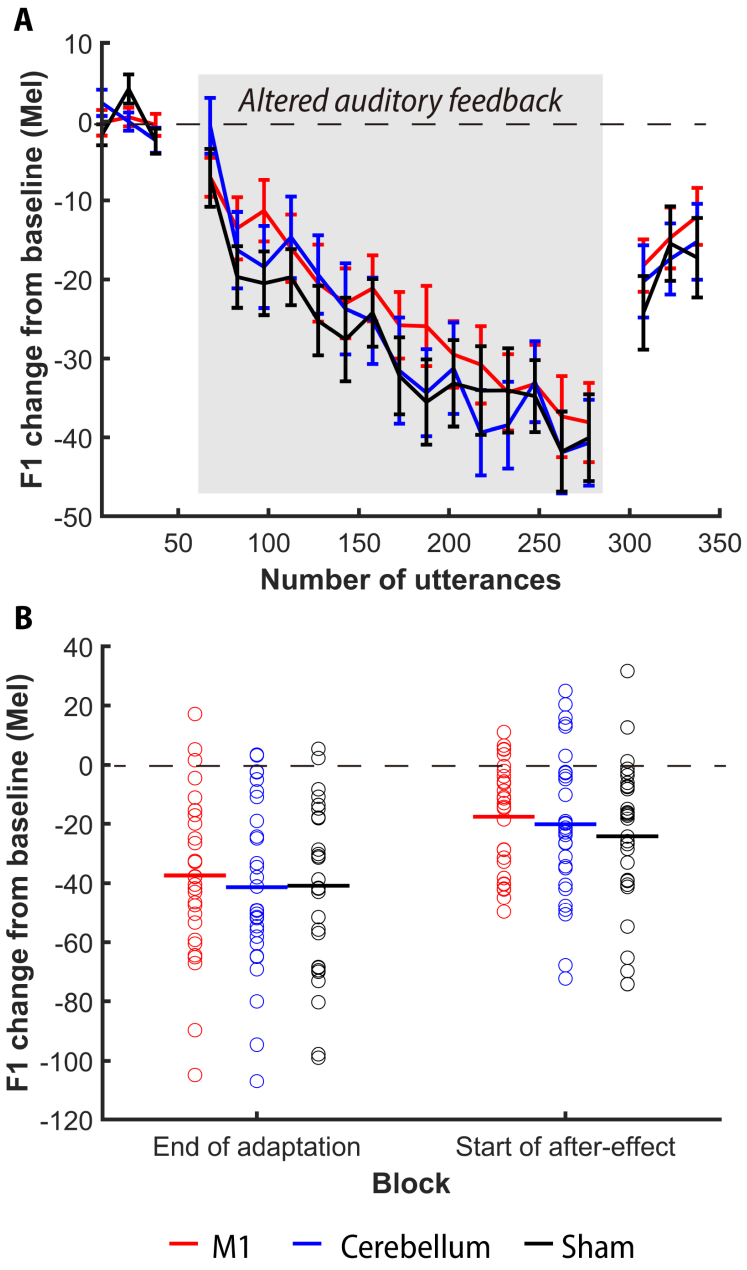

Figure 3.2: Changes in F1 production during adaptation task.
A. Data averaged every 15 trials for baseline, adaptation, and after-effect blocks. Adaptation block with altered auditory feedback is marked in grey. Error bars show ± standard error. B. Average F1 change for individual participants (open circles) for each group at the end of the adaptation block (last 30 trials with altered feedback), and initial 15 trials of the after-effect block. Horizontal lines indicate the mean for the group. Red – anodal-M1; Blue – anodal-cerebellum; Black – sham. All changes from baseline were significant but did not differ among groups (see text for details).

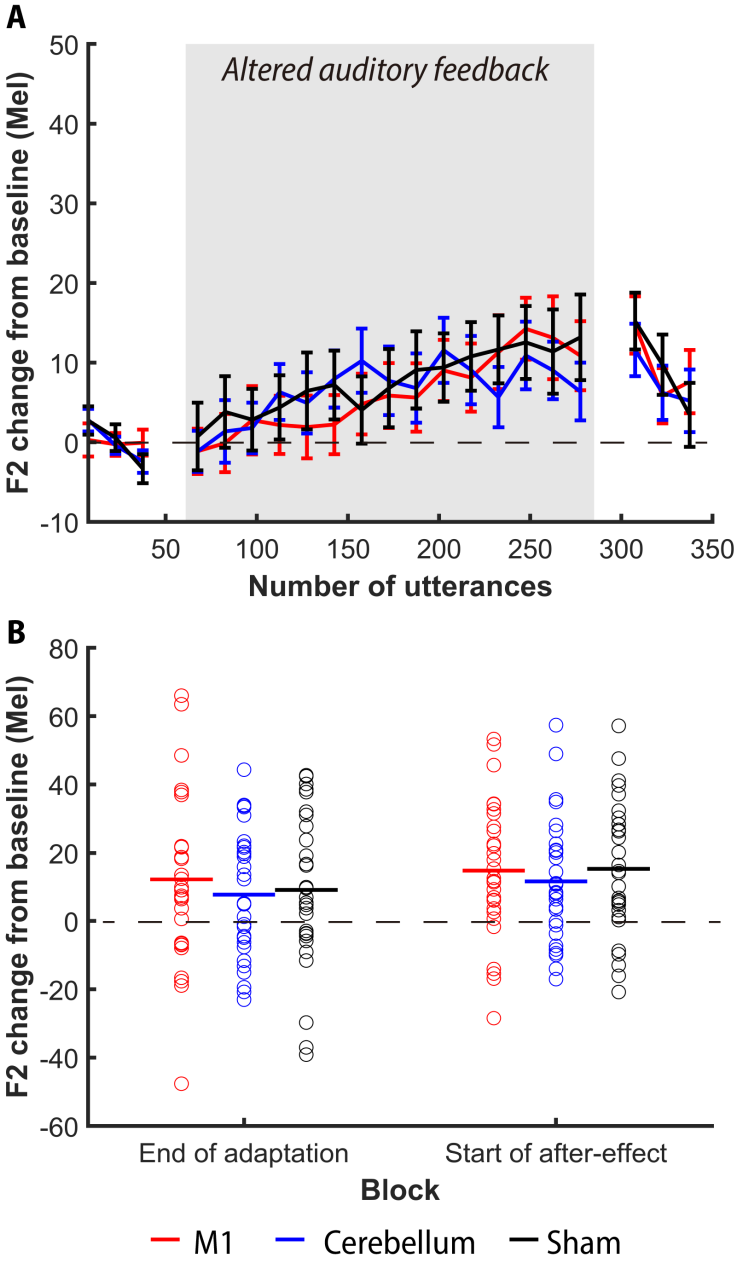

Figure 3.3: Changes in F2 production during adaptation task.
See caption to Figure 3.2 for details. All changes from baseline were significant but did not differ among groups (see text for details).

adaptation or start of after-effect blocks and a significantly greater change in F2 from baseline in the anodal-M1 and sham groups relative to the anodal-cerebellar group for the after-effect block. For the post-adaptation noise-masked block only the anodal-cerebellum and the sham groups showed significant changes in F2 (*p*s

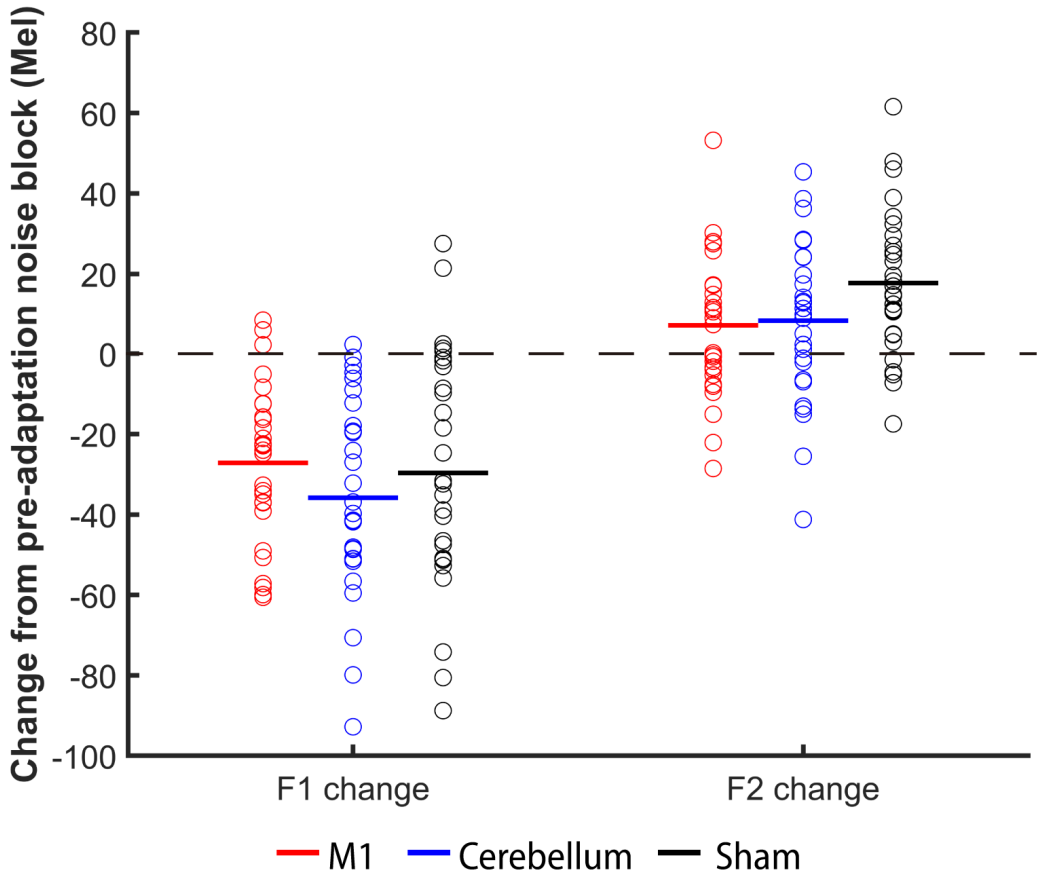

Figure 3.4: Changes in F1 and F2 production from the pre- to the post-adaptation noise block.
See caption to Figure 3.2 for details. All changes were significant for F1 and did not differ among groups. For F2, only cerebellum and sham stimulation groups showed significant adaptation (see text for details).

< 0.05); the change in F2 for the anodal-M1 group was not significantly different from zero ($t(29) = 1.35$, $p = 0.188$, Cohen's $d = 0.246$). However, the ANOVA found no group differences in the magnitude of F2 change ($F(2,82) = 2.81$, $p = 0.066$, $\eta_p^2 = 0.0641$; Figure 3.4).

3.4.2 Exploratory analysis

We performed an exploratory analysis of the data in the adaptation block of the experiment using linear mixed-effects modelling by adding the random effects of

participant and word into the model using group (M1, cerebellum, & sham) and trial (number of utterances) as fixed effects to predict the changes over time in speech adaptation. We did not include random slopes for participants, because this model failed to converge, presumably due to sample size.

The analysis was conducted using the lme4 package (Bates et al., 2015) and the significance was estimated using the lmerTest package (Kuznetsova et al., 2017) in R. First, participants' F1 and F2 for the adaptation trials were normalised by subtracting the average F1 and F2 at baseline. We created two new categorical variables: Group 1 (2/3, −1/3, −1/3; refers to the group mean of M1 stimulation relative to the group mean of sham stimulation) and Group 2 (−1/3, 2/3, −1/3; refers to the group mean of cerebellum stimulation relative to the group mean of sham stimulation) to estimate the group differences. Analysis of pairwise comparisons of estimated marginal means was conducted using the emmeans package (Lenth et al., 2023).

Linear mixed-effects modelling showed that all three groups significantly adjusted their F1 and F2 in response to the altered feedback (significant main effect of time, F1: $t = -45.9$, $p < 0.001$; F2: $t = 16.5$, $p < 0.001$). The M1 tDCS group were slower to adapt in both F1 and F2 compared with the sham group (significant interaction between group 1 (M1 vs sham) and time: F1: $t = -4.49$, $p < 0.001$; F2: $t = -2.36$, $p = 0.018$). However, pairwise comparisons between groups were not significant.

Therefore, although there is evidence suggesting that M1 tDCS might lead to slower speech adaptation compared with the sham stimulation (see Figure 3.2), this result needs further investigation in larger samples of participants.

3.4.3 The effect of tDCS on speech perception

For the speech perception tasks, there were no unexpected differences among the three groups in their baseline (pre-adaptation) measures derived from the identification and discrimination tests of categorical perception (see Table 3.2). There were no differences among the three groups for the post-adaptation measures too (i.e. no main effect of group or interaction).

Post-adaptation, there was a significant change in the position of the category boundary for the identification function; people were more likely to report "had" post-adaptation compared with pre-adaptation; $F(1,80) = 29.2$, $p < 0.001$, $\eta_p^2 = 0.267$, Figure 3.5 A). The slope of the logistic curve increased numerically in all groups from pre- to post-adaptation but this change was not significant ($F(1,77) = 3.46$, $p = 0.067$, $\eta_p^2 = 0.0430$). For the discrimination task, hit rate decreased from pre- to post-adaptation ($F(1,71) = 11.8$, $p < 0.001$, $\eta_p^2 = 0.143$, Figure 3.5 B) as did the rate of false alarms ($F(1,71) = 4.14$, $p = 0.046$, $\eta_p^2 = 0.0551$, Figure 3.5 B), consistent with a reduction in response bias, which was also significantly changed (i.e. the liberal bias shown by participants pre-adaptation was reduced; participants were slightly less likely to say "different" to all pairs resulting in fewer hits but also fewer false alarms, $F(1,69) = 17.1$, $p < 0.001$, $\eta_p^2 = 0.200$, Figure 3.5 B). This was not due to any change in sensitivity since d' was unchanged from pre- to post-adaptation ($F(1,71) = 0.674$, $p = 0.414$, $\eta_p^2 = 0.00940$, Figure 3.5 B). In sum, we found that performing the adaptation task increased the size of the category for "head" relative to "had" compared with the category sizes before adaptation. Participants also became less liberal in their responding after the

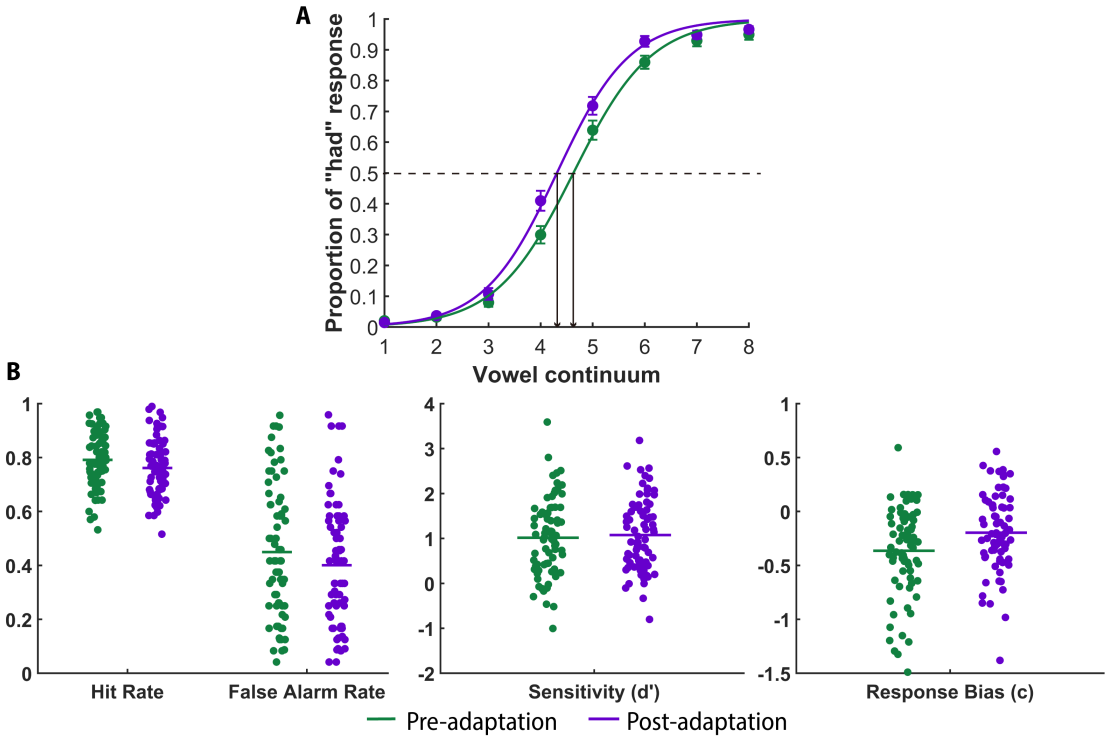

Figure 3.5: Changes in speech perception measures from before to after speech adaptation.

A. Averaged logistic functions for the identification task showing a shift in the vowel boundary between "head" and "had" on the identification task. Circles represent the proportion of "had" response on each vowel continuum. The vowel boundary shifted from pre- to post-adaptation (indicated by down arrows) in the direction of 'head'. Error bars show +/− standard error. **B.** Hit rate, false alarm rate, sensitivity (d'), and response bias (c) for individual participants (circles) on the discrimination task. Horizontal lines indicate the mean for the session. Green – pre-adaptation session; Purple – post-adaptation session. Means and standard errors of each session and of each stimulation group are provided in Table 3.2.

adaptation task, exhibiting an increase in their criterion for signal detection on the discrimination test, which resulted in fewer hits but also fewer false alarms. As for speech production, however, there were no differences among groups in any of these speech perception measures; stimulation had no effect.

3.4.4 Correlation between perception and production

We found no relationship between performance measures of speech perception at baseline and the amount of adaptation in the speech production task (all correla-

Table 3.2: Behavioural measures of perception tasks.

Means±SEs shown.

Group	*df*	M1	Cerebellum	Sham	All
Position					
Pre	86	4.72±0.127	4.44±0.157	4.58±0.129	4.59±0.0790
Post	85	4.53±0.135	4.00±0.151	4.28±0.126	4.28±0.0815
Slope					
Pre	85	14.9±2.30	16.6±2.76	15.8±2.74	15.7±1.49
Post	84	23.7±5.22	25.0±7.18	18.5±3.43	22.3±3.10
Hits					
Pre	74	0.824±0.0214	0.777±0.0236	0.779±0.140	0.791±0.0116
Post	74	0.807±0.0284	0.742±0.0216	0.744±0.150	0.761±0.0126
False Alarms					
Pre	74	0.484±0.0633	0.430±0.0463	0.440±0.0434	0.449±0.0290
Post	74	0.367±0.0617	0.416±0.0468	0.412±0.0354	0.401±0.0270
Sensitivity (d′)					
Pre	74	1.06±0.231	1.03±0.171	0.966±0.137	1.02±0.101
Post	74	1.43±0.194	0.940±0.165	0.931±0.118	1.07±0.0932
Response Bias (c)					
Pre	74	−0.485±0.109	−0.327±0.0811	−0.307±0.0638	−0.364±0.0483
Post	72	−0.134±0.097	−0.228±0.0790	−0.211±0.0545	−0.197±0.0431

tions between measures were not significant, see Table 3.3 for statistics). Similarly, the magnitude of adaptation during the speech production task was unrelated to the changes from pre- to post-adaptation in the various measures of speech perception (all correlations between measures were not significant, see Table 3.4 for statistics). These findings contrast with those reported in previous studies (see introduction for details). Due to the lack of significant relationships between production and perception measures, we did not explore any modulatory effects due to stimulation (i.e. group differences).

Table 3.3: Correlation between speech perception at baseline and F1/F2 change in the last 30 trials of the adaptation block.

Baseline perception	**N**	***r***	***p***
F1 change			
Position	86	−0.208	0.054
Slope	85	0.004	0.972
Hits	74	−0.100	0.395
False Alarms	74	0.017	0.883
Sensitivity (d')	74	−0.052	0.660
Response Bias (c)	74	0.017	0.886
F2 change			
Position	86	−0.196	0.070
Slope	85	0.081	0.461
Hits	74	0.209	0.074
False Alarms	74	0.055	0.641
Sensitivity (d')	74	0.043	0.718
Response Bias (c)	74	−0.133	0.258

3.4.5 Assessment of participant and experimenter blinding

At the end of the experiment, we asked the participants and experimenters to report whether they thought they had received real or sham stimulation (see Table 3.5). We conducted a Chi-squared test on their responses for the two montages separately (M1 and Cerebellum). Even though the association between actual stimulation received and participant report was not significant ($\chi^2_{M1} = 2.500$, $p = 0.114$; $\chi^2_{Cerebellum} = 1.800$, $p = 0.180$), it is worth noting that most participants who received active stimulation were correct in reporting it (26/30 for anodal-M1 and 22/30 for anodal-cerebellum; see Table 3.5). We therefore combined data across montages and compared reports for the active and sham groups. This revealed a significant association between reported stimulation and actual stimulation re-

Table 3.4: Correlation between change of perception before/after speech adaptation and F1/F2 change in the last 30 trials of the adaptation block.

Perception change	**N**	***r***	***p***
F1 change			
Position	83	0.115	0.300
Slope	80	0.108	0.339
Hits	74	0.022	0.852
False Alarms	74	−0.008	0.949
Sensitivity (d')	74	−0.001	0.994
Response Bias (c)	72	0.024	0.842
F2 change			
Position	83	−0.051	0.650
Slope	80	−0.104	0.360
Hits	74	−0.019	0.874
False Alarms	74	−0.097	0.412
Sensitivity (d')	74	0.078	0.509
Response Bias (c)	72	0.132	0.270

ceived ($\chi^2 = 4.091$, $p = 0.043$). For experimenters, there was no significant association between actual stimulation given and their report when looking at the montages separately ($\chi^2_{M1} = 0.720$, $p = 0.396$; $\chi^2_{Cerebellum} = 1.800$, $p = 0.180$) or looking across montages by combining the active groups ($\chi^2 = 2.250$, $p = 0.134$). We conclude that allocation to receive active or sham stimulation was unsuccessfully concealed from participants. However, we think the impact of this awareness on our results was negligible given the lack of significant difference in any of our measures of speech production or perception among groups. Importantly, experimenters involved in the data analysis were unaware of whether the participant had received active or sham stimulation.

Table 3.5: Stimulation reported by participants and experimenters under M1 and cerebellum montages (Active vs. Sham).

	M1 montage		Cerebellum montage	
	Active (30)	Sham (15)	Active (30)	Sham (15)
Participant				
Active	26	10	22	8
Sham	4	5	8	7
Experimenter				
Active	18	7	12	3
Sham	12	8	18	12

3.4.6 Test of tDCS adverse effect

Participants completed an adverse effects questionnaire (Brunoni et al., 2011) on a 4-point scale (0-absent, 1-mild, 2-moderate, 3-severe). Nonparametric independent samples Kruskal-Wallis tests were conducted. Groups differed significantly in their reports of scalp pain ($p = 0.009$), tingling ($p < 0.001$), burning sensation ($p = 0.006$), and skin redness ($p = 0.004$). In short, participants who received anodal-M1 stimulation experienced higher levels of scalp pain, tingling, and burning sensation. Participants who received cerebellum tDCS experienced higher levels of skin redness most likely due to the visible redness over the right upper arm. It is worth noting that the mean ratings for most adverse effects were mild or moderate (see Table 3.6).

3.5 Discussion

The current study aimed to replicate previous findings that anodal tDCS over left motor cortex or right cerebellum enhanced speech adaptation (Lametti et al.,

Table 3.6: Post-tDCS report of adverse effects.

Number of participants self-reporting each adverse effect (0 = absent, 1 = mild, 2 = moderate, 3 = severe). N = 30 for each group.

M1 tDCS group	**Absent**	**Mild**	**Moderate**	**Severe**
Headache	18	7	4	1
Neck pain	29	0	1	0
Scalp pain	17	7	6	0
Tingling	1	9	16	4
Itching	8	9	9	4
Burning sensation	8	11	8	3
Skin redness	28	0	2	0
Sleepiness	10	10	7	3
Trouble concentrating	12	11	3	4
Acute mood change	26	1	3	0
Cerebellum tDCS group				
Headache	21	8	1	0
Neck pain	28	2	0	0
Scalp pain	23	6	1	0
Tingling	4	20	6	0
Itching	4	10	14	2
Burning sensation	19	8	3	0
Skin redness	22	5	3	0
Sleepiness	12	9	7	2
Trouble concentrating	8	16	6	0
Acute mood change	29	1	0	0
Sham group				
Headache	24	4	2	0
Neck pain	28	1	1	0
Scalp pain	27	2	1	0
Tingling	7	16	7	0
Itching	8	15	6	1
Burning sensation	15	9	5	1
Skin redness	30	0	0	0
Sleepiness	14	10	6	0
Trouble concentrating	17	10	3	0
Acute mood change	25	4	1	0

2018a). Specifically, in that study, anodal tDCS over both sites increased the magnitude of adaptation in the first formant in response to a shift in the frequency of that formant. In seeking to replicate these findings, we increased the sample size to three groups of 30, from three groups of 20 and made improvements to the experimental design. We anticipated that these modifications would further our

understanding of the effects of non-invasive brain stimulation in enhancing sensorimotor learning of speech and potentially inform future studies aimed at using brain stimulation to support clinical interventions. Even though we saw robust adaptation in each of the three groups of 30 participants, the magnitude of this adaptation was not increased by anodal tDCS over either the left motor cortex or the right cerebellum. Furthermore, exploratory analysis of our data using linear mixed effects modelling suggests that anodal tDCS over left motor cortex might even reduce adaptation. We failed, therefore, to replicate the findings of the previous study.

Below, we discuss potential explanations for this failure-to-replicate based on the changes made to the study. However, we believe we can rule each of these explanations out and instead argue that the original finding was based on unusual behaviour by participants in the sham group in that study, which led to the erroneous conclusion that anodal stimulation enhanced performance on this task.

3.5.1 Differences between previous study and current one

The current study involved several modifications to the methods used in the original study. For stimulation over the cerebellum, we placed the cathode on the upper right arm rather than the right cheek. The reason for this change addressed a shortcoming noted by the authors of the previous study. In attaching the electrode to the cheek, bandages were wrapped around the head and jaw, which could have affected speech production. Since the frequency of the formants relates to the resonances of the oral cavity, the position of the tongue within the oral cavity is critical. By not wrapping bandages around the jaw, participants were freely able to alter tongue position, and this ability did not differ among groups. We think it is

unlikely that the change to the montage reduced the efficacy of the right cerebellum anodal stimulation. Previous studies have shown this montage to be effective in enhancing procedural and visuomotor learning (Ferrucci et al., 2013; Weightman et al., 2023). Moreover, this change cannot explain why the anodal stimulation over left motor cortex failed to replicate, since that montage was unchanged and all other stimulation parameters were the same in the two studies.

We used a MATLAB program to record, alter and feedback speech to participants in our study rather than the analogue filters and acoustical effects processor used by Lametti et al. (2018a). Again, we think this change is unlikely to have affected results since all participants successfully adapted to a similar degree to that seen in previous work (see Appendix A.1). Furthermore, the "Audapter" programme used here has been used extensively by researchers in other studies, including our own and behavioural performance is not noticeably different to that seen when using the older system (Cai et al., 2023; Tang et al., 2021).

Additional behavioural measures were introduced in the current study that were not used previously. These included three blocks of noise-masked feedback that required participants to repeat the words "head", "bed", and "dead" five times each while feedback was masked by noise scaled to the volume of speech production. Data from the first familiarisation block were excluded. This was performed before the baseline block of the speech production task. We compared the change in formant frequencies during noise-masked production before and after the adaptation block of the experiment to provide a new measure of adaptation that was acquired under the same feedback conditions. The previous study measured adaptation by the amplitude of change from baseline when feedback was normal

to the end of the adaptation block (last 30 trials) when feedback was altered, and in the first 15 trials of the after-effect block when feedback was returned to normal. This last block is often referred to as "wash-out" or the "unlearning" block as participants use the restored normal feedback to return their production to the baseline. Since feedback is masked during the noise-masked blocks, the data are unaffected by these differences or by ongoing learning processes. Analysis of the data in the noise-blocks showed that all three groups adapted significantly by reducing the frequency of F1 (Figure 3.4). Anodal stimulation over either left motor cortex or right cerebellum had no effect on the magnitude of this adaptation, which is consistent with the results of the analyses of other measures. We conclude that our failure to replicate the findings of the previous study was unrelated to any of the modifications made to the experimental protocol. In the next section, we propose that the previous finding was a false positive driven by unusual behaviour in the sham stimulation group.

3.5.2 Explaining the failure to replicate Lametti, Smith, Freidin and Watkins (2018) study

In the current study, all groups adapted to a similar extent, averaging a change in F1 of 36% of the applied shift in F1 feedback by the end of the adaptation block. This amount of adaptation is consistent with reports from several previous studies (see Appendix A). In contrast, the magnitude of the adaptation in F1 seen in the Lametti et al. (2018a) study was smaller, averaging 29% for the motor cortex, 26% for the cerebellar, and only 10% for the sham stimulation group. The size of the adaptation in the sham group was unusually small and about a third

of that seen in the sham group in the current study. We suggest, therefore, that the apparent performance-enhancing effects of tDCS over motor cortex and cerebellum seen previously were due to comparison with the unusually low amount of adaptation seen in the sham group.

To investigate this idea further, we requested the data from the first author (DL) of the original study. We examined the number of participants in each study who showed an increase in F1 in response to the F1 up-shift in feedback rather than the expected decrease. In the original study, nine individuals (out of 60 = 15%) showed an unexpected increase in F1 by the end of the adaptation block; two (out of 20 = 10%) in the cerebellum group and seven (out of 20 = 35%) in the sham group. In the current study, seven individuals showed an increase in F1 (out of 90 = 8%) and these were evenly distributed across the three groups: three in the anodal-M1, two in the anodal-cerebellum, and two in the sham group. Reanalysis of the data in the original study removing the individuals who showed an increase in F1 reduced the sham group size to 13 and revealed no differences among the three groups in the amount of F1 adaptation. We suggest that the previously reported positive effects of tDCS on F1 adaptation were due to very low F1 adaptation in the sham stimulation group, caused by an unusually large proportion of people who did not adapt that were randomly assigned to that group. It is worth noting that the sample sizes of 20 participants per group in the original study were typical for speech adaptation studies and exceeded those in many of them (Deroche et al., 2017; Scott et al., 2020; Shum et al., 2011). These findings suggest larger sample size is needed to investigate the effect of individual differences.

3.5.3 The effects of anodal tDCS on F2 adaptation

In addition to the findings for adaptation in F1 described above, the Lametti et al. (2018a) study showed that cerebellar tDCS abolished the F2 adaptation, which was seen in the anodal-M1 and sham stimulation groups. In the current study, however, all groups showed moderate adaptation in F2 and we again failed to replicate the significant difference among groups previously described. This effect was not due to any obvious individual differences among groups in F2 adaptation. Reanalysis of the data excluding the participants described above had no effect of the results.

It is worth noting that changes in F2, the unperturbed formant, in response to a shift in F1 feedback are generally weaker and do not always occur (Purcell & Munhall, 2006; Rochet-Capellan & Ostry, 2011; Villacorta et al., 2007); cf. (MacDonald et al., 2011; Mitsuya et al., 2011).

3.5.4 The relationship between speech perception and speech adaptation

The speech perception measures were introduced to extend the expected findings of this replication. Unlike previous studies, which reported that individuals with better speech discrimination indices show more F1 adaptation (Villacorta et al., 2007), we found no relationship between such measures in our study. Adaptation to the feedback in the first formant during production of "head", "bed", and "dead" changed the position of the category boundary between "head" and "had" by increasing the size of the "head" category relative to baseline. This replicates the findings of previous reports (Deroche et al., 2017; Lametti et al., 2014a).

On discrimination tasks, post-adaptation response bias was changed also. Participants reduced their pre-adaptation 'liberal bias' consistent with the idea that their criterion threshold for detection of a difference increased. This meant a slight reduction in detection of differences (fewer hits and false alarms). We found no relationship between the amount of speech adaptation and changes in measures of speech perception, which was previously reported by others (Deroche et al., 2017; Lametti et al., 2014b).

3.5.5 Unsuccessful blinding

Before concluding, it is worth noting that the random assignment to receive active or sham stimulation was unsuccessfully concealed from the participants. This was most likely due to the use of 2-mA direct current since our experience at 1-mA is that participants are blind to stimulation (Chesters et al., 2018; Wiltshire & Watkins, 2020). In our opinion, the awareness by participants as to whether they were receiving active or sham stimulation was unlikely to have resulted in our failure to replicate group differences. All participants were naïve to the purpose of the experiment. Even when participants are made aware of the altered feedback, and instructed to avoid compensating for it, performance is unaffected (Munhall et al., 2009). Nevertheless, it is worth noting that extra measures may be necessary for future studies using 2-mA tDCS to ensure participants are blind to stimulation status.

3.5.6 Considerations for future experiments

Because a non-significant ANOVA cannot on its own support the null hypothesis of no effect, we supplemented the analysis of the primary F1 outcome with a Bayesian

ANOVA (JZS prior), which yielded a Bayes factor of $BF_{10} = 0.012$, providing strong evidence for the null model over a model including tDCS group. Given the small effect sizes observed ($\eta_p^2 = 0.003$–0.022) and our sample of 90 participants, a Bayesian analysis lets us quantify support for the absence of an effect rather than merely fail to reject a null hypothesis. The resulting Bayes factor indicates that our data are more consistent with the absence of a group effect. This strengthens our interpretation of the null result as evidence against, rather than merely a failure to detect, an effect of anodal tDCS on speech adaptation, consistent with the broader argument in this chapter for the no effect of tDCS on speech adaptation.

One possible explanation regarding why anodal tDCS failed to enhance performance on this speech production task relates to the possibility that the young healthy participants in our study might be performing at "ceiling", i.e. they have reached a saturation point or plateau in their learning. We think this is unlikely. Examination of Figure 3.2 A reveals a gradual trajectory of change in F1 for each group, which reflects the use of three different words during adaptation rather than one. Similar trajectories were observed in other adaptation studies using multiple words (Houde & Jordan, 1998; Houde & Jordan, 2002; Lametti et al., 2018b; Villacorta et al., 2007). We chose this task specifically because adaptation is gradual and on average only reaches about 36.2% of the perturbation applied to feedback. Even so, it is well known that individuals vary considerably in how much they adapt (see Figure 3.2 B) probably dependent on how much participants trade-off auditory and somatosensory feedback when compensating for the perturbation (Lametti et al., 2012). It would be interesting to consider how modulating sensory areas using non-invasive brain stimulation might alter this trade-off in future experiments. Even if

participants were at ceiling for extent of adaptation, anodal tDCS could enhance the learning rate or retention of the adapted state.

A related explanation for the failure of tDCS to alter performance, especially in young healthy brains, is the idea that when tDCS is applied over a fixed period of time, the brain changes its excitability to counteract the effects and achieve homeostasis (Amadi et al., 2015). A recent study used event-related tDCS, timed to be synchronous with movement, to enhance motor learning (Weightman et al., 2022). Synchronised cerebellum-tDCS enhanced motor performance, while M1-tDCS and asynchronous stimulation did not. It would be interesting to explore event-related tDCS over the cerebellum in our task.

We chose the primary motor cortex and the cerebellum as sites of stimulation because of their known roles in sensorimotor learning and because previous studies using tDCS over these regions had shown successful enhancement of motor learning (Reis et al., 2009) and visuomotor adaptation (Galea et al., 2011); cf. (Jalali et al., 2017; Panouillères et al., 2015). However, there are numerous examples of inconsistent effects of cerebellar tDCS, notably in visuomotor adaptation (Galea et al., 2011); cf. (Jalali et al., 2017; Panouillères et al., 2015) and on other language tasks (e.g. Bongaerts et al., 2022). The expected polarity dependent modulation of tDCS on excitability demonstrated in the motor cortex may not hold for the cerebellum (Oldrati & Schutter, 2017) and transcranial alternating current stimulation (tACS) may be a more promising approach to target the oscillation of cortico-cerebellar network of sensorimotor adaptation. One study showed cerebellar tACS in the gamma range increased visuomotor adaptation (Figueroa-Taiba et al., 2024).

Another point for discussion when considering our null results regards the use of relatively large electrodes (35 cm^2), which might lead to large current spread for both motor cortex tDCS (Bikson et al., 2018) and cerebellar tDCS (Klaus & Schutter, 2021). It is worth noting that Scott et al. (2020) used more focal high-definition anodal tDCS over left ventral sensorimotor cortex and increased speech adaptation during altered auditory feedback. It would be important to replicate this finding.

Finally, the negative finding of tDCS in speech adaptation might be related to the stimulation site. As discussed in Chapter 2, previous brain imaging studies have suggested that speech adaptation to formant perturbation engages more right-lateralised frontal and temporal activity (Floegel et al., 2020). This cortical activity is also believed to be correlated with the activity in the left cerebellum. In fact, in our later brain imaging study (Chapter 4), we observed brain activity corresponding to speech adaptation to formant perturbation in sensorimotor areas of both hemispheres and in the left cerebellum. Future studies may stimulate the right speech motor cortex and left cerebellum to test the effect of anodal tDCS on speech adaptation to formant perturbation.

3.6 Summary

In the current study, we aimed to replicate the enhancement of tDCS over left speech motor cortex and right cerebellum on speech adaptation. By recruiting a larger sample size (30 participants for each group), we did not replicate the enhancement of tDCS on speech adaptation described by Lametti et al. (2018a). We conclude that anodal tDCS over either M1 or the cerebellum has no effect on

speech adaptation. This failure to replicate adds to the body of literature demonstrating inconsistent effects of tDCS on cognitive performance in young, healthy participants. In the language and motor learning domains, specifically, there is a growing body of literature highlighting the unreliable nature of previously described effects (Klaus & Hartwigsen, 2020; Westwood et al., 2017; Wiltshire & Watkins, 2020). Effect sizes are likely to be small and therefore even larger sample sizes are needed to provide sufficient power to detect these effects. We suggest that further studies could exploit the benefits of better temporal and spatial resolution afforded by event-related tDCS and high-definition tDCS, or use tACS to modulate activity in a frequency-specific manner. Another possibility is the stimulated areas (left speech motor cortex and right cerebellum) might not be the areas involved in speech adaptation. In the study presented in the next chapter, we used behavioural and brain imaging methods to investigate the neural correlates of speech adaptation during altered auditory feedback.

4

The neural correlates of speech adaptation

4.1 Abstract

Here, we developed a Mandarin version of the speech adaptation task previously tested in English. Seventy-six native Mandarin speakers were asked to produce Mandarin consonant-vowel syllables while receiving real-time auditory feedback in which the frequency of the first vowel formant (F1) was either increased or decreased. We compared vowel production during altered and normal feedback. Results showed that participants adapted in the direction opposite to the shift, as seen in English speakers (see Chapters 2 and 3).

Functional MRI was then used to investigate the neural correlates of sensorimotor adaptation of speech in Mandarin. A further 48 native Mandarin speakers

were asked to produce Mandarin consonant-vowel syllables with altered auditory feedback where the F1 was increased from baseline where feedback was normal. Images of the whole brain were collected to capture activity during performance of this task. As participants adapted to the altered auditory feedback by reducing the frequency of F1 production, activity increased in the left posterior cerebellum and decreased in sensorimotor cortex, consistent with the notion that these areas play a critical role in forward modelling and learning of refined feedforward motor commands, respectively.

4.2 Introduction

There is growing interest in the adaptive behaviour of human cognition, particularly in how individuals adjust to changes in their environment. This type of adaptation often involves motor and sensory networks in the brain. Speech adaptation has been studied by perturbing the sound of a voice when speech is produced (Houde & Jordan, 1998; Houde & Jordan, 2002), see Chapter 2. In the task, speech is recorded and fed back to the speaker in near real-time. Speech is perturbed by altering the frequencies of the vowel formants and measuring the effects of such perturbation on subsequent speech production. The frequencies of the first (F1) and second vowel formants (F2) determine vowel identity (Ladefoged & Johnson, 2014). Participants generally respond by adjusting their speech in response to this formant perturbation. Such adaptation can occur implicitly and involuntarily (Munhall et al., 2009).

This form of feedback control suggests that the brain continuously compares the planned speech outcome with auditory feedback. If there is a mismatch between the intended auditory target and the actual perceived speech output, the brain initiates immediate corrective adjustments to minimise errors. Adapted speech is often retained after a period of learning with sustained altered auditory feedback (Munhall et al., 2009; Villacorta et al., 2007) possibly reflecting changes in an internal model of speech production. Speech adaptation involves both forward and inverse modelling of speech to provide motor-to-sensory and sensory-to-motor transformations, respectively. The forward model predicts the sensory consequences of an intended motor command and uses sensory feedback to update ongoing movements. The inverse model learns the mapping between sensory outcomes and motor commands. Inverse modelling provides feedforward control allowing accurate execution of articulatory commands at speed. Sensory feedback is used to update internal models and maintain accurate speech production.

So far, there is little empirical work using brain imaging to investigate the neural correlates involved in speech adaptation. As introduced in Chapter 1, the DIVA (Directions Into Velocities of Articulators) model (Tourville et al., 2008) proposed a left lateralised brain network to support feedforward control, and a right lateralised brain network to support feedback control in speech production. Brain imaging studies using unpredictable altered auditory feedback tasks have shown increased activity in posterior auditory cortex bilaterally, right ventral motor and premotor cortex and anterior medial cerebellum (Guenther et al., 2006; Tourville & Guenther, 2011). An error signal network including right angular gyrus, right supplementary motor area, and cerebellum, and a sensorimotor integration network within inferior

frontal gyrus bilaterally were also identified to support speech motor control (Zheng et al., 2013). However, adaptation to unpredictable altered auditory feedback does not involve the feedforward control system as feedback is only altered within each trial and returns to normal after each perturbation.

Findings from EEG studies suggest that adaptation involves systematic reorganisation of functional coupling across distributed networks. Phase-coherence analyses show redistribution of theta–gamma coupling during learning, with increased coherence over central regions and decreased coherence over fronto-temporal regions; critically, late-phase adaptation was positively related to central coherence and negatively related to frontal coherence, consistent with the establishment of a new sensorimotor mapping (Sengupta & Nasir, 2015). Related work shows that late-phase adaptation is reflected in power changes across theta, beta, and gamma bands, and that these bands interact rather than contributing independently (Sengupta & Nasir, 2016). ERP evidence similarly links adaptation to auditory prediction processes: behavioural adaptation to an F1 increase covaries with speaking-induced suppression of the P2 component under altered feedback (Sato & Shiller, 2018).

Recently, MEG has provided converging evidence that prediction-error signals within auditory cortex track speech motor learning (Kim et al., 2025). Speaking-induced suppression is reduced (corresponding to larger auditory prediction errors) during early adaptation relative to unaltered feedback and the magnitude of this reduction predicted behavioural adaptation. The findings from this study highlighted the putative role of auditory cortex in prediction-error computations during speech adaptation.

A handful of studies have explored the roles of specific brain areas in speech adaptation using noninvasive brain stimulation. Transcranial magnetic stimulation (TMS) revealed the left inferior parietal lobe (Deroche et al., 2017) and supramarginal gyrus (Shum et al., 2011) as key players in speech adaptation during sustained altered auditory feedback. Temporarily inhibiting the tongue representation in the left speech motor cortex using TMS abolished speech adaptation, while inhibiting the hand representation had no effect (Tang et al., 2021). Similarly, transcranial direct current stimulation (tDCS) studies suggest that increasing excitability in sensorimotor networks can enhance adaptation. Anodal tDCS over the left inferior parietal lobe increased adaptation to the F1 perturbation compared to cathodal and sham stimulation (Deroche et al., 2017). High-definition anodal tDCS over the left ventral sensorimotor cortex also increased adaptation in F1 during perturbed auditory feedback (Scott et al., 2020). Anodal tDCS over the right cerebellum led to larger pitch compensation for pitch errors in vowel production (Peng et al., 2021).

Functional brain imaging studies have begun to characterise the neural mechanisms that support speech adaptation under sustained altered auditory feedback. FMRI evidence indicates partly separable responses to temporal (delayed feedback) versus spectral (formant-shift) perturbations, broadly consistent with hemispheric specialisation in early auditory processing (right-biased spectral; left-biased temporal) (Floegel et al., 2020). Spectral adaptation was associated with increased activity in inferior frontal gyrus and superior temporal sulcus, predominantly in the right hemisphere, whereas temporal adaptation was associated with increased activity in superior temporal gyrus and superior temporal sulcus, predominantly in the

left; inferior frontal gyrus activity showed a corresponding rightward bias for spectral perturbations and leftward bias for temporal perturbations. Building on the proposal of bilateral auditory feedback comparison interacting with left-lateralised frontal feedforward control in the DIVA model, these fMRI results motivated a domain-tuned refinement in which mismatch processing in superior temporal cortex is right-biased for spectral errors and left-biased for temporal errors, with mismatch signals conveyed via temporo-frontal pathways to generate corrective commands (Floegel et al., 2020). These findings challenge a simple right-lateralised feedback versus left-lateralised feedforward division (Tourville et al., 2008).

To date, the vast majority of studies exploring speech adaptation have been conducted in English. To the best of our knowledge, the effects of formant perturbation have not yet been explored in tonal languages like Mandarin. In tonal languages, small deviations in fundamental frequency can alter lexical tone categories. Accordingly, the effects of pitch perturbation on speech production have been explored in Mandarin speakers. Robust compensatory responses to upward and downward pitch shifts were seen, along with after-effects when normal feedback was restored, indicating both online feedback correction and short-term sensorimotor adaptation (Jones & Munhall, 2002; Xu et al., 2004).

In the current study, we had two main aims: (i) to develop a speech adaptation paradigm using formant perturbation suitable for Mandarin speakers; (ii) to examine the neural correlates of this form of speech adaptation in Mandarin. We predicted that Mandarin speakers would show compensatory changes in vowel production in response to sustained formant perturbation, shifting their productions in the direction opposite to the perturbation (as seen previously in other

non-tonal languages). We further predicted that speech adaptation would change brain activity in auditory, somatosensory, and motor areas, and in the cerebellum.

4.3 Methods – Behavioural study

4.3.1 Participants

76 native Chinese speakers (38 females) aged from 18 to 27 years participated in the study. Most were students at Beijing Normal University and were given a small monetary reimbursement for participation. All participants reported right-handedness, normal or corrected-to-normal vision, and no speech or hearing difficulty. The study was approved by the Human Participants Research Ethics Committee of State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University (ICBIR_D_0084_001) in April 2024. All participants provided informed consent before taking part.

Half of the participants (38 participants) received auditory feedback where F1 frequency was increased (F1 upward shift), and the other half (38 participants) received auditory feedback where F1 was decreased (F1 downward shift). We excluded two participants from the analysis due to incorrect or unusual pronunciation of the syllables. We retained data for analysis in 37 participants (19 males, 18 females) in each group.

4.3.2 Experimental procedure

Nine Mandarin consonant-vowel syllables (Pinyin) were chosen, i.e. “dī”, “mī”, “tī”, “dē”, “mē”, “tē”, “dā”, “mā”, “tā”. These syllables are commonly used to articulate

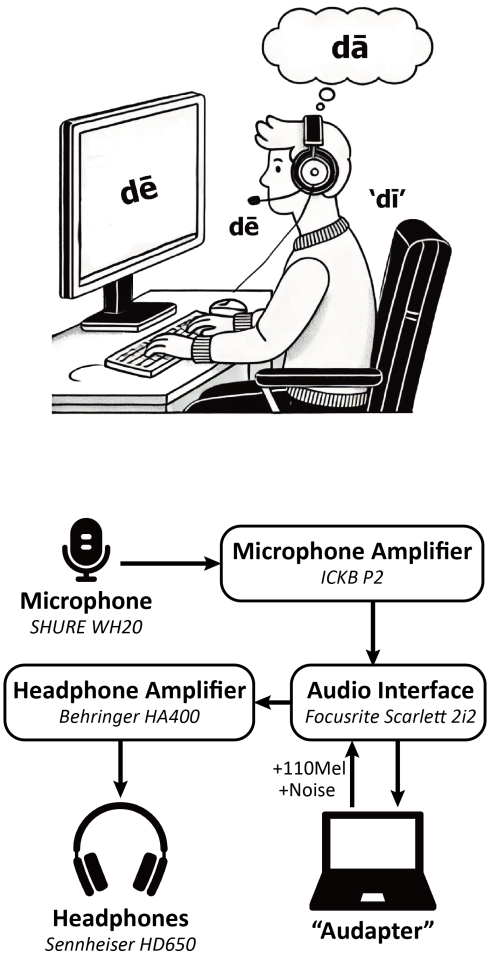

Figure 4.1: Setup of behavioural study
The upper figure shows the setup in the lab. It was generated with the help of ChatGPT. The bottom figure shows how the signal is recorded, processed, and fed back to the speaker.

Chinese characters. All syllables were presented and pronounced in the first tone (indicated by the macron on the top of the vowel). Participants were seated in a quiet room with a screen displaying the syllable in front of them. They wore headphones (HD650, Sennheiser, Germany) and a microphone (WH20, SHURE, USA), which were connected to a headphone amplifier (HA400, Behringer, Music Tribe, China) and a microphone amplifier (P2, ICKB, China), respectively. Both amplifiers were connected to a Focusrite Scarlett 2i2 audio interface (Focusrite

Audio Engineering Ltd., High Wycombe, UK). Speech was recorded and fed back to them using "Audapter" (Cai et al., 2008; Tourville et al., 2013). See Figure 4.1.

The speech production task included four blocks (vowel exploration, baseline, adaptation, and after-effect). In each trial, participants read words presented for 1500-ms at 750-ms intervals. During the vowel exploration block, participants produced the nine syllables with unperturbed feedback. Syllables were presented in a random order, 15 times each. During baseline and after-effect blocks, three syllables ("dē", "mē", "tē") were each repeated 20 times with normal feedback. During the adaptation block the three syllables were repeated 75 times with altered feedback. Breaks of 30 seconds occurred every 45 utterances.

Auditory feedback was altered using "Audapter" (Cai et al., 2008; Tourville et al., 2013). In response to the altered auditory feedback, participants typically adjust their utterances to adapt to the perceived altered auditory feedback. For example, participants say "mē" but with the increase in F1, the word fed back over headphones sounds closer to the frequency of the vowel in "mō". Participants may adapt differently: by reducing F1 frequency in their production, which when shifted sounds like the target vowel, or by reducing F1 and simultaneously increasing F2 frequencies to produce a familiar tongue position for the vowel in "dī". Feedback was perturbed by increasing or decreasing F1 frequency by 110 Mel, mixing it with the rest of the unaltered speech signal and playing it back through headphones with an imperceptible delay. Pink noise at 60 dB was added to both altered and normal feedback.

4.3.3 Statistical analysis

For all utterances, formant frequencies were extracted and analysed using MATLAB as described in Chapter 3. Specifically, the average F1 and F2 frequencies were extracted across a 50-ms time frame at the centre of each vowel using Praat (Boersma & Weenink, 2022) with linear predictive coding. F1 and F2 values exceeding three standard deviations from the mean in each block were excluded (F1: 0.883% and F2: 2.66% of the data). Formant frequency changes (F1 and F2) were calculated for each trial in the baseline, adaptation, and after-effect blocks by subtracting the average value of the baseline trials for each formant in each participant.

The speech adaptation effect was tested by analysis of F1 and F2 changes from baseline in the last 30 trials of adaptation block and in the first 15 trials of after-effect block (consistent with previous analyses of similar data in English speakers – see Chapter 3). One sample t-tests compared the formant frequency changes against zero. We also examined the adaptation effect separately for each consonant (d, m, & t). First, we extracted trials for each consonant separately and formant frequency changes were calculated by subtracting the average value of the baseline trials for each consonant-vowel syllable for each formant in each participant. Adaptation magnitude was then measured by averaging all trials in adaptation block (75 trials for each consonant-vowel syllable). Adaptation magnitudes were then compared using repeated-measures ANOVA for F1 upward and downward groups separately.

4.4 Results – Behavioural study

4.4.1 Behavioural results

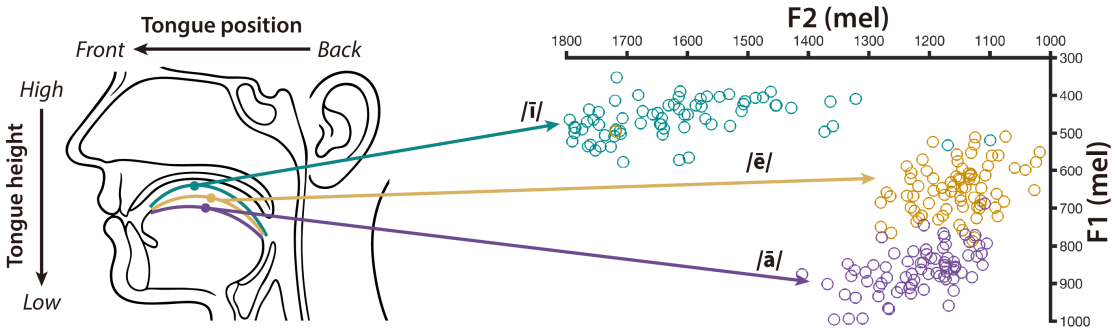

Figure 4.2: Vowel exploration - Behavioural study
Schematic showing the relationship between tongue height and position in the vocal tract (left) and the corresponding frequencies of the first and second formants of three vowels (right). Right figure shows vowel production averaged for each participant. Averaged F1 and F2 frequencies of each trial utterance are plotted in yellow for the vowel /ē/ in "dē", "tē", and "mē", green for the vowel /ī/ in "dī", "tī", "mī" and purple for the vowel /ā/ in "dā", "tā", and "mā".

First, we compared the frequencies associated with the production of vowels /ī/, /ē/, and /ā/ in Mandarin speakers. The three vowels were significantly different from each other in both F1 and F2 production (ps < 0.001). Examination of F1-F2 vowel space in Figure 4.2 indicates that the vowel /ī/ has much higher F2 frequencies than vowels /ē/ and /ā/, which have similar values for F2.

For the behavioural data, changes in F1 and F2 production from baseline during all phases of the experiment are shown in Figure 4.3. For the group receiving an upward shift in F1 feedback, there was a significant reduction in F1 in the last 30 trials of adaptation relative to baseline ($t(37) = -7.10$, $p < 0.001$, Cohen's $d = -1.15$), and this was maintained in the first 15 trials of the after-effect block ($t(37) = -7.71$, $p < 0.001$, Cohen's $d = -1.25$). For the group receiving a downward shift in F1 feedback, there was a significant increase in F1 in the last 30 trials of

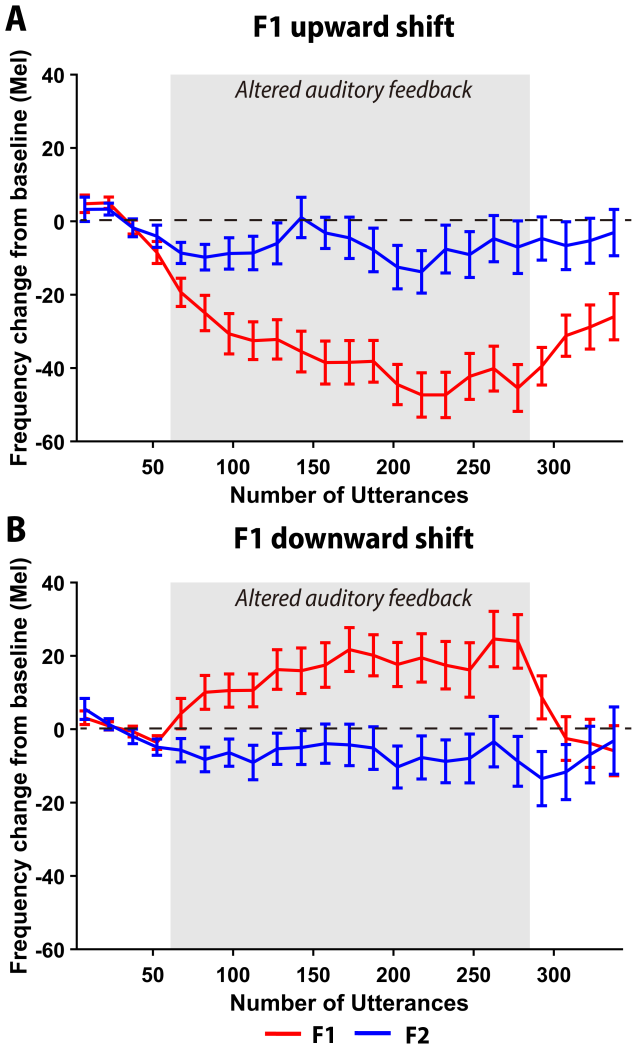

Figure 4.3: F1 and F2 changes in response to F1 upward and downward shift **A.** Frequency change from baseline in response to F1 upward shift. **B.** Frequency change from baseline in response to F1 downward shift. Frequency change from baseline was averaged for every block of 15 trials. Adaptation phase with altered auditory feedback is marked in grey. The solid lines indicated the group mean changes in frequency and error bars indicate ± one standard error of the mean. Red – F1; Blue – F2.

adaptation ($t(37) = 3.30$, $p = 0.001$, Cohen's $d = 0.535$) but this weaker effect was not maintained in the first 15 trials of the after-effect block ($t(37) = 1.48$, $p = 0.074$, Cohen's $d = 0.240$). Participants did not show significant change in F2 in response to either upward or downward shifts in F1 ($ps > 0.05$).

As is clear from Figure 4.3, adaptation in response to the upward shift in F1

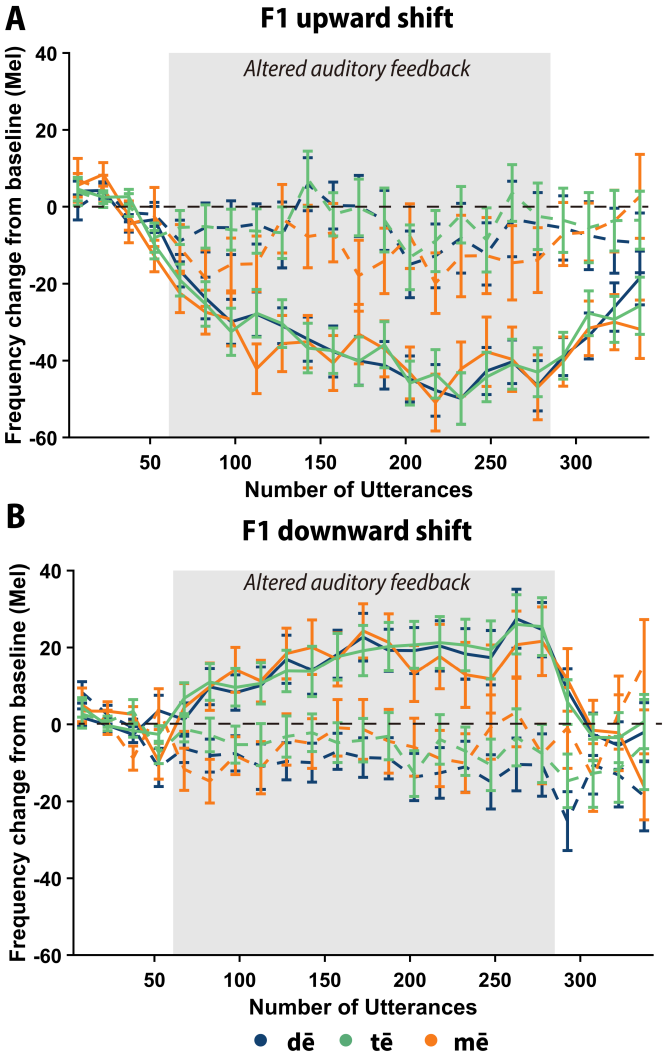

Figure 4.4: F1 change from baseline for each consonant-vowel syllable
Formant frequency change for each consonant-vowel syllable in response to F1 upward shift (**A**) and F1 downward (**B**) shift. Solid lines indicate Frequency change in F1 from baseline and dashed lines indicate frequency change in F2 from baseline. See legend to Figure 4.3 for details. Blue – dē; Green – tē; Orange – mē.

was larger in magnitude compared with the response to the downward shift ($t(74) = 7.05$, $p < 0.001$, Cohen's $d = 1.62$) and this change was maintained for longer once feedback had been returned to normal in the after-effect block ($t(74) = 6.18$, $p < 0.001$, Cohen's $d = 1.42$).

We also tested the size of the F1 and F2 adaptation effect for each consonant-vowel syllable separately. Results showed that three consonant-vowel syllables

showed similar amplitudes and trajectories of speech adaptation in response to both upward and downward shifts (ps > 0.05, see Figure 4.4).

Based on the results of the behavioural study, the syllable "dē" was selected as the speech stimulus, and an upward formant shift was used as the perturbation in the subsequent functional MRI study. Anecdotal observations during behavioural testing revealed that candidate syllables differed in their suitability for repeated production during adaptation. The syllable "dī" showed instability in some speakers, who sometimes produced it closer to "dā", making it less appropriate as a stable vowel target. The syllable "tē" was also not ideal because its first-tone form does not correspond to a common Mandarin word. By contrast, first-tone "dē" is associated with a highly frequent Mandarin word and therefore provides a familiar and stable target for repeated production. In addition, the consonant-vowel (CV) syllable was used rather than the consonant–vowel–consonant (CVC) structure often used in English versions of this paradigm (see Chapter 3), because CVC syllables are not part of Mandarin syllable structure. Finally, as the behavioural experiment showed more robust speech adaptation to the upward than the downward formant shift, the upward shift was chosen for the fMRI experiment.

4.5 Methods – Brain imaging study

4.5.1 Participants

Forty-eight (23 males and 25 females) Mandarin Chinese speakers were recruited aged between 18 and 33 years. Most were students at Beijing Normal University and were given a small monetary reimbursement or class credit for participation.

Exclusion criteria included a history of neurological disorders, speech or hearing impairment, and MRI contraindications. All participants passed a hearing screening test with no hearing loss at or above 1000 Hz and 20 dB. All were naive to the task (i.e. they had not previously participated in the behavioural study).

4.5.2 Experimental procedure

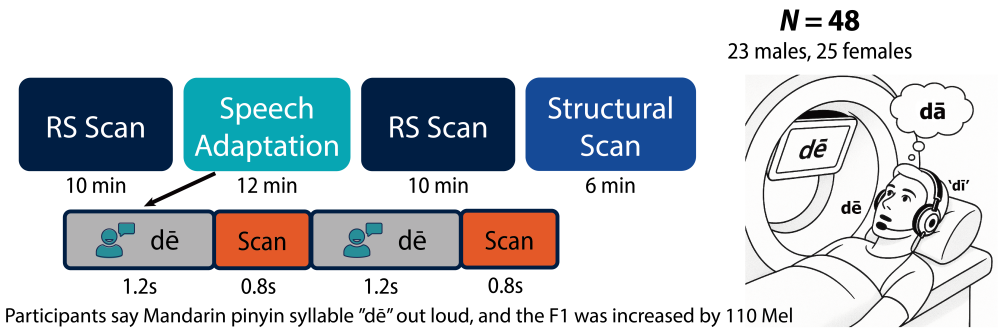

Figure 4.5: Setup and procedure of the brain imaging study
Illustration of procedure (left) and setup (right) of the brain imaging study. The setup figure was generated with the help of ChatGPT.

MRI scans were collected using the Siemens 3T MAGNETOM Prisma scanner and 32-channel head coil at the Brain Imaging Centre of Beijing Normal University.

For each participant, we ensured that they could hear the sound of their own production of the syllable "dē" played back to them clearly through the headphones and were comfortable with the sound level. Minimum sound levels (from the Optoacoustics control panel) were not below 90 dB. Participants practised production of "dē", "dī" & "dā" three times each to familiarise themselves with the task, speaking into the microphone and hearing themselves via the headphones comfortably inside the scanner.

Fieldmap images were collected first to capture the inhomogeneity of the magnetic field. Next, participants were required to complete the vowel exploration task for five and a half minutes, during which 160 echo-planar images were collected. Participants produced the syllable presented on the screen during a 1.2 s interval between echo-planar volume acquisitions. Subsequently, they were required to fixate on a cross at the centre of the screen for ten minutes while resting-state fMRI images (750 scans) were acquired. Participants next completed the altered auditory feedback task for 12 minutes (360 scans). They produced the syllables displayed on the screen while auditory feedback was either normal (baseline: 40 trials) or altered (adaptation: 200 trials) during the 1.2 s silent interval between image acquisitions. Then, another resting-state scan session was conducted for ten minutes. Finally, T1-weighted structural images were collected (6 minutes). See Figure 4.5 for illustration.

4.5.3 Scan parameters

Fieldmap images were acquired using the following sequence: TR = 589 ms, TE_1 = 4.92 ms, TE_2 = 7.38 ms, flip angle = 60°, FoV = 216 mm, voxel size = 2.4 mm isotropic.

FMRI images (360 volumes) were acquired using the multi-band accelerated EPI sequence: TR = 800 ms, TE = 30 ms, flip angle = 80°, FoV = 216 mm, multiband acceleration = 6, phase encoding direction = anterior to posterior, voxel size = 2.4 mm isotropic. A single-band image was collected before each multi-band sequence which was used to correct the multi-band induced distortions to the images for registration with the structural image and standard space template.

T1-weighted structural images were collected using a MPRAGE sequence in sagittal orientation: TR = 2530 ms, TE = 2.27 ms, flip angle = 7°, FoV = 256 mm, voxel size = 1 mm isotropic.

4.5.4 Vowel exploration, baseline, and altered auditory feedback task

Experiments ran in MATLAB, using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). During the vowel exploration task, participants produced Mandarin syllables "dē", "dī", and "dā" eight times during each block (16 seconds), interspersed with a 16-seconds rest block featuring a fixation cross at the centre of the screen (equivalent to eight trials of production). The use of one consonant "d" resulted in a consistent place of articulation ("d" – alveolar) across the in-scanner task. These four blocks (three syllable production blocks and a rest block) were repeated four times. So, each syllable was produced 32 times in four blocks of eight trials.

During the speech adaptation task, participants produced the syllable "dē" while receiving either normal (baseline phase) or altered (adaptation phase) auditory feedback. Blocks of 20 trials (40 s) were followed by rest blocks for 20 seconds during which participants fixated a cross at the centre of the screen. The baseline phase comprised two blocks of 20 trials (40 trials in total) where auditory feedback was normal and each was followed by a rest block. During the adaptation phase, participants completed 10 blocks of 20 trials, each followed by a rest block. Feedback was perturbed by increasing F1 frequency by 110 Mel, mixing it with the rest of the

unaltered speech signal and playing it back through headphones with an imperceptible delay. Pink noise at 60 dB was added to both altered and normal feedback to mask the abrupt noise induced by "Audapter" at the start and the end of each trial.

Participants used the MR-compatible OptoActive II headphones and FORMRI III+ microphone (Optoacoustics Ltd., Or Yehuda, Israel) inside the scanner. We sent the recording of each syllable collected from the microphone to a Focusrite Scarlett 2i2 audio interface (Focusrite Audio Engineering Ltd., High Wycombe, UK) and fed back the recording to the headphones with or without the perturbation and mixed with noise. Participants did not wear earplugs so that the intensity and frequency of the sound perceived by them was maintained. This arrangement was pre-approved by the ethics committee and agreed with each participant before scanning. Another monitoring set of Sennheiser HD 650 headphones (Sennheiser, Germany) was connected to the sound output of the audio interface for the experimenter to monitor the effectiveness of auditory feedback throughout the experiment.

4.5.5 Behavioural analysis

To analyse participants' behavioural performance when inside the scanner, the fundamental (F0), first and second formant frequencies from each recording were extracted using Praat (Boersma & Weenink, 2022) with the linear predictive coding (LPC) method. Further data analysis was performed in MATLAB. The average fundamental and formant frequencies were calculated from a 50-ms time window at the centre of each vowel. F0, F1, and F2 values exceeding three standard deviations from the mean were excluded from further analysis (the percentage of outliers excluded across all participants in F0 was 0.369%, F1 was 0.004%, and

F2 was 0%). Fundamental and formant frequency changes across all trials were calculated by subtracting the average value of the baseline for each formant in each participant. We compared the averaged F0, F1, and F2 in the last 40 trials of adaptation phase against the average for the baseline (40 trials).

4.5.6 Task fMRI analysis

Task fMRI analysis was conducted using FEAT (FMRI Expert Analysis Tool) Version 6.00. Image pre-processing of the task functional images involved: removal of nonbrain tissues using BET (Smith, 2002), motion correction using MCFLIRT (Jenkinson et al., 2002), high-pass temporal filtering (90-s), and fieldmap unwarping using FUGUE to correct for distortions in the images and improve registration (Jenkinson, 2002); linear registration to the participant's structural T1-weighted image using FLIRT's Boundary-Based Registration (Jenkinson & Smith, 2001; Jenkinson et al., 2002) and non-linear registration via the high resolution structural image to the standard 2-mm MNI-152 template image using FNIRT (Andersson et al., 2007); and spatially smoothed using an FWHM = 5 mm Gaussian kernel. Single-subject general linear model (GLM) analyses were carried out with local autocorrelation correction (Woolrich et al., 2001). Standard motion correction parameters (6) plus their temporal derivatives (6), and squares of the above (12) were included as regressors in the GLM model (standard + extended motion parameters). The extended motion parameters option was selected to better regress out potential movement caused by speaking inside the scanner.

For the vowel exploration task, separate explanatory variables (EVs) were set to production of each syllable "dī", "dē", and "dā". We then modelled brain activity

evoked by the production of each syllable separately and obtained a map of the F-statistic across the effect size maps for the three syllables.

For the altered auditory feedback task, we modelled six phases of the task; each phase comprised 40 trials. This allowed us to capture the temporal change in brain activity under the altered auditory feedback relative to the baseline phase. We also included F1 and F2 change from baseline of each trial for each subject as two additional EVs. This allowed us to 1) model brain activity by comparing adaptation (altered feedback) with baseline (normal feedback) directly; 2) examine for correlates of F1 and F2 change reflecting individual levels of speech adaptation.

Data for both tasks were averaged at the group level. Significance was inferred using cluster statistics with an initial height threshold of $Z > 3.1$ and a corrected cluster extent threshold of $p < 0.05$ (Jezzard et al., 2001).

4.6 Results – Brain imaging study

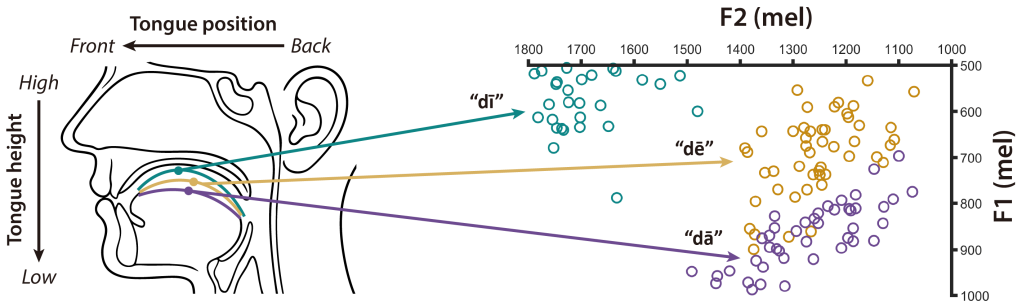

Figure 4.6: Vowel exploration - Brain imaging study
Schematic showing the relationship between tongue height and position in the vocal tract (left) and the corresponding frequencies of the first and second formants of three vowels (right). Right figure shows vowel production averaged for each participant. Averaged F1 and F2 frequencies of each trial utterance are plotted in yellow for the vowel in "dē", green for the vowel in "dī" and purple for the vowel in "dā".

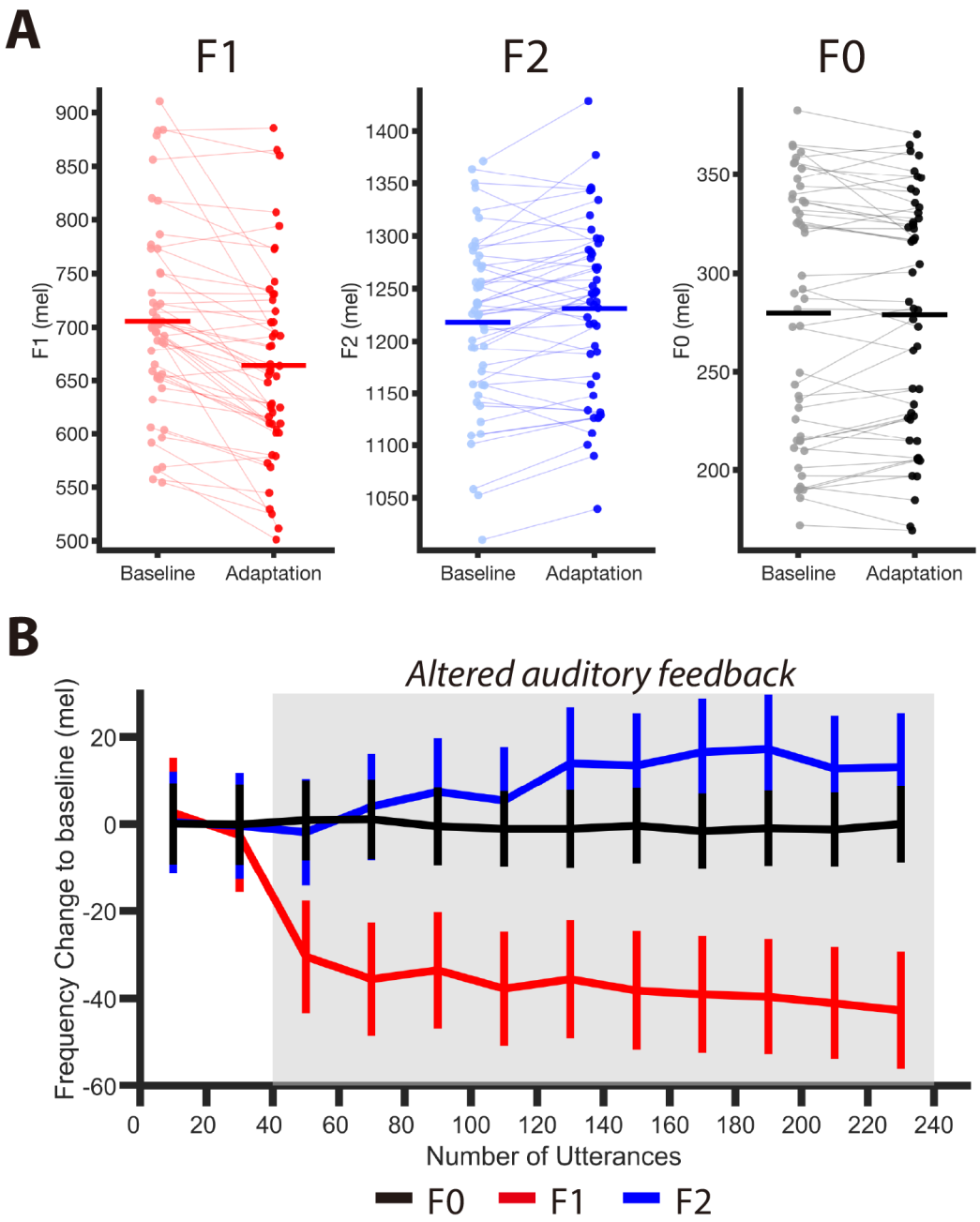

Figure 4.7: Change of formant and fundamental frequencies during altered auditory feedback

A. Averaged F1 and F2 during baseline (all 40 trials) and last 40 trials of the adaptation phase, horizontal lines indicate the group mean. **B.** Frequency change from baseline averaged every block of 20 trials for baseline and adaptation phases. Adaptation phase with altered auditory feedback is marked in grey. Error bars show ± standard error. Black – F0; Red – F1; Blue – F2.

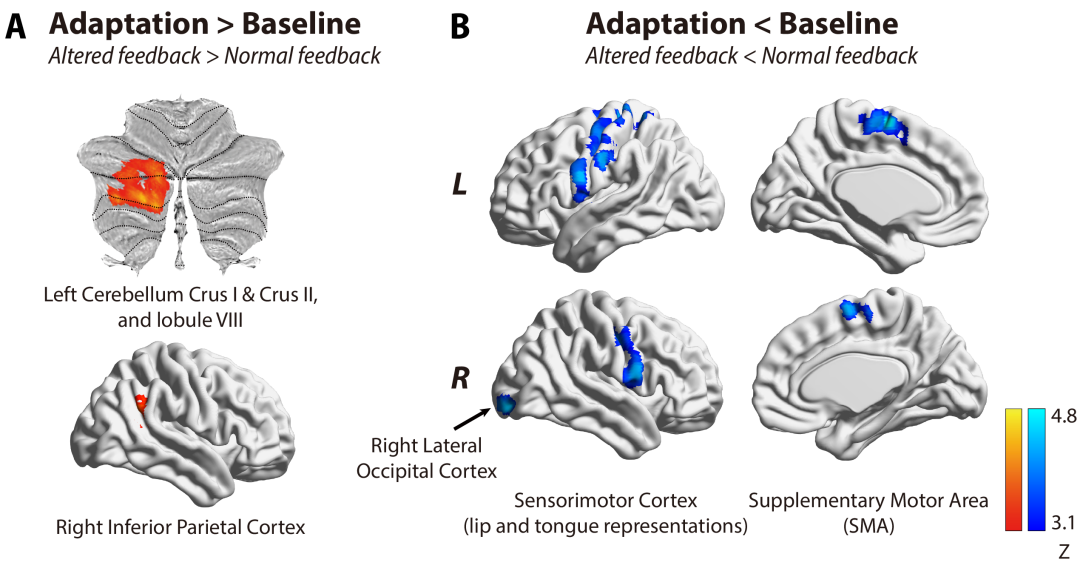

Figure 4.8: Changes in brain activity during altered auditory feedback **A.** Increased brain activity (red-yellow) during altered feedback (adaptation) compared with normal feedback (baseline) is shown in a map of the Z-statistic (thresholded at a height of 3.1 with cluster extent $p < 0.05$ corrected) overlaid on a cortical surface of the right hemisphere from the MNI-152 template and a flat map of the cerebellum (SUIT surface-based map (Diedrichsen & Zotow, 2015)). **B.** Reduced brain activity (blue–light blue) during altered feedback (adaptation) compared with normal feedback (baseline) is shown in a map of the Z-statistic overlaid on lateral and medial cortical surfaces. Surface rendering was created using a "maximum voxel" approach from the BrainNet Viewer toolbox (Xia et al., 2013).

4.6.1 Vowel exploration: production of vowel-consonant syllables

First, we tested the production of vowels in syllables "dī", "dē", and "dā". The three vowels were significantly different from each other in F1 production (ps $<$ 0.001). The vowel in "dī" had significantly higher F2 frequency than the other two vowels (ps < 0.001), which did not differ ($p = 0.442$). Averaged F1-F2 vowel space is shown in Figure 4.6.

Contrasts comparing brain activity evoked by production of the different consonant-vowel syllable pairs did not reveal any significant differences. An expected network of brain areas was activated by syllable production including ventral premotor,

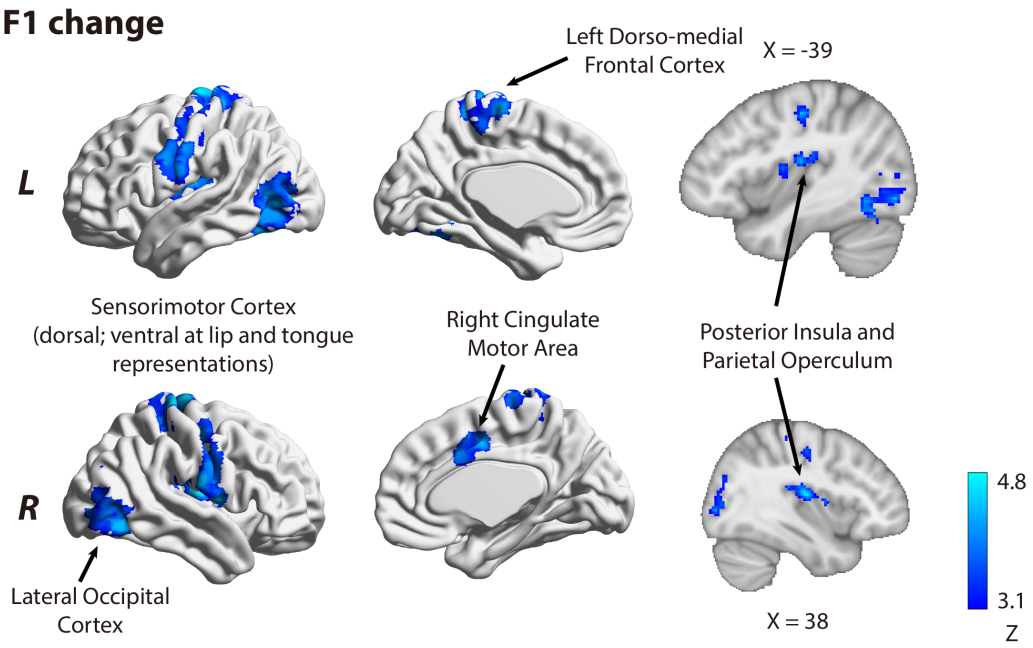

Figure 4.9: Brain activity correlating with the magnitude of F1 reduction Negative linear relationship between change in brain activity and change from baseline in F1. See caption to Figure 4.8 for details of surface rendering. Sagittal images were created using FSLeyes.

sensorimotor, and superior temporal cortex bilaterally.

4.6.2 Change of formant and fundamental frequencies during altered auditory feedback

On average, participants adapted to the shift in F1 by reducing the frequency of F1 and increasing the frequency of F2 in their production during the adaptation phase. Reduction in frequency during the last 40 trials of adaptation compared with baseline was significant for F1 ($t(47) = -5.91$, $p < 0.001$, Cohen's $d = -0.853$) and was significantly increased for F2 ($t(47) = 3.03$, $p = 0.004$, Cohen's $d = 0.439$). Although both changes were reliably seen across participants the magnitude of the change in F1 was considerably bigger than that in F2, as seen in the beha-

vioural study. There was no change in F0 ($t(47) = -0.394$, $p = 0.696$, Cohen's $d = -0.057$). See Figure 4.7.

4.6.3 Changes of brain activity during altered auditory feedback

During altered feedback (adaptation), activity increased relative to normal feedback (baseline) in Crus I, Crus II, and lobule VII of the left posterior cerebellum and right inferior parietal cortex (Figure 4.8 A), and decreased in the ventral sensorimotor cortex, at the level of the lip and tongue representations, in the supplementary motor area (SMA) bilaterally, and in the right lateral occipital cortex (Figure 4.8 B).

The magnitude of F1 reduction in individual participants during the adaptation phase was negatively correlated with change in activity in a similar network of brain regions. However, this more sensitive analysis additionally revealed adaptation-related changes in the sensorimotor cortex, both dorsal and ventral (at the level of the lip and tongue representations), posterior insula and parietal operculum, all bilaterally, and right cingulate motor area (Figure 4.9). Activity was not significantly correlated with change in F2 in any brain area.

4.7 Discussion

Here, we examined speech adaptation in response to a shift in formant frequency in auditory feedback in a large sample of Mandarin Chinese speakers. Our behavioural study revealed a similar pattern of behaviour to that previously and extensively observed in Western non-tonal languages (e.g. English). We then examined the

neural correlates of adaptation using functional MRI. Our results revealed that as participants adapted to the altered auditory feedback by reducing the frequency of the first formant in their production, activity increased in the left posterior cerebellum and decreased in the sensorimotor cortex. Individual differences in the magnitude of F1 reduction were also negatively correlated with adaptation-related activity in a similar network of regions.

In the behavioural study, Mandarin Chinese speakers exhibited adaptive responses to F1 formant perturbation, extending previous work in Mandarin that has focused primarily on pitch perturbation (Jones & Munhall, 2002; Xu et al., 2004). Both upward and downward F1 shifts elicited group-averaged compensatory responses in the direction opposite to the perturbation, with larger and more reliable adaptation to upward shifts, consistent with findings in English speakers (Villacorta et al., 2007). This upward-shift effect was also replicated in the MRI environment, demonstrating that Mandarin speakers adapt to sustained formant perturbation even during scanning.

Unlike some findings in English, where F1 perturbation can be accompanied by smaller changes in F2 (MacDonald et al., 2010), Mandarin speakers did not show systematic F2 adaptation following the F1 shift. This difference may reflect the specific vowel structure of the stimuli used here. Increasing F1 during production of "dē" shifts the perceived vowel toward "dā", but these two vowels did not substantially differ in F2 in our speakers (Section 4.6.1). Thus, the perturbation may have produced little or no F2 mismatch to drive compensatory adjustment. More broadly, because Mandarin is a tonal language in which lexical tone contrasts are carried by F0, these findings provide an important test of auditory–motor control

models developed largely from non-tonal languages. The results suggest that the neural correlates observed here are broadly consistent with prior English-language work, while extending formant-adaptation findings to Mandarin and strengthening the cross-linguistic generality of auditory–motor control in speech.

The results of this behavioural study allowed us to optimise the experimental design of the subsequent brain-imaging study. Only one consonant-vowel syllable ("dē") and the F1 upward shift were used to ensure a large amplitude and reliable trajectory of adaptation in F1 production. We first compared brain activity during altered feedback (adaptation) with normal feedback (baseline). Results showed increased activity in the Crus I, Crus II, and lobule VII of the left posterior cerebellum and right inferior parietal cortex for adaptation relative to baseline. Results also showed decreased activation during adaptation in the ventral sensorimotor cortex, at the level of the lip and tongue representations, the supplementary motor area (SMA), both bilaterally, and in the right lateral occipital cortex.

Individual differences in the magnitude of F1 reduction were also negatively correlated with adaptation-related activity in a similar network of regions. This more sensitive analysis additionally revealed bilateral changes in sensorimotor cortex at the level of the lip and tongue representations, posterior insula, and parietal operculum, as well as the right cingulate motor area. No brain areas showed activity correlated with changes in F2.

In the following sections, we first discuss the role of the left posterior cerebellum in speech adaptation, interpreting the increased activity there as continuous engagement in predicting the sensory consequences of motor commands. We then consider

the role of sensorimotor areas, where decreased activity is interpreted as a learning-related reduction reflecting recalibrated feedforward control. Next, we discuss how the magnitude of F1 reduction across individuals tracked activity in this network and extend it to secondary somatosensory and cingulate motor regions. Finally, we compare these findings with prior fMRI work on sustained formant perturbation.

4.7.1 The role of left posterior cerebellum in speech adaptation

In the current study, we observed increased activity in the left posterior cerebellum (Crus I, Crus II, and lobule VII) during altered feedback. Speech production can be conceptualised as an adaptive control system: auditory feedback is not used primarily to correct errors online, but instead to update the control policy when a mismatch is detected (Neilson & Neilson, 1987). A central component of this process is the ability to generate rapid predictions of the sensory consequences of motor commands known as a forward model. The forward model is a network that predicts the sensory consequences of a motor command. For example, when you raise your tongue to produce a low F1 vowel sound in response to increased F1 feedback, the forward model uses the motor command to predict the resulting low-F1 sound before the sound actually reaches your ears. One long-standing idea is that the prediction comes from an "efference copy" of motor commands (Miall & Wolpert, 1996). Specifically, when the motor command is generated and sent to articulators, a prediction of its sensory outcome is simultaneously sent to the sensory areas in the brain to compare with sensory input.

The cerebellum has been widely implicated in computing forward models that support speech adaptation (Golfinopoulos et al., 2011; Tourville et al., 2008). Under sustained formant shifts, participants repeatedly experienced a discrepancy between predicted and perceived auditory feedback, which would be expected to engage predictive computations and model updating. One influential account proposes that the cerebellum implements Smith predictors via two forward-model components (Miall et al., 1993). The first component generates rapid predictions of auditory and somatosensory outcomes from speech motor commands. The second component introduces a delayed version of that prediction, enabling comparison with the temporally lagged sensory feedback. This architecture can support both moment-to-moment adjustment and longer-term updating of internal models across changing sensorimotor states (Albert et al., 2009; Kelly & Garavan, 2004; Sami & Miall, 2013). Within this framework, the heightened activity in the left posterior cerebellum observed here is consistent with increased reliance on predictive processing and internal model updating during sustained auditory feedback perturbation.

4.7.2 The role of sensorimotor areas in speech adaptation

We found reduced activity during altered feedback compared with normal feedback in sensorimotor cortex, including regions corresponding to larynx, lip, and tongue representations (Eichert et al., 2021), as well as in the supplementary motor area (SMA) bilaterally. Moreover, activity within a broader set of sensorimotor regions (bilateral sensorimotor cortex, posterior insula and parietal operculum, SMA, and right cingulate motor area) was linearly related to change in F1, linking neural

modulation to the behavioural expression of adaptation. Changes in brain activity of sensorimotor areas were described previously when comparing altered with unaltered auditory feedback. Tourville et al. (2008) reported increased brain activity in ventral motor and somatosensory cortex during unexpected altered feedback compared with unaltered feedback. However, Niziolek and Guenther (2013) found that as speakers adapt more to the altered auditory feedback, they showed decreased brain activity in ventral motor and somatosensory cortex bilaterally.

The involvement of the SMA in speech adaptation aligns with its role in initiating and regulating speech motor commands (Riecker et al., 2005; Tourville & Guenther, 2011). According to the DIVA model, the feedforward control system relies on the SMA to trigger the appropriate motor program at the correct time via cortico–basal ganglia–thalamo–cortical loops that include an SMA "initiation map" on the medial frontal wall. This loop is thought to evaluate whether the current cognitive and sensorimotor context is suitable for launching the next speech sound. When the appropriate context is detected, the basal ganglia engage thalamic relays to activate the relevant SMA initiation unit, which then triggers "readout" of the learned motor program for the current sound. Beyond movement initiation, the SMA also receives auditory information during speech (Dai et al., 2022) and supports transformation of sensory error signals into motor updates through interactions with inferior parietal and primary motor regions (Hickok, 2012; Hickok et al., 2022; Tourville & Guenther, 2011). The SMA has also been implicated in monitoring auditory feedback, especially under demanding conditions (Christoffels et al., 2007; van de Ven et al., 2009). In this context, reduced activity in

sensorimotor cortex and SMA bilaterally is consistent with a shift toward a more efficient control as learning progresses.

One plausible interpretation of these changes in activity is that early in adaptation, the unexpected feedback requires larger articulatory adjustments (e.g. involving laryngeal, lip, and tongue representations) to counteract the perceived formant shift. With continued exposure, feedforward motor commands are recalibrated, enabling speakers to pre-emptively produce the desired acoustic outcome with reduced reliance on feedback-driven correction. The DIVA model provides a principled framework for this interpretation: auditory errors drive updates to feedforward commands in premotor cortex, which then reduce the need for subsequent feedback-based adjustment (Tourville & Guenther, 2011). Learning-related reductions in sensorimotor activity have been reported across motor learning paradigms, and are often interpreted as reflecting increased neural efficiency as a new motor mapping is acquired (Albert et al., 2009; Kelly & Garavan, 2004; Sami & Miall, 2013). Accordingly, reduced activation in articulatory motor cortex may reflect recalibrated motor commands that achieve the target vowel with less corrective effort, complementing the cerebellar forward-modelling processes described above. More broadly, these results support a transition from feedback control to updated feedforward control during sensorimotor adaptation (Houde & Nagarajan, 2011).

4.7.3 Brain activity tracks individual differences in F1 adaptation

Beyond the categorical contrast of altered versus normal feedback, the magnitude of F1 reduction in individual participants was linearly correlated with adaptation-

related changes in brain activity. This parametric analysis recapitulated the network revealed by the categorical contrast (left posterior cerebellum, ventral sensorimotor cortex, SMA, and right inferior parietal cortex) but additionally exposed adaptation-related modulation in sensorimotor cortex bilaterally at the level of the lip and tongue representations, posterior insula and parietal operculum bilaterally, and the right cingulate motor area.

The recruitment of posterior insula and parietal operculum—components of the secondary somatosensory network—is consistent with a role for somatosensory monitoring of articulatory state as feedforward commands are updated (Golfinopoulos et al., 2011; Tourville & Guenther, 2011). The right cingulate motor area, identified across neuroimaging studies as part of the medial-wall motor system that contributes to action selection and the adjustment of motor commands under changing demands (Amiez & Petrides, 2014; Picard & Strick, 1996), is well placed to support the updating of ongoing speech motor commands as auditory targets and articulator outputs are reconciled across trials.

By contrast, no brain regions showed activity that correlated with change in F2. This dissociation aligns with the behavioural finding that F1 was the acoustic dimension along which sustained perturbation drove robust compensatory change, and reinforces the interpretation that this network specifically supports learning-driven recalibration of feedforward speech motor commands rather than a generic response to altered auditory feedback.

4.7.4 Comparison with Floegel et al. (2020) study

The present results can be situated within the small fMRI literature on sustained altered auditory feedback, most directly the work of Floegel et al. (2020). Floegel et al. (2020) compared spectral (formant-shift) and temporal (delayed-feedback) perturbations and reported partly separable, domain-tuned responses. Spectral adaptation was associated with increased activity in inferior frontal gyrus and superior temporal sulcus, predominantly in the right hemisphere, whereas temporal adaptation was associated with predominantly left-lateralised activity in superior temporal gyrus and superior temporal sulcus, with a corresponding left-lateralised pattern of activity in inferior frontal gyrus. Therefore, Floegel et al. (2020) proposed a domain-tuned refinement of the DIVA model in which mismatch processing in superior temporal cortex is right-lateralised for spectral errors and left-lateralised for temporal errors, with mismatch signals conveyed via temporo-frontal pathways to generate corrective commands—thereby challenging the previously described simple right-lateralised feedback versus left-lateralised feedforward division (Tourville et al., 2008).

The current findings converge with Floegel et al. (2020) in implicating a distributed cortical network during sustained formant adaptation but diverge on the hemispheric lateralisation. Where their spectral perturbation produced predominantly increased activity in the middle portion of the right superior temporal sulcus and the posterior right inferior frontal gyrus, the present whole-brain analysis revealed changes in brain activity bilaterally within ventral sensorimotor cortex, posterior insula, and parietal operculum, alongside a left-lateralised increase in posterior

cerebellum (Crus I, Crus II, lobule VII) and a right-lateralised increase in inferior parietal cortex (angular gyrus). Several methodological differences may contribute to this discrepancy. First, Floegel et al. (2020) adopted a region-of-interest approach restricted to superior temporal and prefrontal cortex, whereas the present analysis was unconstrained at the whole-brain level. Second, the present parametric analysis scaled brain activity with the magnitude of F1 change, providing a more direct link between change in brain activity and behavioural performance rather than a categorical contrast alone.

4.8 Summary

In this study, we developed a novel Mandarin version of the altered auditory feedback task with formant perturbation and showed that Mandarin Chinese speakers robustly adapt by decreasing F1 production under sustained increase of auditory feedback in F1. This adaptation showed a similar amplitude and trajectory to that seen in English speakers. Using fMRI, we identified a brain network of this short-term speech adaptation: as participants learned to decrease F1 to counteract an upward F1 perturbation in their feedback, activity increased in the left posterior cerebellum and decreased in ventral sensorimotor cortex and supplementary motor area. Activity in sensorimotor cortex, at the level of the lip and tongue representations, posterior insula and parietal operculum, all bilaterally, and right cingulate motor area also showed correlation with the amplitude of F1 decrease during the task. This pattern supports a model in which the cerebellum plays a central role in generating and updating predictions about the sensory consequences of speech

motor commands during sustained mismatch in auditory feedback, while sensorimotor regions show learning-related reductions consistent with the recalibration and increasingly efficient deployment of feedforward speech motor commands.

Together, these behavioural and brain imaging results demonstrate that altered auditory feedback engages a distributed network supporting predictive control and adaptive updating in speech. They extend evidence for formant-based sensorimotor learning to a tonal-language population, strengthening the cross-linguistic generality of core auditory–motor control mechanisms, and clarifying how the cerebellum's role in sensory-motor prediction and sensorimotor cortex's role in motor implementation interact to maintain accurate speech output in the face of persistent sensory perturbation.

5

Changes in resting-state brain connectivity due to speech adaptation

5.1 Abstract

Functionally related brain networks are engaged even without task demands. This resting-state brain connectivity can be captured by low-frequency fluctuations in brain activity measured using functional MRI. One major proposal of resting-state brain connectivity is that the resting brain actively processes previous experience. Here, we tested whether speech adaptation can change resting-state brain connectivity. We scanned 48 native Mandarin Chinese speakers while they adapted to altered auditory feedback. Two sequences of resting-state images (10 minutes each)

were collected before and after the adaptation task to compare brain connectivity in the resting state. Speech adaptation modulated resting-state brain connectivity by increasing functional coupling within three major networks: a superior temporal auditory network, a dorsomedial sensorimotor network, and a network connected to the left posterior cerebellum. The results suggest a reorganisation of coordinated brain activity across networks beyond the immediate demands of the task. Importantly, we show that these resting-state networks are modulated by prior speech motor learning, pointing to a connection between learning-related neuroplasticity and resting-state activity.

5.2 Introduction

A growing body of work suggests that speech adaptation to altered auditory feedback is supported by distributed sensorimotor networks. Task-based studies using unexpected perturbations have shown that the brain rapidly detects mismatches between intended and perceived speech sounds and recruits sensorimotor areas and cerebellum to generate corrective commands (Tourville et al., 2008; Zheng et al., 2013). More recent work further indicates that the neural implementation of feedback-based correction may depend on the acoustic domain of the perturbation, with distinct lateralisation patterns for spectral (i.e. formant perturbation) and temporal (i.e. delayed auditory feedback) deviations (Floegel et al., 2020).

Beyond haemodynamic measures, electrophysiological studies have highlighted that speech adaptation involves dynamic interactions across frequency bands and changes in inter-regional communication, consistent with the establishment of a new

sensorimotor mapping during learning (Sato & Shiller, 2018; Sengupta & Nasir, 2015, 2016). Evidence from clinical populations further supports this network view: patterns of reduced adaptation alongside enhanced online compensation in spinocerebellar ataxia (Houde et al., 2019; Li et al., 2019; Parrell et al., 2017) and Parkinson's disease have been interpreted as shifts in the weighting of feedforward versus feedback control following disruption to the cerebellum or basal ganglia (Mollaei et al., 2013, 2016). Together with structural imaging evidence linking individual variability in speech adaptation to morphology in parietal–temporal and frontal regions (Chen et al., 2021), these findings motivate the need to characterise not only task-evoked responses, but also the intrinsic network architecture that may reflect speech adaptation.

Large-scale brain networks are active and dynamically organised even in the absence of explicit task demands (Raichle, 2010). Early evidence for such intrinsic organisation came from the observation that low-frequency brain fluctuations (0.1–0.01 Hz) were highly correlated across regions within the motor cortex (Biswal et al., 1995). These low-frequency fluctuations in the blood-oxygen-level-dependent (BOLD) signal can be used to delineate sets of spatially distributed brain areas whose activity is temporally correlated, commonly referred to as resting-state brain networks (Damoiseaux et al., 2006; Greicius et al., 2009). Resting-state fMRI (rs-fMRI) provides a standard way to study these networks by collecting functional MRI data while participants rest quietly without performing an explicit task.

Resting-state fMRI examines intrinsic brain activity that is not time-locked to specific behavioural events. This approach has revealed reproducible large-scale

networks in healthy adults (Smith et al., 2011). Importantly, alterations in resting-state network organisation and connectivity have also been consistently reported across a range of neurological disorders (Tahmasian et al., 2017). Resting-state connectivity therefore provides a useful window onto the organisation and integrity of distributed functional brain systems.

Beyond characterising stable network connectivity, rs-fMRI can also capture learning-related changes in intrinsic brain activity. For example, Albert et al. (2009) showed that visuomotor learning modulated resting-state connectivity within a fronto-parietal network and a cerebellar network. Such learning-induced changes were driven by motor learning not by motor performance. Both of these networks were engaged during learning of visuomotor tasks. The result was interpreted as modulation of resting-state brain networks by prior visuomotor learning. In speech adaptation, Floegel et al. (2020) provided initial evidence that adaptation to formant perturbation is associated with changes in resting-state connectivity, such that greater behavioural adaptation was linked to increased fronto-temporal connectivity in the right hemisphere and increased temporo-parietal connectivity in the left hemisphere.

Sustained altered auditory feedback during speech production can induce sustained modifications of the motor commands used to produce speech sounds (as discussed in Chapter 2). Analogous to visuomotor learning (Albert et al., 2009), this form of speech motor learning may therefore alter resting-state connectivity within networks that support the detection of auditory mismatch, the transformation of sensory prediction errors into corrective motor commands, and the stabilisation of updated speech motor plans. Accordingly, in the current study we collected

two ten-minute rs-fMRI scans, one before and one after participants performed the speech adaptation task. We hypothesised that sensorimotor adaptation of speech would be followed by an updated pattern of resting-state brain connectivity, particularly within auditory and sensorimotor areas and their interactions with broader control networks implicated in speech motor learning.

5.3 Materials and Methods

5.3.1 Participants

The same forty-eight (23 males and 25 females) Mandarin Chinese speakers described in Chapter 4 took part in this study. See Chapter 4 for details.

5.3.2 Experimental procedure

Two sessions of ten minutes (750 scans) of resting-state fMRI images were collected immediately before and after the speech adaptation task. Participants were required to keep their eyes open and look at a fixation cross on the screen. For the complete experimental procedure, please see Chapter 4.

5.3.3 Resting-state scan parameters

Resting-state fMRI images (750 volumes) were acquired using the multi-band accelerated echo-planar sequence: TR = 800 ms, TE = 30 ms, flip angle = 80°, FoV = 216 mm, multiband acceleration = 6, phase encoding direction = anterior to posterior, voxel size = 2.4 mm isotropic. A single-band image was collected before each multi-band sequence and was used to correct the multi-band induced

distortions to the image.

5.3.4 ICA and dual-regression analysis

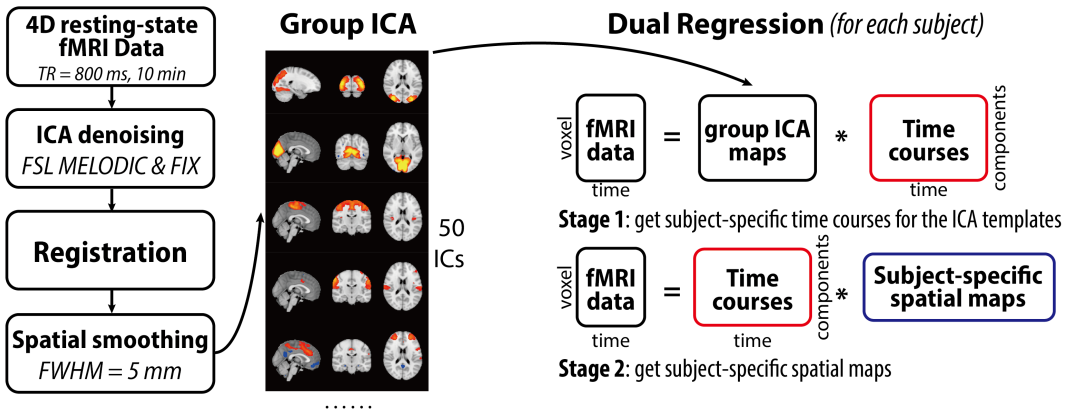

Figure 5.1: ICA and dual-regression analysis
An illustration of preprocessing, group ICA, and dual regression analysis.

Data analysis was conducted using FSL (FMRIB Software Library, Jenkinson et al., Jenkinson et al., 2012). First, all resting-state functional images underwent these initial steps: removal of skull and scalp from the images using BET (Smith, 2002), motion correction using MCFLIRT (Jenkinson et al., 2002), high-pass temporal filtering (100-s), and fieldmap unwarping using FUGUE to correct for distortions in the images and improve registration (Jenkinson, 2002). Then, individual resting-state datasets were analysed by subject-level independent component analysis (ICA) using FSL's MELODIC (Multivariate Exploratory Linear Optimized Decomposition into Independent Components) and cleaned using FIX (FMRIB's ICA-based Xnoiseifier) to remove noise components (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014). FIX was trained using the signal and noise components identified from 10 subjects and classified by the author (QY) using

the protocol described by Griffanti et al. (2017). The output was then applied to all participants in all sessions. All denoised images were non-linearly registered to the 2-mm MNI-152 template using FNIRT and spatially smoothed using an FWHM = 5 mm Gaussian kernel.

Group ICA was performed based on the spatially smoothed data using MELODIC to extract 50 independent components (ICs, 46 signal, 4 noise). Then, dual regression analysis (Beckmann et al., 2009) was conducted. This involves using the group-spatial maps of each IC to extract a subject-specific time course from each individual's data and then, in a second regression step, the subject-specific time course was used to extract a subject-specific spatial map of that IC. These subject-specific spatial maps were statistically compared from before to after the speech adaptation task. Non-parametric permutation tests (5000 permutations) and threshold-free cluster enhancement (TFCE) were used to infer significant changes in connectivity ($p < 0.05$, corrected). See Figure 5.1.

To test the relationship between individual behavioural adaptation of F1 and task-related changes in resting-state connectivity, we extracted the median connectivity value from clusters that showed significant pre–post task differences. We then conducted correlation analyses to examine whether the change in produced F1 in response to the F1 shift was associated with changes in resting-state connectivity.

5.3.5 Seed-based connectivity analysis

Beyond whole-brain group ICA and dual regression analysis, we also conducted seed-based connectivity analysis. In Chapter 4, we found increased activity in left

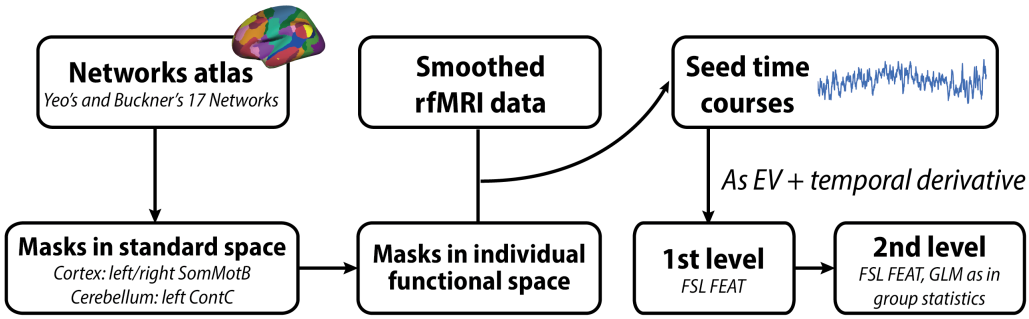

Figure 5.2: Seed-based connectivity analysis
An illustration diagram of seed-based connectivity analysis.

posterior cerebellum and decreased activity in ventral sensorimotor cortex bilaterally. Based on this finding, we selected three seed regions from the 17-network atlas of Thomas Yeo et al. (2011) and its connecting network within cerebellum from Buckner et al. (2011). Specifically, we selected somatomotor network B (Network 4; SomMotB) from Yeo's 17-network cortical parcellation in both hemispheres. This network includes tongue-related sensorimotor representations and auditory-processing regions in the superior temporal gyrus, and it overlapped with areas showing decreased activity during the task. We also selected control network C (Network 13; ContC) from Buckner's 17-network cerebellar parcellation in the left cerebellum, corresponding to cognitive-control regions within Crus I and Crus II; this network overlapped with regions where we found increased activity during the speech adaptation task.

First, both cortical and cerebellar atlases in FSL standard MNI-152 2-mm space were divided into left and right hemispheres. The hemispheric masks were then registered to each participant's native functional space using FLIRT. Individual mean time courses were extracted from each seed region. First-level GLM analyses

were subsequently performed for each participant using FSL FEAT (statistics-only mode). The spatially smoothed functional data were used as the input dataset. The extracted seed time course, along with its temporal derivatives, were entered into the model as explanatory variables (EVs). Group second-level analysis were performed using paired t-tests to compare the two time points. Significance was inferred using cluster statistics with an initial height threshold of $Z > 3.1$ and a corrected cluster extent threshold of $p < 0.05$ (Jezzard et al., 2001). See Figure 5.2.

5.4 Results

5.4.1 Increased resting-state connectivity in superior temporal cortex

We looked for changes in resting-state networks that reflected acquisition of the adapted state, which had occurred in the interval between the two resting-state scans. Following adaptation, connectivity increased between the network encompassing the posterior superior temporal cortex bilaterally, known as the "auditory" network (Smith et al., 2009) and the right posterior insula, right Heschl's gyrus, and parietal operculum (Figure 5.3 A). It is worth noting that the spatial extent of this "auditory" network is not limited to the superior temporal cortex but encompasses the inferior parietal cortex dorsally and extends medially to the posterior insula cortex. These latter two areas in the right hemisphere were where functional connectivity was increased following adaptation. There was an equivalent increase in the resting-state connectivity in the left hemisphere at a subthreshold level. However, changes in this region did not survive the cluster-level thresholding based on extent.

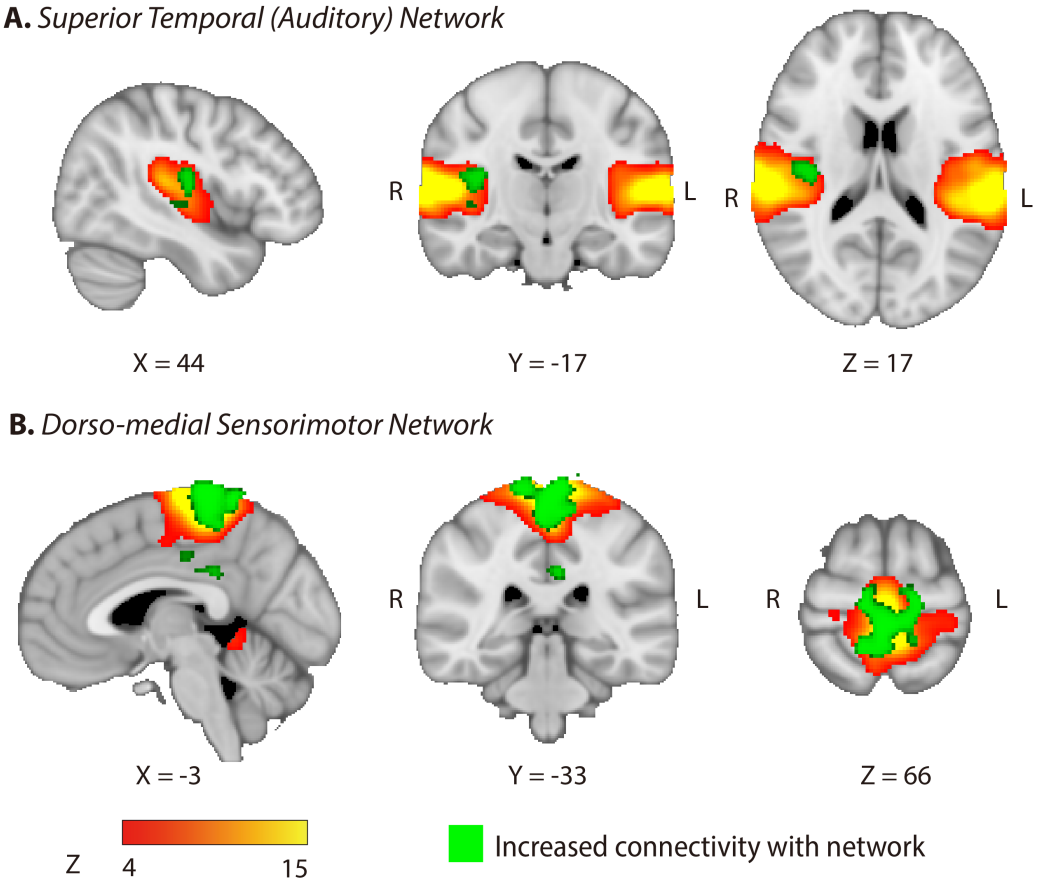

Figure 5.3: Increased resting-state connectivity in superior temporal and dorsomedial sensorimotor networks

Dual-regression results showing higher resting-state connectivity after speech adaptation. Following adaptation, connectivity increased between right posterior insula and parietal operculum and the rest of the "auditory network" involving the superior temporal cortex bilaterally (A), and in a portion of a sensorimotor network located dorsomedially (B). The two group networks identified using ICA are shown in the red–yellow Z-statistic maps overlaid on the MNI-152 template image and thresholded at $Z > 4$. The clusters in green show voxels where connectivity with these networks was significantly changed with $p < 0.05$ with threshold-free cluster enhancement (TFCE). The figure was created using FSLeyes from FSL. L–left; R–right. The relevant coordinate in standard space is provided below each brain slice.

In the correlational analysis with behavioural data, we found that individuals with larger compensation (i.e. greater F1 decrease in response to the increased F1 feedback), exhibited larger changes in resting-state connectivity in right superior temporal cortex ($r = -0.298$, $p = 0.020$, Figure 5.4).

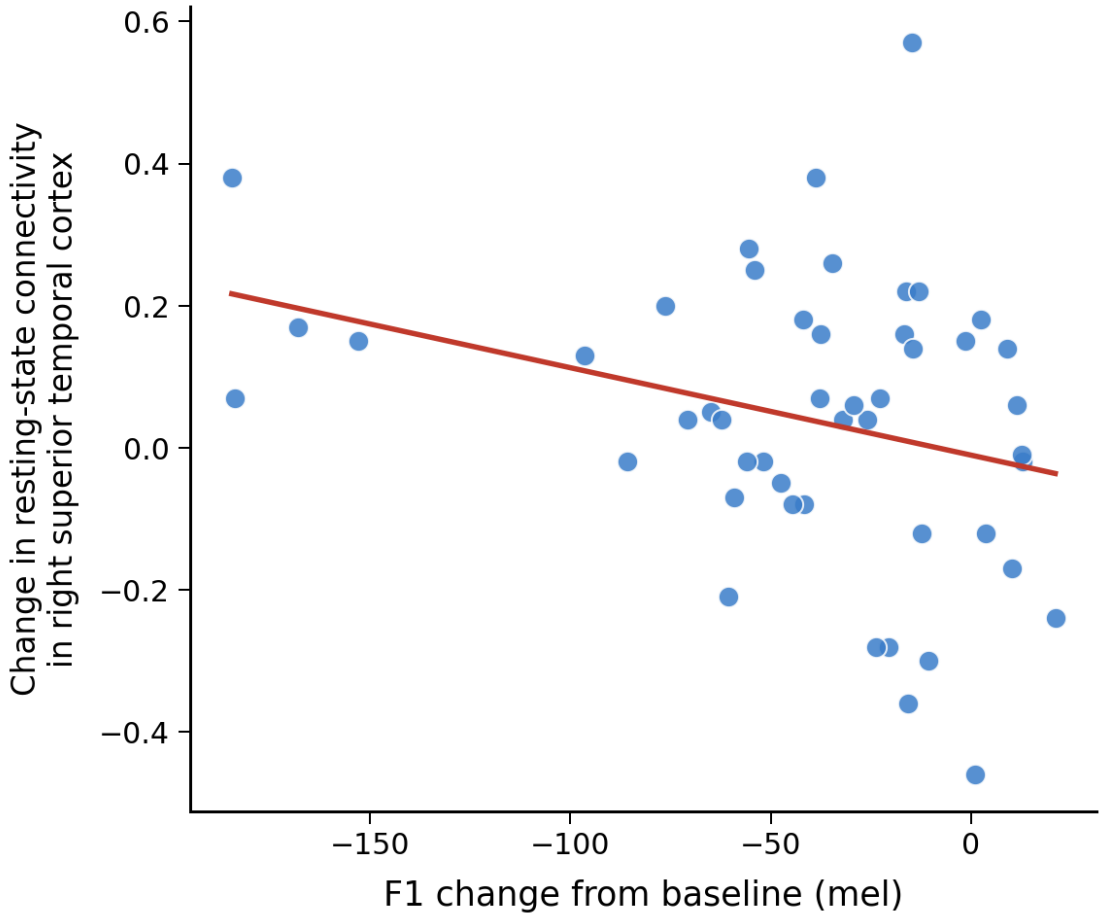

Figure 5.4: Correlation between F1 change from baseline and change in resting-state connectivity in right superior temporal cortex
Each circle represents data for one participant. The red line shows the linear regression fit. A significant negative correlation was observed between the magnitude of F1 change from baseline (in the last 40 trials during adaptation block) and change in resting-state connectivity in a cluster located in the right superior temporal cortex ($r = -0.298$, $p = 0.020$). See text for details.

Notably, four participants showed change in F1 adaptation that exceeded the 110 Mel perturbation itself. These "over-compensators" sit at the extreme end of the distribution of F1 change and could arise from a stronger reliance on auditory relative to somatosensory feedback, producing a larger corrective response for a given perceived error, or from greater instability of sensorimotor integration to the perceived mismatch itself. Because a correlation can be disproportionately influenced by a small number of extreme points, we tested whether the correlation between

F1 change and change in right superior temporal connectivity was driven by these over-compensators. After excluding participants whose F1 change exceeded the 110 Mel perturbation, the correlation was no longer significant ($r = -0.195$, $p = 0.206$).

5.4.2 Increased resting-state connectivity in dorsomedial sensorimotor cortex

Following adaptation, connectivity also increased in a portion of a sensorimotor network located dorsomedially. It included primary sensorimotor cortex bilaterally. Connectivity between this network and a portion of the right posterior cingulate cortex that lay outside the network was also increased (Figure 5.3 B).

No significant correlation was found between F1 change and the change of resting-state connectivity within either the superior temporal network or the dorsomedial sensorimotor network.

5.4.3 Increased resting-state connectivity in occipital cortex

We also found increased connectivity between occipital cortex and the rest of the occipital resting-state networks bilaterally (Figure 5.5).

5.4.4 Changes in functional connectivity: seed-based analysis

Seed-based analysis revealed increased functional connectivity from the ventral sensorimotor cortex and superior temporal gyrus with the right medial occipital cortex (Figure 5.6 A).

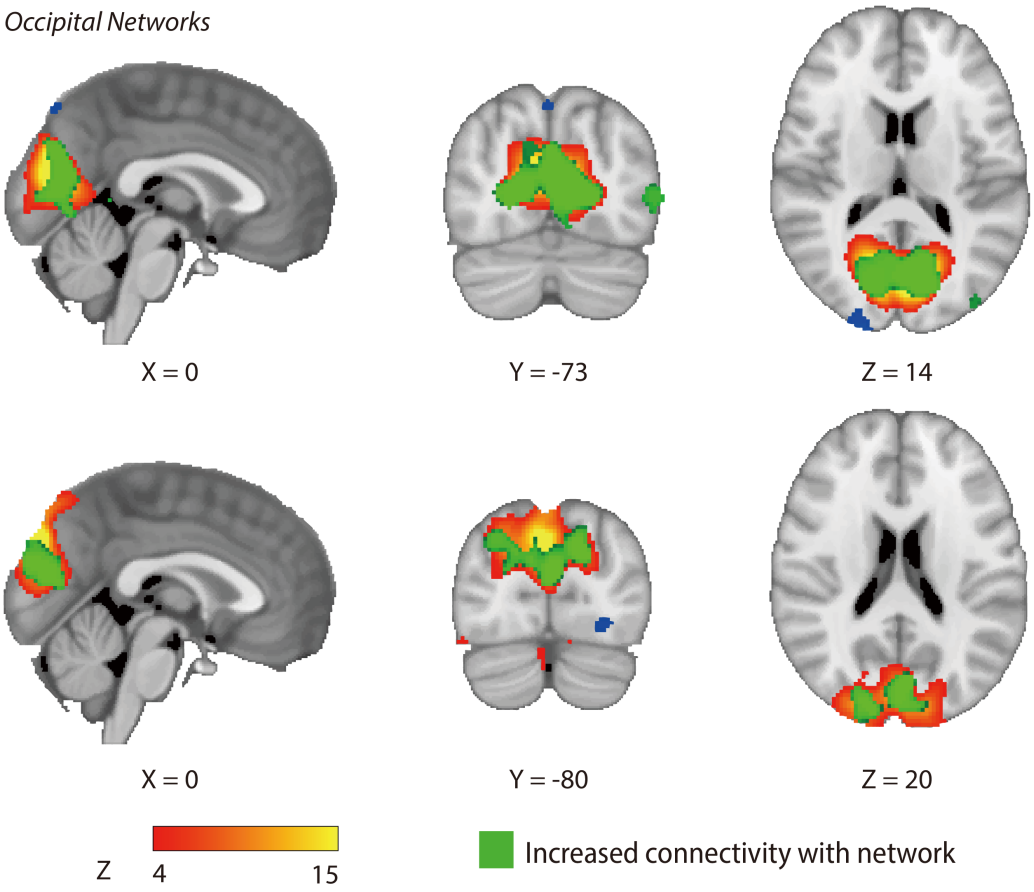

Figure 5.5: Increased resting-state connectivity in occipital cortex
Dual-regression results showing higher resting-state connectivity after the speech motor learning. Both figures show increased connectivity between occipital cortex (green) and the rest of the bilateral occipital visual network (positive: red–yellow; negative: blue). The cluster in green shows voxels with $p < 0.05$ with threshold-free cluster enhancement (TFCE). See caption to Figure 5.3 for details.

Seed-based analysis revealed increased functional connectivity within the left posterior cerebellar control network, as well as between the left posterior cerebellum and the left dorsal somatosensory cortex and right medial occipital cortex (Figure 5.6 B).

There were no significant reductions in resting-state connectivity that were evident in either the group ICA and dual-regression or the seed-based connectivity analyses.

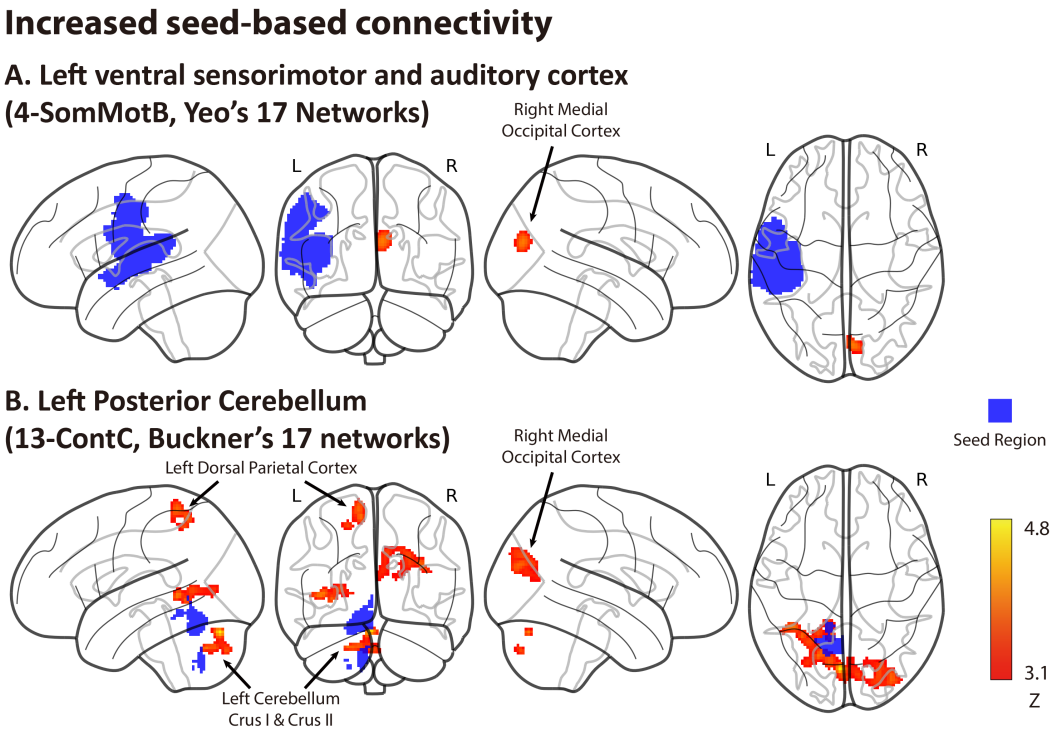

Figure 5.6: Increased seed-based connectivity
Seed-based connectivity results showing increased connectivity after speech adaptation. **A.** Increased connectivity between the left ventral somatosensory-motor network (blue) and the right medial occipital cortex (red–yellow). **B.** Increased connectivity between the left posterior cerebellar network (blue), and the left cerebellum Crus I and Crus II (within cerebellum), and left dorsal parietal cortex, and right medial occipital cortex (red–yellow). The seed regions (blue) and the statistical maps (red–yellow) are overlaid on "glass" brains using Nilearn.

5.5 Discussion

Here, we examined whether speech adaptation reshaped intrinsic brain activity. Analysis of resting-state fMRI scans collected before and after the adaptation phase revealed network-level reorganisation across multiple brain networks. Specifically, we observed (i) increased connectivity within a right-lateralised superior temporal ("auditory") network, (ii) stronger coupling involving dorsomedial sensorimotor cortex bilaterally, and (iii) increased resting-state activity within an occipital ("visual") network. Extending these network-level effects, seed-based

analyses showed increased coupling from ventral sensorimotor cortex and superior temporal gyrus to right medial occipital cortex, as well as strengthened connectivity within the left posterior cerebellar control network and between left posterior cerebellum and left dorsal somatosensory cortex and right medial occipital cortex. Together, these findings suggest that a relatively brief period of speech adaptation to sustained altered auditory feedback can induce measurable changes in intrinsic functional brain connectivity, extending beyond the task-evoked neural responses observed during speech adaptation itself. More broadly, the results support the view that sensorimotor learning is accompanied by reconfiguration of large-scale networks that may reflect consolidation, re-weighting of sensory information, or altered readiness to process altered auditory feedback.

5.5.1 Updated superior temporal ("auditory") network

Following speech adaptation, we observed increased resting-state connectivity within the superior temporal network, limited to the right hemisphere, including the right posterior insula and parietal operculum. Also, speakers who showed more change in F1 in response to increase in F1 during auditory feedback exhibited larger changes in resting-state connectivity in the superior temporal network. We suggest that adaptation, which required sustained attention to auditory feedback and continuous processing of vowel formant errors, led to a short-term strengthening of functional integration within this auditory-related system. Repeated exposure to altered feedback and recurrent detection of sensory discrepancies may have increased the efficiency with which auditory information is integrated with regions implicated in monitoring internal sensory states (Hickok et al., 2011; Tourville et al., 2008).

The observed post-adaptation increase in connectivity can therefore be interpreted as a candidate marker of an individual's sensitivity and short-term plasticity in processing the mismatch in auditory feedback: auditory areas, together with posterior insula/opercular areas, may have been transiently upregulated by the need to monitor perturbed feedback and support corrective responses.

The involvement of the right posterior insula and parietal operculum is consistent with their broader roles in sensorimotor processing. Although these regions are not traditionally framed as core speech areas, both have been implicated in integrating sensory signals with action and in representing bodily states relevant for speech motor control. However, the left anterior insula has been considered as a part of the speech production network, and prior work suggests that it participates in speech motor planning and execution (Dogil et al., 2002; Dronkers, 1996; Moser et al., 2009). Speech-related activity in the right anterior insula is sometimes reported in the context of emotional prosody (Sammler et al., 2015) and patients with Broca's aphasia recruit insular cortex bilaterally during speech entrainment, in which mimicking audiovisual speech stimuli enables fluent speech production in real time (Fridriksson et al., 2018; Fridriksson et al., 2012). Similarly, parietal and posterior opercular regions have been linked to multisensory integration and sensorimotor transformations relevant to action (Mountcastle, 1995). In our task, participants likely relied not only on auditory error signals but also on proprioceptive and somatosensory cues used to stabilise articulation, and this multisensory context may have increased functional coupling between superior temporal auditory regions and insular/opercular components of the network. From this perspective, the connectivity increases may reflect tighter coordination between

perceptual analysis of auditory errors and internal modelling processes that support sensorimotor adaptation.

Notably, the effects observed in the current study were most robust in the right hemisphere. This pattern of lateralisation is consistent with evidence that right auditory cortex is more strongly involved in spectral processing of speech (vowel formants), whereas left auditory cortex is more strongly associated with temporal processing (Albouy et al., 2020; Flinker et al., 2019; Floegel et al., 2020). Importantly, this right auditory region overlapped with the area that showed F1 change related modulation during the task fMRI analysis reported in Chapter 4, providing converging evidence that the same spectral-processing resources engaged during online compensation for the F1 perturbation also exhibit altered resting-state connectivity. The strengthened right auditory network connectivity observed at rest may therefore reflect an after-effect of repeatedly engaging these spectral processing resources, such that functional integration within the relevant network remains elevated following learning. More generally, this finding suggests that resting-state measures can capture short-term imprints of speech learning in networks selectively recruited by the specific error dimension of the perturbation.

5.5.2 Updated dorsomedial sensorimotor network

Following speech adaptation, we also observed stronger resting-state functional connectivity involving portions of the dorsomedial sensorimotor cortex network bilaterally, which extended to include increased connectivity with the left posterior cingulate cortex. This pattern indicates that speech adaptation can strengthen

interactions within sensorimotor circuits even in the absence of ongoing task demands. During sustained altered auditory feedback, participants repeatedly adjusted articulatory output to minimise perceived errors, implying iterative updating of the mapping between auditory error signals and motor commands. The post-adaptation increase in dorsomedial sensorimotor connectivity may therefore index a more tightly integrated circuit supporting the transformation of sensory feedback into corrective motor adjustments. In other words, the network-level changes observed at rest may reflect the lingering organisation of neural resources that were repeatedly co-activated to support corrective motor learning during adaptation.

Notably, the observed changes were centred on dorsomedial rather than ventrolateral sensorimotor cortex, the latter being more associated with lip and tongue representations (Eichert et al., 2020). Jossinger et al. (2026) showed that ventral sensorimotor cortex is involved when distinguishing syllables by place of articulation whereas activity in dorsal sensorimotor cortex was more sensitive to the voice onset time in syllable production. This distinction suggests that dorsal regions encode higher-order temporal and coordinative parameters of speech rather than articulator-specific representations. The increased resting-state connectivity within dorsomedial sensorimotor cortex following adaptation seen here may also reflect the integration of auditory feedback with motor commands.

The involvement of cingulate cortex is compatible with its role at the interface of motor control and higher-order performance monitoring. Cingulate motor areas have direct access to the skeletomotor apparatus (Paus, 2001), and cingulate regions implicated in vocal control also receive input from auditory association

cortex (Paus et al., 1993). A complementary perspective comes from recent evidence that primary motor cortex contains interleaved, non-effector-specific zones: Gordon et al. (2023) identified a somato-cognitive action network (SCAN) that occupies the dorsomedial gaps between hand, foot, and face representations and is tightly coupled to cingulo-opercular control regions, insula, and specific cerebellar territories, and is proposed to support integrative aspects of action rather than the execution of effector-specific movements. Our findings of changes in resting-state activity in dorsomedial sensorimotor cortex, cingulate cortex, and the seed-based results showing increased connectivity between left posterior cerebellum and left dorsal parietal cortex are together consistent with reorganisation of a SCAN-like inter-effector system rather than of orofacial effector cortex. In this context, increased coupling between dorsomedial sensorimotor regions and cingulate cortex following speech adaptation may reflect strengthened coordination between motor output, sensory monitoring, and the control processes that support the evaluation and updating of speech actions under altered auditory feedback.

Overall, these updates to the sensorimotor network align with the broader motor learning literature showing that brief training can induce measurable changes in resting-state connectivity, often interpreted as reflecting early consolidation or stabilisation of newly acquired motor skills (Albert et al., 2009; Farrens et al., 2023). Although much of this evidence comes from visuomotor learning, the present findings support the extension of these principles to speech motor learning.

A further possibility to consider is that increased dorsomedial sensorimotor coupling reflects a transient increase in "readiness" of the sensorimotor system to respond to auditory feedback. Following adaptation, participants may remain

biased toward monitoring and correcting speech output, such that the intrinsic organisation of sensorimotor networks is temporarily altered in a way that supports rapid corrective responses. Although resting-state data cannot directly adjudicate between consolidation and heightened readiness accounts, both converge on the idea that speech adaptation to sustained altered auditory feedback leaves measurable after-effects in intrinsic auditory-motor integration. Behaviourally, for example, we often see retention of compensatory responses when the normal feedback is restored (i.e. after-effect, see results in Chapter 4). Also, speech adaptation can generalise to sounds that were not produced under altered feedback (Houde & Jordan, 1998; Jones & Munhall, 2005; Villacorta et al., 2007). Future work could test these alternatives by tracking the time course of resting-state changes, or by relating individual differences in post-adaptation connectivity shifts to behavioural indices of retention or after-effects.

5.5.3 Updated cerebellum network

We observed increased seed-based connectivity within the left posterior cerebellar control network (Crus I and Crus II), as well as enhanced connectivity between the left posterior cerebellum and left dorsal parietal cortex. The cerebellum has been widely implicated in computing forward models that support speech adaptation (Golfinopoulos et al., 2011; Tourville et al., 2008). Under sustained formant shift, participants repeatedly experienced a discrepancy between predicted and perceived auditory feedback, a manipulation expected to engage predictive computations and drive updating of internal models.

One influential account proposes that the cerebellum implements a Smith predictor architecture comprising two forward-model components (Miall et al., 1993). The first generates rapid predictions of auditory and somatosensory consequences from outgoing speech motor commands. The second produces a temporally delayed version of that prediction, enabling comparison with delayed sensory feedback. Such an architecture supports both rapid, online correction and longer-term recalibration of internal models across changing sensorimotor states (Albert et al., 2009; Kelly & Garavan, 2004; Sami & Miall, 2013).

Within this framework, the increased intrinsic connectivity within left posterior cerebellum and between the cerebellum and dorsal parietal cortex suggests plastic changes in a feedforward control network repeatedly engaged during adaptation. Speech adaptation is typically incomplete: while auditory feedback is immediately shifted outside its target region, somatosensory feedback initially remains within its expected bounds. As feedforward commands gradually adapt to compensate for the auditory perturbation, the resulting articulatory changes can push somatosensory feedback away from its original target. This creates a secondary mismatch that recruits somatosensory feedback control mechanisms, which may partially counteract auditory-driven corrections (Guenther, 2016). The observed strengthening of connectivity within the left posterior cerebellar network and its coupling with dorsal parietal cortex is therefore consistent with sustained recalibration of forward models and updating of somatosensory state representations. Rather than reflecting transient task engagement, these resting-state changes likely index persistent reorganisation of cerebellar–parietal connectivity that mediates predictive control and sensorimotor integration during speech motor learning.

5.5.4 Updated occipital ("visual") network

Increased resting-state activity was observed within an occipital ("visual") network following speech adaptation. We also found increased connectivity between the seed of left posterior cerebellum and right medial occipital cortex. Changes in the visual system were not anticipated following this task, which primarily places demands on the motor, auditory, and somatosensory systems. A plausible explanation is that these connectivity changes within the occipital cortex reflect the non-essential visual processing during the task, as participants continuously viewed words presented on the screen during speech production. In this sense, the resting-state visual network connectivity increases may represent a modality-specific after-effect of sustained visual stimulation and repeated engagement of visual processing. Prior studies have similarly reported enhancements in resting-state activity or connectivity within visual networks following exposure to visual input or visual tasks (Bang et al., 2018; Guidotti et al., 2015; Stevens et al., 2009).

5.5.5 Limitations

A lack of control limitation applies to the resting-state analyses. Because the pre- and post-adaptation scans were separated by a fixed period of speech adaptation task, changes in connectivity are confounded with the simple passage of time and with non-specific effects of fatigue, arousal, or vigilance across the scanning session, rather than being uniquely attributable to speech adaptation. This concern is evidenced by the finding that connectivity within the occipital "visual" network also increased after adaptation, an effect that cannot easily be explained by auditory-motor learning and is more plausibly attributed to sustained visual fixation on

the on-screen syllable throughout the session. If a similarly non-specific process contributed to the visual network change, it is difficult to rule out a comparable contribution to the auditory and sensorimotor network changes of primary interest. This limitation could be addressed by including a control group who complete two resting-state scans separated by an equivalent period of syllable production under normal, unperturbed feedback, isolating connectivity changes specifically attributable to adaptation from those driven by time in the scanner or sustained task engagement more generally.

5.6 Summary

In this chapter, we examined resting-state brain activity before and after speech adaptation to sustained altered auditory feedback. By focusing on intrinsic functional connectivity, this approach allowed us to capture learning-related changes that extend beyond transient task-evoked responses. Our results revealed network-level reorganisation following adaptation, indicating that speech motor learning leaves measurable changes in resting-state networks.

Specifically, speech adaptation increased resting-state connectivity within a right-lateralised superior temporal ("auditory") network, particularly involving the right posterior insula and parietal operculum. This pattern is consistent with strengthened integration mechanisms supporting spectral error monitoring and aligns with the established right-hemisphere lateralisation for spectral auditory processing. Adaptation also enhanced connectivity within a cerebellar control network, including increased coupling within the left posterior cerebellum (Crus I and Crus

II) and between left posterior cerebellum and left dorsal sensorimotor cortex. In addition, we observed strengthened bilateral connectivity in dorsomedial sensorimotor cortex, extending to the left cingulate cortex, suggesting reinforcement of circuits involved in motor planning, monitoring, and error-based updating of forward models.

Together, these findings indicate that repeated recalibration of the mapping between perturbed auditory feedback and articulatory commands induces persistent modifications in both feedback and feedforward control systems. Rather than being confined to task execution, these changes appear to reorganise intrinsic sensorimotor learning networks at rest. Increased connectivity within an occipital ("visual") network likely reflects sustained visual demands during syllable reading, further demonstrating that resting-state networks reflect recently engaged task experience.

Overall, the results demonstrate that speech adaptation to altered auditory feedback can produce measurable changes in resting-state functional connectivity across auditory, cerebellar, and sensorimotor networks. These "after-effects" reflect task-specific reconfiguration of intrinsic brain networks, suggesting that resting-state connectivity provides a sensitive index of the neuroplasticity supporting speech motor control.

6

General Discussion

6.1 Summary of findings

The aim of this thesis was to investigate the neural correlates of speech adaptation during altered auditory feedback. I first used transcranial direct-current stimulation to test whether modulating brain activity in the left speech motor cortex and right cerebellum can enhance speech adaptation. The study was designed to replicate the results of Lametti et al. (2018a), in which stimulation to the left speech motor cortex enhanced adaptation in both F1 and F2, whereas stimulation to the right cerebellum enhanced F1 adaptation only. We used a randomised, double-blind, sham-controlled design; methods and analysis were pre-registered. The original study used a cerebellar montage that placed the cathode on the right cheek

and secured the electrodes with bandages around the jaw, potentially affecting speech production. To avoid this issue, we used a different cerebellar tDCS montage validated in previous studies (Ferrucci et al., 2013; Weightman et al., 2023): the anode was positioned over the right cerebellum and the cathode was placed on the right deltoid muscle (upper arm). We also increased power by increasing the sample size from 20 to 30 participants per group. We replaced the original hardware-based feedback perturbation system with the MATLAB MEX-based program "Audapter" (Cai et al., 2008; Tourville et al., 2013), allowing precise, low-latency perturbation of auditory feedback. However, we found no evidence that anodal tDCS over either the left motor cortex or the right cerebellum enhanced speech adaptation. Furthermore, reanalysis of data from Lametti et al. (2018a) suggested that the previously reported positive effects were likely driven by atypical performance in the original sham group rather than by stimulation effects.

Given the failure to replicate the enhancement of tDCS over left speech motor cortex or right cerebellum on speech adaptation during altered auditory feedback, we next turned to brain imaging to directly examine brain activity associated with speech adaptation. Before scanning participants, we developed and validated a Mandarin version of the speech adaptation task previously tested in English. Seventy-six native Mandarin speakers produced consonant-vowel syllables while receiving real-time auditory feedback in which the first formant (F1) of the vowel was either increased or decreased. As seen previously in English speakers, Mandarin speakers adapted their productions in the direction opposite to the formant perturbation, confirming that the formant perturbation paradigm reliably elicits speech adaptation in Mandarin speakers.

In the brain imaging study, functional MRI was used to investigate the neural correlates of speech adaptation in Mandarin speakers under altered auditory feedback. Another 48 native Mandarin speakers produced syllables while F1 was increased in auditory feedback. Whole-brain imaging captured task-evoked brain activity during speech adaptation. As participants compensated by lowering F1 in their production of the vowel, activity increased in the left posterior cerebellum and decreased in sensorimotor cortex bilaterally. These findings are consistent with a role for the cerebellum in forward modelling and prediction error processing, and with sensorimotor cortex involvement in updating and refining feedforward motor commands.

In addition to task-evoked responses, we examined resting-state brain activity before and after speech adaptation to assess learning-related changes beyond immediate task performance. Speech adaptation produced network-level reorganisation in intrinsic functional connectivity. Connectivity increased within a right-lateralised superior temporal ("auditory") network, particularly involving the right posterior insula and parietal operculum, consistent with strengthened mechanisms for spectral error monitoring and right-hemisphere dominance for fine-grained auditory processing. Adaptation also enhanced connectivity within a cerebellar network, including coupling within the left posterior cerebellum (Crus I and Crus II) and between the left posterior cerebellum and left dorsal sensorimotor cortex. Additionally, dorsomedial sensorimotor cortex bilaterally showed increased coupling with the left cingulate cortex, suggesting reinforcement of motor planning.

Together, these findings indicate that speech adaptation to altered auditory feedback engages auditory and sensorimotor systems during task and produces

measurable after-effects in task-free intrinsic brain networks. In the following sections, I discuss the broader implications of these findings, beginning with methodological considerations related to the speech adaptation paradigm and brain stimulation and then turning to the functional contributions of auditory cortex, motor and somatosensory cortex, and cerebellum to speech adaptation.

6.2 Discussing the speech adaptation paradigm

The studies in this thesis are based on the speech adaptation paradigm using altered auditory feedback, in which a speaker's acoustic output is recorded and perturbed in near real time. Over the past decades, the speech adaptation paradigm has been implemented by many research groups for a wide range of research questions. Review of this literature and findings from studies in our own research group highlighted several critical issues that I suggest should be considered and addressed in future work.

First, speech adaptation studies often differ substantially in design (e.g. sample size, the amplitude and direction of the shift, the number of trials, and the perturbation schedule). These methodological variations make direct comparisons across studies difficult and underline the need for clearer reporting standards, as well as more systematic investigation of how specific design factors influence behavioural outcomes (Caudrelier & Rochet-Capellan, 2019).

Second, individual differences in speech adaptation are large. Although speech adaptation studies typically focus on group-level statistics, speakers often show substantially different learning trajectories: some adapt rapidly within the first

few trials and then plateau, whereas others show more gradual change across trials. Also, atypical responses are often observed, including small numbers of "followers" (speakers whose productions adapt in the same direction as the perturbation) and "non-adapters" (speakers showing little or no adaptation). Such heterogeneity has been reported repeatedly in the speech adaptation literature and complicates interpretation when analyses rely primarily on group averages or when specific response patterns are treated as outliers (Kitchen et al., 2022). Pre-registration specifying whether analyses will include or exclude these individuals would help.

Third, the relationship between perceptual acuity and adaptation needs careful measurement, both to explain individual differences and to constrain interpretations. Two directions of this relationship should be distinguished. The first concerns whether individual differences in perceptual acuity predict the magnitude of speech adaptation: speakers with finer discrimination of formant differences may detect smaller mismatches between predicted and perceived feedback and therefore adapt more strongly. The second concerns whether undergoing adaptation reshapes perception, such that sustained exposure to altered feedback reshaped speech perception. In the tDCS study (Chapter 3), we measured each speaker's speech perception using individually tailored stimuli and obtained evidence relevant to both. We did not observe a correlation between perceptual performance and the magnitude of speech adaptation, suggesting that perceptual acuity did not predict the size of the compensatory response in this sample; however, speech perception improved following the adaptation task, indicating that adaptation itself can recalibrate perception. The wider literature on the first direction is mixed:

some studies report that speakers with better perceptual acuity show larger adaptation responses during altered auditory feedback (Nault & Munhall, 2020), whereas others find weaker, null, or opposite relationships (Martin et al., 2018). Together, these results suggest that the link from acuity to adaptation magnitude is unlikely to be straightforward, whereas the recalibration of perception by adaptation may be a more reliable consequence of exposure to altered feedback.

Fourth, this thesis offered an opportunity to compare speech adaptation across two languages with different F1-F2 vowel spaces. The tDCS study (Chapter 3) tested English speakers producing "head", "bed", and "dead", whose vowel /ε/ sits between /i/ in "hid" and /æ/ in "had" along both F1 and F2 frequency scales (Figure 2.2 A in Chapter 2). The behavioural and brain imaging studies (Chapter 4) tested Mandarin speakers producing vowels /ī/, /ē/, and /ā/, which are separated primarily in F1, with /ē/ and /ā/ falling closer in F2 (Figure 2.2 B). Both English and Mandarin speakers showed robust F1 adaptation in the direction opposite to the F1 shift, supporting the cross-linguistic generalisation of speech adaptation to formant perturbation. English and Mandarin speakers diverged in F2 adaptation: English speakers showed compensatory changes in F2 alongside the F1 change, whereas Mandarin speakers showed no significant F2 adaptation. This dissociation can be understood from the geometry of the target language's vowel space. An upward F1 shift in English "head" pushes the perceived vowel toward "had". However, no English vowel differs from "head" in F1 alone: "had" is distinguished from "head" in both F1 and F2 (Figure 2.2 A). The F1 perturbation therefore creates a perceptual shift that does not map cleanly onto a single neighbouring vowel category, and

speakers can offset it either by reducing F1 alone, or by reducing F1 and concurrently changing F2 in the direction of an adjacent vowel category (MacDonald et al., 2010). In Mandarin, /ē/ and /ā/ are separated primarily along F1, with comparable F2 (Figure 2.2 B). The F1 perturbation therefore produces a perceptual shift that maps cleanly between the two vowel categories along F1 alone, and F1-only compensation fully addresses the mismatch with no pressure for an F2-based adjustment. The cross-linguistic difference therefore reflects the geometry of the target language's vowel space rather than a difference in adaptation mechanism, and supports the view that speech adaptation is a language-general process whose acoustic signature is shaped by the structure of the speaker's phonological inventory.

Finally, Mandarin is a tonal language in which the fundamental frequency (F0) contour carries lexical contrasts that do not exist in non-tonal languages (e.g. English), yet the formant-perturbation studies in this thesis manipulated only the first formant (F1) and did not manipulate F0 during production. This design choice allowed direct comparison with the English, but means the present data cannot speak to whether, or how, tone production interacts with formant adaptation in a tonal language. Because tone is carried by dynamic F0 trajectories on the vowel, and tone category could in principle constrain how far speakers can shift their frequency without threatening intelligibility, this remains an open question. Future studies could focus on pitch perturbation in which a syllable is produced, to test whether the magnitude or trajectory of pitch adaptation differs across Mandarin's lexical tone categories, or apply combined F0-and-F1 perturbations to establish whether pitch and formant adaptation interact or are happened independently.

A productive direction for future speech adaptation studies is to characterise individual differences explicitly rather than treating them as residual variance. This could include modelling participant-specific speech adaptation parameters (e.g. learning rate, asymptote, and trajectory), reporting full response distributions, and testing predictors of adaptation magnitude such as auditory perceptual acuity and baseline production variability. Also, greater standardisation of methods (particularly regarding perturbation magnitude and direction, timing and exposure duration, and technical factors such as latency of altered auditory feedback) will improve comparability across studies and help identify which aspects of the paradigm drive the observed diversity of behavioural outcomes.

6.3 Discussing failure of using tDCS to modulate speech adaptation

TDCS has been widely used to modulate motor learning and adaptation and has been proposed as a potential therapeutic method in clinical populations. However, the reliability of behavioural effects in young, healthy adults has been questioned. Studies in non-clinical samples have repeatedly concluded that single-session tDCS produces, at best, small and highly variable effects on motor and cognitive tasks, with many outcomes indistinguishable from sham at the meta-analytic level (Horvath et al., 2015). More broadly, the field faces substantial reproducibility challenges that likely stem from the combination of weak effect sizes and pervasive underpowering. Even when effects exist, typical sample sizes (about 20–30 subjects per group) may be insufficient to detect them robustly, inflating both false

positives and unstable estimates (Minarik et al., 2016).

A core methodological issue is the large number of degrees of freedom in how stimulation is delivered and analysed. The induced electric field depends not only on nominal parameters (intensity, duration, and polarity) but also on montage choices (electrode size, shape, and placement) and participant-specific anatomy. Therefore, nominally "similar" protocols can yield meaningfully different intracranial electric fields across individuals and studies, contributing to heterogeneity in individual response to stimulation (Li et al., 2015; Vergallito et al., 2022). This heterogeneity is compounded by state- and task-dependence: tDCS effects are often contingent on baseline performance, ongoing neural activity, and ceiling constraints that are especially relevant in healthy participants who may already perform at ceiling, leaving limited room for measurable behavioural improvement (Horvath et al., 2015; Vergallito et al., 2022). Therefore, inconsistent outcomes are unsurprising, even when studies adopt comparable stimulation parameters.

In the speech domain, anodal tDCS over left motor cortex failed to enhance production of tongue twisters in young adults (Wiltshire & Watkins, 2020), an effect previously described in older adults (Fiori et al., 2014; Wong et al., 2019). Similarly, the findings of our current tDCS study failed to demonstrate enhanced speech motor learning from either M1 or cerebellar stimulation, in contrast with other findings (Peng et al., 2021; Scott et al., 2020). Also, anodal tDCS over left prefrontal cortex has been reported to improve verbal fluency (Klaus & Schutter, 2018), but these benefits are not consistently reproduced, with multiple studies reporting null effects or substantial inter-individual variability despite similar protocols (Klaus & Hartwigsen, 2020; Vannorsdall et al., 2016; Westwood et al., 2017). Taken

together, these considerations motivate caution in interpreting positive findings and suggest that establishing reliable tDCS modulation of speech adaptation may require substantially larger samples, electric-field-informed intensity control, and tighter standardisation of task and stimulation parameters.

Beyond the speech domain, the broader motor-learning literature shows a similar pattern of inconsistency. Galea et al. (2011) reported that anodal cerebellar tDCS enhanced visuo-motor adaptation, reflected in a faster reduction of movement errors. In contrast, anodal tDCS over primary motor cortex did not enhance the rate of adaptation but enhanced retention of a learned movement. These findings suggested dissociable contributions of cerebellum and motor cortex to adaptation and consolidation. However, a large-scale follow-up investigation across seven visuomotor adaptation experiments, involving a total of 192 young healthy adults, failed to replicate the enhancing effects of cerebellar tDCS across a range of task parameters (Jalali et al., 2017). Similarly, Panouillères et al. (2015) found that tDCS over primary motor cortex enhanced motor adaptation, whereas stimulation of the lateral cerebellum did not modulate motor adaptation performance. A more recent study used cerebellar event-related tDCS (ER-tDCS) to successfully enhance motor learning (Weightman et al., 2022). TDCS had a short duration ($<$ 3 s) and was timed to occur simultaneously with the movement. In that study, M1-tDCS did not enhance adaptation and asynchronously-timed cerebellar stimulation also failed to do so. This finding suggested that the cerebellum might adopt a time-dependent window for tDCS enhancement and sheds light on the reactive role of cerebellum in motor control.

Based on the findings of the tDCS study presented in this thesis, several limitations arose that could be addressed in future studies.

First, we used relatively large electrodes (35 cm^2), which may have produced diffuse current spread for both motor cortex (Bikson et al., 2018) and cerebellar (Klaus & Schutter, 2021) stimulation. Cerebellar tDCS is particularly sensitive to electrode placement: small shifts in position alter the local intensity of the induced electric field (Saturnino et al., 2015), so a non-focal montage may dilute or mis-target the intended effect. Scott et al. (2020) used high-definition anodal tDCS over left ventral sensorimotor cortex and reported increased speech adaptation during altered auditory feedback. It would be important to replicate the findings of this particular study.

Second, the "blinding" in our study was not fully effective. Many participants could correctly identify whether they had received active or sham stimulation based on physical sensations they experienced, most likely because we used 2-mA direct current. In other tDCS studies using 1-mA current, participants are typically blind to stimulation (Chesters et al., 2018; Wiltshire & Watkins, 2020). Such awareness can in principle bias behavioural performance through expectancy effects: participants who believe they are receiving active stimulation may invest more effort or attention in the task and produce inflated learning estimates, while those who believe they are receiving sham may disengage. We consider this unlikely to explain our null results, because group means and trajectories of F1 adaptation did not differ across active and sham conditions, and prior work has shown that even explicit awareness of altered auditory feedback does not change compensatory responses (Munhall et al., 2009). Nevertheless, ensuring effective blinding through

more careful sham protocols remains a methodological priority for future tDCS studies at larger, detectable current amplitudes/densities.

Third, our stimulation was delivered as a single fixed-period block spanning baseline and adaptation, rather than time-locked to specific moments during learning. Given the rapid onset of stimulation effects on neuronal firing rates (Landau et al., 1964) and the time-dependent cerebellar windows for event-related tDCS noted earlier, sustained periodical stimulation may not align with the moments at which the cerebellum is most engaged in correcting individual articulatory errors. A more selective stimulation schedule, time-locked to perturbed feedback or to ongoing motor output, may therefore be required to detect effects of cerebellar tDCS on speech adaptation if they exist.

Finally, the choice of stimulation site may not have matched the network most engaged in speech adaptation to formant perturbation. As discussed in Chapter 2, previous brain imaging studies have suggested that speech adaptation to formant perturbation engages more right-lateralised frontal and temporal activity (Floegel et al., 2020), which is thought to be functionally connected to the left cerebellum. Similarly, in our own imaging study (Chapter 4), we observed brain activity corresponding to speech adaptation to formant perturbation in sensorimotor areas of both hemispheres and in the left cerebellum. Our tDCS montages, by contrast, targeted left motor cortex and right cerebellum, following the more conventional assumption of left-lateralised speech motor control and its contralateral cerebellar counterpart. This mismatch between the stimulated and functionally engaged network may explain why anodal tDCS failed to modulate speech adaptation. Future

studies may instead stimulate the right speech motor cortex and left cerebellum to test the effect of anodal tDCS on speech adaptation to formant perturbation.

6.4 Discussing the role of motor, somatosensory, and auditory cortex in speech adaptation

Our brain imaging results showed that speech adaptation to altered auditory feedback engages a distributed network of brain regions including motor, somatosensory, and auditory cortex, with changes in these regions occurring during both task performance and over a longer term (resting-state data). During the speech adaptation task, reduced activity was observed in sensorimotor cortex bilaterally, including regions corresponding to larynx, lip, and tongue representations (Eichert et al., 2021), as well as in the supplementary motor area (SMA) bilaterally. In addition, activity across a broader set of sensorimotor regions, including sensorimotor cortex bilaterally, posterior insula and parietal operculum, SMA, and right cingulate motor area, scaled with the amplitude of F1 change. These effects link changes in brain activity to speech adaptation behaviourally and indicate that altered feedback influences not only auditory processing but also the motor and somatosensory systems involved in generating and adjusting speech production.

Beyond task-evoked responses, we also observed increased resting-state functional connectivity after speech adaptation within brain regions of the dorsomedial sensorimotor cortex bilaterally, extending to increased connectivity with the left posterior cingulate cortex. The results showed that speech adaptation led to measurable changes in intrinsic brain connectivity. Together, the task and resting-state

findings showed that speech adaptation involves not only online adjustments during speaking, but also reconfiguration of sensorimotor networks.

A similar pattern emerged in auditory cortex. The imaging results point to a central role for superior temporal regions, particularly within a right-lateralised network. Increased activity and strengthened resting-state connectivity in this network suggest that speech adaptation depends on auditory regions that detect and evaluate mismatches between predicted and perceived speech outcomes. Specifically, the right superior temporal cortex is especially important when speech processing depends on spectral detail of the sound, as in formant perturbation (Floegel et al., 2020).

Taken together, these findings support the view that speech adaptation is not confined to a single locus of correction, but instead reflects coordinated changes across auditory, motor, and somatosensory systems. Auditory cortex contributes to detecting the discrepancy introduced by altered feedback, while motor and somatosensory regions participate in implementing and stabilising the compensatory response. The resting-state findings further suggest that these learning-related processes change intrinsic network organisation, reflecting the update of feedforward speech motor commands.

Within the sensorimotor system, the pattern of reduced activity during altered feedback reflects a learning-related shift in how control is achieved. One plausible interpretation is that early in adaptation, unexpected feedback elicits relatively effortful corrective responses, recruiting articulatory motor representations more strongly to counter the perceived formant shift. With continued exposure, speakers recalibrate their motor commands so that the desired acoustic output is

generated more efficiently and with less reliance on moment-to-moment correction. Therefore, reduced activity in motor and somatosensory cortex reflects more efficient implementation of an updated mapping between intended articulation and expected auditory outcome.

A limitation of this interpretation is that the reduction in sensorimotor and SMA activity observed across the approximately 40-minute adaptation task cannot be fully disentangled from a general effect of time-on-task or fatigue, since the paradigm did not include a control condition that held scan duration and speech output constant while removing the auditory perturbation itself. A decline in cortical engagement with sustained task performance, independent of any learning process, could produce a similar pattern of reduced activity across the adaptation block. This ambiguity could be addressed in future studies by interleaving "unperturbed feedback" blocks of matched duration and speech (e.g. using a different syllable) with adaptation blocks, so that any time-on-task or fatigue-related decline can be estimated independently of the adaptation-related brain activity and subtracted from it. Here we presented the change of brain activity correlated with F1 change, which would support an adaptation-related brain activity rather than brain activity purely due to time-on-task.

This interpretation is broadly consistent with previous work, although prior studies suggest that the direction of sensorimotor modulation can vary with context and stage of learning. For example, Tourville et al. (2008) reported increased activity in ventral motor and somatosensory cortex during unexpected altered relative to unaltered feedback, whereas Niziolek and Guenther (2013) found that

greater adaptation was associated with decreased activity in these regions bilaterally. Considered together with the current findings, this pattern suggests that sensorimotor responses to feedback perturbation are dynamic: activity may be heightened when errors are unexpected and correction is effortful, but reduced once the system has adapted and can produce more accurate speech outcomes with less online adjustment.

The involvement of SMA supports this interpretation. Rather than reflecting only articulator-specific control, SMA activity may reflect more general control of speech production, including initiation, sequencing, and regulation of speech movements (Riecker et al., 2005; Tourville & Guenther, 2011). Reduced SMA activity during altered feedback is therefore compatible with the idea that, once adaptation progresses, the control demands associated with generating and regulating the response become less intensive.

The post-adaptation connectivity changes were located in dorsomedial sensorimotor cortex rather than the ventrolateral regions associated with lip and tongue representations (Eichert et al., 2020), suggesting that adaptation extends beyond articulator-specific motor maps to regions coordinating and timing speech actions more generally. Consistent with this distinction, Jossinger et al. (2026) showed that activity in ventral sensorimotor cortex distinguished syllables by place of articulation, whereas dorsal sensorimotor cortex activity was more sensitive to voice onset time. The dorsomedial locus of altered connectivity reported in our study also aligns with the somato-cognitive action network identified by Gordon et al. (2023), an inter-effector system in primary motor cortex tightly coupled to cingulo-opercular control regions, insula, and cerebellum. Increased dorsomedial

connectivity after speech adaptation may therefore reflect enhanced integration of control processes for coordinating articulation and transforming sensory feedback into motor adjustments. Consistent with this view, c-TBS over auditory and somatosensory cortex reduced the retention of speech adaptation (Rao et al., 2026), highlighting the central role of sensory networks (both auditory and somatosensory) in the speech motor learning.

It is worth noting that the task and resting-state findings converge on a shared brain network but reveal a shift in cortical locus. During the speech adaptation task, activity in left posterior cerebellum, bilateral posterior insula and parietal operculum, right cingulate motor area, and bilateral ventral sensorimotor cortex at the level of the lip and tongue representations scaled with individual F1 change—an effector-specific pattern centred on the orofacial articulators. At rest, by contrast, the sensorimotor effect appeared in dorsomedial sensorimotor cortex, which coupled more strongly with cingulate cortex. This displacement from ventral, effector-specific cortex during task to dorsomedial, inter-effector cortex at rest maps closely onto the somato-cognitive action network described by Gordon et al. (2023), whose nodes occupy the dorsomedial gaps between effector representations and are functionally coupled to cingulo-opercular regions, the insular, and the cerebellum rather than to a single articulator. While the broader cerebellum–insula–cingulate system was engaged in both phases, the contribution of the sensorimotor network shifts from the effector-specific orofacial region during online adaptation to an effector-non-specific control network reflecting reorganisation once speech adaptation has been learned.

The resting-state changes also align with the motor learning literature showing that even brief training can alter intrinsic functional connectivity, often interpreted as reflecting previous experience or newly acquired skills (Albert et al., 2009; Farrens et al., 2023). Another interpretation is that increased dorsomedial sensorimotor coupling reflects a transient increase in readiness of the system to detect and respond to sensory errors after speech adaptation. However, this view needs to be tested in future studies. The joint engagement of auditory and somatosensory regions is consistent with the view that speech motor learning is guided by a multimodal target combining auditory and somatosensory goals, with somatosensory contributions only partly recovered when auditory feedback is perturbed (Guenther et al., 2006; Perkell et al., 2000). This interpretation is supported by the retention of compensatory responses when normal feedback is restored and generalisation to unperturbed speech sounds (Houde & Jordan, 1998; Jones & Munhall, 2005; Villacorta et al., 2007).

Superior temporal regions are well placed to encode the acoustic mismatch introduced by altered feedback, and the right-lateralised pattern observed here fits evidence that right superior temporal gyrus is especially engaged when speech processing depends on fine-grained spectral detail (Albouy et al., 2020; Floegel et al., 2020; Ramos Nuñez et al., 2020). Therefore, increased activity and connectivity within the superior temporal gyrus reflects enhanced detection of auditory prediction errors during speech adaptation. These error signals could then contribute to both immediate corrective responses and longer-term updating of speech motor commands.

The state-feedback control model (Houde & Nagarajan, 2011) and the FACTS model (Parrell et al., 2019b) provide useful way to interpret the current findings. In this view, auditory prediction errors drive rapid online corrections that stabilise performance in the moment, while the accumulation of such errors across trials supports longer-term updating of internal mappings and feedforward commands. The DIVA model offers a more specific explanation of how these systems interact (Guenther et al., 2006; Tourville & Guenther, 2011): predicted sensory consequences derived from efference copy are compared with incoming auditory feedback in posterior auditory regions, and the resulting mismatch is used to update feedforward commands. The model also assigns SMA a role in initiating learned motor programmes within a cortico-basal ganglia-thalamo-cortical loop and proposes an auditory-motor interface in posterior temporal cortex that transforms auditory error signals into motor adjustments. Although our data do not directly test all of these model components, reduced activity in motor and somatosensory cortex and SMA after speech adaptation indicates a shift away from effortful feedback-driven correction towards a more efficient and updated feedforward programme.

This raises the question of whether the present findings can discriminate between these accounts. I do not think they can. Because DIVA's feedback architecture is itself an implementation of state-feedback control principles, and FACTS embeds a similar state-estimation process within a Task Dynamics framework, the auditory-cortical and cerebellar changes reported in Chapters 4 and 5 (increased cerebellar activity interpreted as forward-model updating, and auditory-cortical involvement interpreted as error detection) are predicted in broadly similar form by all three models and do not, on their own, distinguish between them. The findings of this

thesis therefore support the general predictive-control architecture shared across these three models, rather than favouring DIVA specifically over SFC or FACTS.

6.5 Discussing the role of cerebellum in speech adaptation

Both task-evoked activation and resting-state connectivity changes point to a critical role for the left posterior cerebellum in speech adaptation. Increased cerebellar activity during speech adaptation, together with strengthened connectivity within cerebellar networks and with dorsal sensorimotor cortex, is consistent with views that the cerebellum implements forward models that predict sensory consequences of motor commands.

A notable feature of these cerebellar effects is their left lateralisation. Both task-evoked activity and resting-state connectivity changes were in the left posterior cerebellum, on the opposite side from the right-cerebellar tDCS tested in Chapter 3. This is informative when set alongside the resting-state finding that post-adaptation connectivity increases were in right-lateralised superior temporal cortex (Chapter 5). Right superior temporal regions are thought to be engaged in spectral processing of speech, whereas left superior temporal regions are thought to engaged in temporal processing (Albouy et al., 2020; Flinker et al., 2019; Floegel et al., 2020). Floegel et al. (2020) reported predominantly right-hemisphere activity in superior temporal sulcus and inferior frontal gyrus during formant perturbation specifically. Because corticocerebellar connectivity is predominantly contralateral, the

left cerebellum is the cerebellar partner of the right cerebral hemisphere; the convergence between right-lateralised cortical responses to spectral perturbation and left-lateralised cerebellar engagement is therefore consistent with a right-cerebral and left-cerebellar circuit supporting forward modelling during speech adaptation to formant perturbations.

Forward modelling is increasingly incorporated into computational models of speech motor control. For example, the FACTS model treats the cerebellum as a key substrate for forward prediction and state estimation, complementing cortical error-processing and control (Parrell et al., 2019b). In the DIVA model and its extension to GODIVA, cerebellar loops are proposed to contribute to predicting expected sensory outcomes and to supporting sensorimotor learning during speech acquisition and adaptation (Guenther et al., 2006; Tourville & Guenther, 2011). In an updated hierarchical state feedback control model, Hickok (2012) proposed that the cerebellum performs coordinate transformation between motor-phoneme programmes and somato-phoneme targets.

It is worth noting where these cerebellar activities fall in cerebellar anatomy. The DIVA model attributes cerebellar contributions to speech motor control in superior paravermal regions, particularly lobule VI, which lies in the "motor cerebellum" identified across motor tasks and contains a clear speech representation alongside a second representation in lobule VIIIA (Stoodley & Schmahmann, 2009; Tourville & Guenther, 2011). In the current study, both task-evoked activation and resting-state connectivity changes were instead concentrated in left Crus I, Crus II, and lobule VII, regions that fall within the "cognitive cerebellum" rather than the

"motor cerebellum", and that are functionally linked to prefrontal and parietal control networks rather than to primary motor cortex (Buckner et al., 2011; Stoodley & Schmahmann, 2009). This anatomical distinction is reinforced by the seed-based analysis in Chapter 5, in which the cerebellar network engaged by adaptation corresponded to Buckner's "control network C" rather than to a somatomotor cerebellar network. The localisation observed here is therefore more consistent with the cerebellum contributing to error monitoring, integration of auditory and somatosensory feedback, and updating of internal models than to direct articulator-level control.

In the current study, we showed that activity in the left posterior cerebellum correlated with changes in compensatory responses to shifts in auditory feedback, together with increased connectivity between the left posterior cerebellum and the left dorsal somatosensory cortex. Some researchers have proposed that the cerebellum contributes specifically to somatosensory control during speech production. Evidence for this comes from studies showing that individuals with cerebellar dysfunction produce abnormally exaggerated compensatory responses to altered auditory feedback (Houde et al., 2019; Li et al., 2019). This pattern has been interpreted as evidence that damage to cerebellar contributions to somatosensory feedback control may reduce the reliability of somatosensory information, causing the speech motor system to rely more heavily on auditory feedback (Houde et al., 2019). An alternative view is that impaired cerebellar forward modelling of somatosensory consequences itself produces exaggerated compensation, because less reliable internal predictions lead to larger sensory prediction errors when auditory feedback is altered (Parrell & Niziolek, 2021; Parrell et al., 2017). Our results are broadly consistent with both accounts, given the increased connectivity between

the posterior cerebellum and dorsal somatosensory cortex. At the same time, the association between cerebellar activity and compensation to auditory feedback shifts suggests that cerebellar involvement may not be limited to somatosensory feedback control. Instead, the cerebellum may contribute to the integration of auditory and somatosensory feedback signals, supporting the adjustment of speech motor commands when sensory feedback is perturbed.

6.6 Conclusion

In summary, the investigations conducted in this thesis contributed new knowledge by furthering our understanding of the use of tDCS in speech motor learning, by illuminating the neural correlates of this form of learning using fMRI, and by extending the formant perturbation paradigm to a tonal language. Anodal tDCS over left speech motor cortex and right cerebellum did not enhance speech adaptation under altered auditory feedback. Functional brain imaging results revealed clear contributions of auditory, sensorimotor, and cerebellar systems during speech adaptation, as well as lasting changes in the intrinsic connectivity of the regions following speech adaptation. These findings strengthen current neurocomputational models of speech motor control by demonstrating that adaptation engages distributed predictive and feedback-control circuits spanning sensorimotor cortex and the cerebellum. Speech adaptation also led to measurable plasticity within this network. Together, the results contribute to a clearer understanding of how the human brain maintains flexible and stable speech production in the face of changing sensory input.

Appendices

A

Appendix

A.1 Comparisons of F1 change to previous studies

The average amount of adaptation achieved across these studies was 36.2% (range 20.2% –56%). Our average of 36.4% falls within this range and can be considered typical, therefore. It is important to note that in addition to the variation among these studies in the size of the F1 shift applied, they differed in the number of trials used in the adaptation block and the participants studied. For example, only female participants were tested in some previous studies (Lametti et al., 2014b; Munhall et al., 2009; Shiller et al., 2009) to reduce heterogeneity in the baseline formant frequencies.

Table A.1: Comparisons of F1 change to previous studies

'–' indicates data not reported. M1, Cerebellum, and Sham indicate M1, cerebellum, and sham stimulation groups respectively.

Study	Group	N	F1 Shift Magnitude	Averaged F1 Change	SE	Range (min–max)	Approx.% Adaptation
Current Study (Chapter 3)	M1	30	**+110 Mel**	−37.7 Mel	4.97 Mel	−106 –17.1 Mel	34.3%
	Cerebellum	30		−41.3 Mel	5.20 Mel	−107 –3.48 Mel	37.5%
	Sham	30		−40.9 Mel	5.16 Mel	−98.9 –5.44 Mel	37.2%
Villacorta et al. (2007)		20	+30%, −30%	–	–	–	50%, 35%
Munhall et al. (2009)	Control only	18	+183.2 Hz	−56.6 Hz	6.0 Hz	–	31.0%
Lametti et al. (2012)	F1-shift only	14	∼−125 Hz	+70 Hz	–	–	56%
Lametti, Rochet-Capellan et al. (2014)		13	+176 Hz	−139 Hz	–	–	34.1%
		13	−60 Hz	+50 Hz	–	–	36.0%
Lametti, Krol et al. (2014)		15	+174 Hz	−59 Hz	–	–	34%
Lametti et al. (2018)		10	−49.5 Mel	+15 Mel	–	–	30.3%
		10	+49.5 Mel	−10 Mel	–	–	20.2%
Lametti et al. (2020)	Two groups from Exp. 1	20	+150 Hz	−67.12 Hz	8.92 Hz	–	44.7%
		20	−45.15 Hz	+5.78 Hz	–	–	30.1%
Shiller et al. (2024)		40	70 Mel (F1–F2 space)	22.95–24.11 Mel	–	–	32–34%

References

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology*, *51*(4), 705–723.

Abur, D., Subaciute, A., Daliri, A., Lester-Smith, R. A., Lupiani, A. A., Cilento, D., Enos, N. M., Weerathunge, H. R., Tardif, M. C., & Stepp, C. E. (2021). Feedback and Feedforward Auditory-Motor Processes for Voice and Articulation in Parkinson's Disease. *Journal of Speech, Language, and Hearing Research*, *64*(12), 4682–4694.

Albert, N. B., Robertson, E. M., & Miall, R. C. (2009). The Resting Human Brain and Motor Learning. *Current Biology*, *19*(12), 1023–1027.

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.

Amadi, U., Allman, C., Johansen-Berg, H., & Stagg, C. J. (2015). The Homeostatic Interaction Between Anodal Transcranial Direct Current Stimulation and Motor Learning in Humans is Related to GABAA Activity. *Brain Stimulation*, *8*(5), 898–905.

Amiez, C., & Petrides, M. (2014). Neuroimaging Evidence of the Anatomo-Functional Organization of the Human Cingulate Motor Areas. *Cerebral Cortex*, *24*(3), 563–578.

Andersson, J. L., Jenkinson, M., & Smith, S. (2007). *Non-linear registration, aka Spatial normalisation* (tech. rep.).

Bahners, B. H., Florin, E., Rohrhuber, J., Krause, H., Hirschmann, J., van de Vijver, R., Schnitzler, A., & Butz, M. (2020). Deep Brain Stimulation Does Not Modulate Auditory-Motor Integration of Speech in Parkinson's Disease. *Frontiers in Neurology*, *11*.

Bang, J. W., Sasaki, Y., Watanabe, T., & Rahnev, D. (2018). Feature-Specific Awake Reactivation in Human V1 after Visual Training. *The Journal of Neuroscience*, *38*(45), 9648–9657.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. *Journal of Statistical Software*, *67*(1).

Beckmann, C. F., Mackay, C. E., Filippini, N., & Smith, S. M. (2009). Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. *NeuroImage*, *47*(Suppl 1), S148.

Behroozmand, R., Johari, K., Kelley, R. M., Kapnoula, E. C., Narayanan, N. S., & Greenlee, J. D. (2019). Effect of deep brain stimulation on vocal motor control mechanisms in Parkinson's disease. *Parkinsonism & Related Disorders*, *63*, 46–53.

Belyk, M., Kraft, S. J., & Brown, S. (2014). Stuttering as a trait or state –an ALE meta-analysis of neuroimaging studies. *European Journal of Neuroscience*, *41*(2), 275–284.

Bernard, J. A. (2022). Don' t forget the little brain: A framework for incorporating the cerebellum into the understanding of cognitive aging. *Neuroscience & Biobehavioral Reviews*, *137*, 104639.

Bikson, M., Brunoni, A. R., Charvet, L. E., Clark, V. P., Cohen, L. G., Deng, Z.-D., Dmochowski, J., Edwards, D. J., Frohlich, F., Kappenman, E. S., Lim, K. O., Loo, C., Mantovani, A., McMullen, D. P., Parra, L. C., Pearson, M., Richardson, J. D., Rumsey, J. M., Sehatpour, P., … Lisanby, S. H. (2018). Rigor and reproducibility in research with transcranial electrical stimulation: An NIMH-sponsored workshop. *Brain Stimulation*, *11*(3), 465–480.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine*, *34*(4), 537–541.

Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer.

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural Representations and Mechanisms for the Performance of Simple Speech Sequences. *Journal of Cognitive Neuroscience*, *22*(7), 1504–1529.

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage*, *32*(2), 821–841.

Bongaerts, F. L. P., Schutter, D., & Klaus, J. (2022). Cerebellar tDCS does not modulate language processing performance in healthy individuals. *Neuropsychologia*, *169*, 108206.

Bourguignon, N. J., Baum, S. R., & Shiller, D. M. (2014). Lexical-perceptual integration influences sensorimotor adaptation in speech. *Frontiers in Human Neuroscience*, *8*.

Bourguignon, N. J., Baum, S. R., & Shiller, D. M. (2016). Please say what this word is —Vowel-extrinsic normalization in the sensorimotor control of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(7), 1039–1047.

Bradshaw, A. R., Lametti, D. R., & McGettigan, C. (2021). The Role of Sensory Feedback in Developmental Stuttering: A Review. *Neurobiology of Language*, *2*(2), 308–334.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–6.

Broca, P. (1861). Nouvelle observation d' aphémie produite par une lésion de la moitié postérieure des deuxième et troisième circonvolutions frontales. *Bulletins de la Société anatomique de Paris*, *36*, 398–407.

Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 105–117.

Brunoni, A. R., Amadera, J., Berbel, B., Volz, M. S., Rizzerio, B. G., & Fregni, F. (2011). A systematic review on reporting and assessment of adverse effects associated with transcranial direct current stimulation. *International Journal of Neuropsychopharmacology*, *14*(8), 1133–1145.

Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, *25*(5), 663–678.

Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C., & Yeo, B. T. T. (2011). The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(5), 2322–2345.

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*, *103*(6), 3153–3161.

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak Responses to Auditory Feedback Perturbation during Articulation in Persons Who Stutter: Evidence for Abnormal Auditory-Motor Transformation. *PLoS ONE*, *7*(7), e41830.

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 the 8th ISSP*, 65–68.

Cai, X., Ouyang, M., Yin, Y., & Zhang, Q. (2023). Sensorimotor Adaptation to Formant-Shifted Auditory Feedback Is Predicted by Language-Specific Factors in L1 and L2 Speech Production. *Language and Speech*, *67*(3), 846–869.

Caudrelier, T., & Rochet-Capellan, A. (2019). Changes in speech production in response to formant perturbations: An overview of two decades of research. In *Speech production and perception: Learning and memory* (pp. 11–75).

Chen, N., Zhao, C., Wang, M., Jones, J. A., Liu, P., Chen, X., Gong, G., & Liu, H. (2021). Linking Cortical Morphology to Interindividual Variability in Auditory Feedback Control of Vocal Production. *Cerebral Cortex*, *31*(6), 2932–2943.

Chen, Z., Liu, P., Wang, E. Q., Larson, C. R., Huang, D., & Liu, H. (2012). ERP correlates of language-specific processing of auditory pitch feedback during self-vocalization. *Brain and Language*, *121*(1), 25–34.

Chesters, J., Mottonen, R., & Watkins, K. E. (2018). Transcranial direct current stimulation over left inferior frontal cortex improves speech fluency in adults who stutter. *Brain*, *141*(4), 1161–1171.

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Human Brain Mapping*, *28*(9), 868–879.

Coemans, S., Struys, E., Vandenborre, D., Wilssens, I., Engelborghs, S., Paquier, P., Tsapkini, K., & Keulen, S. (2021). A Systematic Review of Transcranial Direct Current Stimulation in Primary Progressive Aphasia: Methodological Considerations. *Frontiers in Aging Neuroscience*, *13*.

Connally, E. L., Ward, D., Howell, P., & Watkins, K. E. (2014). Disrupted white matter in language and motor tracts in developmental stuttering. *Brain and Language*, *131*, 25–35.

Dai, G., Chen, M., Chen, X., Guo, Z., Li, T., Jones, J. A., Wu, X., Li, J., Liu, P., Liu, H., & Liu, D. (2022). A causal link between left supplementary motor area and auditory-motor control of vocal production: Evidence by continuous theta burst stimulation. *NeuroImage*, *264*, 119767.

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S.-E. (2017). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science*, *21*(2).

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006). Consistent resting-state networks

across healthy subjects. *Proceedings of the National Academy of Sciences*, *103*(37), 13848–13853.

Demirel, B., Lametti, D., Gilboa, N. A., Stagg, C. J., & Watkins, K. E. (2025). The Effects of Theta-Gamma Peak Stimulation on Sensorimotor Learning During Speech Production. *Neurobiology of Language*, *6*, NOL.a.22.

Deroche, M. L. D., Nguyen, D. L., & Gracco, V. L. (2017). Modulation of Speech Motor Learning with Transcranial Direct Current Stimulation of the Inferior Parietal Lobe. *Frontiers in Integrative Neuroscience*, *11*, 35.

Diedrichsen, J., & Zotow, E. (2015). Surface-Based Display of Volume-Averaged Cerebellar Imaging Data. *PLoS One*, *10*(7), e0133402.

Dimov, S., Katseff, S., & Johnson, K. (2012, July). Social and personality variables in compensation for altered auditory feedback. In M.-J. Solé & D. Recasens (Eds.), *The Initiation of Sound Change: Perception, production, and social factors* (pp. 185–210). John Benjamins Publishing Company.

Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., Riecker, A., & Wildgruber, D. (2002). The speaking brain: A tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. *Journal of Neurolinguistics*, *15*(1), 59–90.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, *384*(6605), 159–161.

Duffy, J. R. (1995). *Motor speech disorders: Substrates, differential diagnosis, and management* (1st ed.). Mosby.

Eichert, N., Papp, D., Mars, R. B., & Watkins, K. E. (2020). Mapping Human Laryngeal Motor Cortex during Vocalization. *Cerebral Cortex*, *30*(12), 6254–6269.

Eichert, N., Watkins, K. E., Mars, R. B., & Petrides, M. (2021). Morphological and functional variability in central and subcentral motor cortex of the human brain. *Brain Structure and Function*, *226*(1), 263–279.

Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *The Journal of the Acoustical Society of America*, *70*(1), 45–50.

Elsner, B., Kugler, J., Pohl, M., & Mehrholz, J. (2019). Transcranial direct current stimulation (tDCS) for improving aphasia in adults with aphasia after stroke. *Cochrane Database of Systematic Reviews*, *2019*(5), CD009760.

Farrens, A. J., Vahdat, S., & Sergi, F. (2023). Changes in Resting State Functional Connectivity Associated with Dynamic Adaptation of Wrist Movements. *The Journal of Neuroscience*, *43*(19), 3520–3537.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Fedorenko, E., Ivanova, A. A., & Regev, T. I. (2024). The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews Neuroscience*, *25*(5), 289–312.

Feng, Y., Gracco, V. L., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of Neurophysiology*, *106*(2), 667–679.

Feng, Y., Xiao, Y., Yan, Y., & Max, L. (2018). Adaptation in Mandarin tone production with pitch-shifted auditory feedback: Influence of tonal contrast requirements. *Language, Cognition and Neuroscience*, *33*(6), 734–749.

Ferrucci, R., Brunoni, A. R., Parazzini, M., Vergari, M., Rossi, E., Fumagalli, M., Mameli, F., Rosa, M., Giannicola, G., Zago, S., & Priori, A. (2013). Modulating Human Procedural Learning by Cerebellar Transcranial Direct Current Stimulation. *The Cerebellum*, *12*(4), 485–492.

Figueroa-Taiba, P., Alvarez-Ruf, J., Ulloa, P., Bruna-Melo, T., Espinoza-Maraboli, L., Burgos, P. I., & Mariman, J. J. (2024). Potentiation of Motor Adaptation Via Cerebellar tACS: Characterization of the Stimulation Frequency. *Cerebellum*, *23*(6), 2487–2496.

Fiori, V., Cipollari, S., Caltagirone, C., & Marangolo, P. (2014). "If two witches would watch two watches, which witch would watch which watch?" tDCS over the left frontal region modulates tongue twister repetition in healthy subjects. *Neuroscience*, *256*, 195–200.

Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, *2*(12), e1600723.

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.

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.

Folkins, J. W., & Abbs, J. H. (1975). Lip and Jaw Motor Control during Speech: Responses to Resistive Loading of the Jaw. *Journal of Speech and Hearing Research*, *18*(1), 207–220.

Fridriksson, J., den Ouden, D. B., Hillis, A. E., Hickok, G., Rorden, C., Basilakos, A., Yourganov, G., & Bonilha, L. (2018). Anatomy of aphasia revisited. *Brain*, *141*(3), 848–862.

Fridriksson, J., Hubbard, H. I., Hudspeth, S. G., Holland, A. L., Bonilha, L., Fromm, D., & Rorden, C. (2012). Speech entrainment enables patients with Broca's aphasia to produce fluent speech. *Brain*, *135*(12), 3815–3829.

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.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125.

Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2008). A Neuroimaging Study of Premotor Lateralization and Cerebellar Involvement in the Production of

Phonemes and Syllables. *Journal of Speech, Language, and Hearing Research*, *51*(5), 1183–1202.

Golfinopoulos, E., Tourville, J., & Guenther, F. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage*, *52*(3), 862–874.

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.

Goodglass, H., & Geschwind, N. (1976). Language disorders (aphasia). In *Handbook of perception* (pp. 389–428, Vol. 7). Academic Press New York.

Gordon, E. M., Chauvin, R. J., Van, A. N., Rajesh, A., Nielsen, A., Newbold, D. J., Lynch, C. J., Seider, N. A., Krimmel, S. R., Scheidter, K. M., Monk, J., Miller, R. L., Metoki, A., Montez, D. F., Zheng, A., Elbau, I., Madison, T., Nishino, T., Myers, M. J., … Dosenbach, N. U. F. (2023). A somato-cognitive action network alternates with effector regions in motor cortex. *Nature*, *617*(7960), 351–359.

Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-State Functional Connectivity Reflects Structural Connectivity in the Default Mode Network. *Cerebral Cortex*, *19*(1), 72–78.

Griffanti, L., Douaud, G., Bijsterbosch, J., Evangelisti, S., Alfaro-Almagro, F., Glasser, M. F., Duff, E. P., Fitzgibbon, S., Westphal, R., Carone, D., Beckmann, C. F., & Smith, S. M. (2017). Hand classification of fMRI ICA noise components. *NeuroImage*, *154*, 188–205.

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., Zsoldos, E., Ebmeier, K. P., Filippini, N., Mackay, C. E., Moeller, S., Xu, J., Yacoub, E., Baselli, G., Ugurbil, K., Miller, K. L., & Smith, S. M. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, *95*, 232–247.

Guenther, F. H. (2016). *Neural Control of Speech*. MIT Press.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280–301.

Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., & Tourville, J. A. (2004). Representation of Sound Categories in Auditory Cortical Maps. *Journal of Speech, Language, and Hearing Research*, *47*(1), 46–57.

Guidotti, R., Del Gratta, C., Baldassarre, A., Romani, G. L., & Corbetta, M. (2015). Visual Learning Induces Changes in Resting-State fMRI Multivariate Pattern of Information. *Journal of Neuroscience*, *35*(27), 9786–9798.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407–22.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145.

Hickok, G., Venezia, J., & Teghipco, A. (2022). Beyond Broca: Neural architecture and evolution of a dual motor speech coordination system. *Brain*, *146*(5), 1775–1790.

Horvath, J. C., Forte, J. D., & Carter, O. (2015). Quantitative Review Finds No Evidence of Cognitive Effects in Healthy Populations From Single-session Transcranial Direct Current Stimulation (tDCS). *Brain Stimulation*, *8*(3), 535–550.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213–6.

Houde, J. F., Gill, J. S., Agnew, Z., Kothare, H., Hickok, G., Parrell, B., Ivry, R. B., & Nagarajan, S. S. (2019). Abnormally increased vocal responses to pitch feedback perturbations in patients with cerebellar degeneration. *The Journal of the Acoustical Society of America*, *145*(5), EL372–EL378.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor Adaptation of Speech I. *Journal of Speech, Language, and Hearing Research*, *45*(2), 295–310.

Houde, J. F., & Nagarajan, S. S. (2011). Speech Production as State Feedback Control. *Frontiers in Human Neuroscience*, *5*, 82.

Huang, X., Fan, H., Li, J., Jones, J. A., Wang, E. Q., Chen, L., Chen, X., & Liu, H. (2019). External cueing facilitates auditory-motor integration for speech control in individuals with Parkinson's disease. *Neurobiology of Aging*, *76*, 96–105.

Ito, M. (1970). Neurophysiological aspects of the cerebellar motor control system. *International journal of neurology*, *7*(2), 162–176.

Jacobson, L., Koslowsky, M., & Lavidor, M. (2011). tDCS polarity effects in motor and cognitive domains: A meta-analytical review. *Experimental Brain Research*, *216*(1), 1–10.

Jalali, R., Miall, R. C., & Galea, J. M. (2017). No consistent effect of cerebellar transcranial direct current stimulation on visuomotor adaptation. *Journal of Neurophysiology*, *118*(2), 655–665.

Jenkinson, M. (2002). Fast, automated, *N*-dimensional phase-unwrapping algorithm. *Magnetic Resonance in Medicine*, *49*(1), 193–197.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156.

Jezzard, P., Matthews, P. M., & Smith, S. M. (2001). *Functional Magnetic Resonance Imaging*.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, *108*(3), 1246–1251.

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: Studies of Mandarin tone production. *Journal of Phonetics*, *30*(3), 303–320.

Jones, J. A., & Munhall, K. (2005). Remapping Auditory-Motor Representations in Voice Production. *Current Biology*, *15*(19), 1768–1772.

Jossinger, S., Arafat, B., & Diedrichsen, J. (2026, January). Neural representations of speech production in neocortical and cerebellar regions.

Katseff, S., Houde, J., & Johnson, K. (2012). Partial Compensation for Altered Auditory Feedback: A Tradeoff with Somatosensory Feedback? *Language and Speech*, *55*(2), 295–308.

Kelly, A. M. C., & Garavan, H. (2004). Human Functional Neuroimaging of Brain Changes Associated with Practice. *Cerebral Cortex*, *15*(8), 1089–1102.

Kent, R. D. (2004). The uniqueness of speech among motor systems. *Clinical Linguistics & Phonetics*, *18*(6-8), 495–505.

Kim, H., Wright, D. L., Rhee, J., & Kim, T. (2023a). C3 in the 10-20 system may not be the best target for the motor hand area. *Brain Research*, *1807*, 148311.

Kim, K. S., Gaines, J. L., Parrell, B., Ramanarayanan, V., Nagarajan, S. S., & Houde, J. F. (2023b). Mechanisms of sensorimotor adaptation in a hierarchical state feedback control model of speech. *PLOS Computational Biology*, *19*(7), e1011244.

Kim, K. S., Hinkley, L. B., Brent, K., Gaines, J. L., Pongos, A. L., Gupta, S., Dale, C. L., Nagarajan, S. S., & Houde, J. F. (2025). Neurophysiological Evidence of Sensory Prediction Errors Driving Speech Sensorimotor Adaptation. *Journal of Neuroscience*, *45*(27).

Kitchen, N. M., Kim, K. S., Wang, P. Z., Hermosillo, R. J., & Max, L. (2022). Individual sensorimotor adaptation characteristics are independent across orofacial speech movements and limb reaching movements. *Journal of Neurophysiology*, *128*(3), 696–710.

Klaus, J., & Hartwigsen, G. (2020). Failure to Improve Verbal Fluency with Transcranial Direct Current Stimulation. *Neuroscience*, *449*, 123–133.

Klaus, J., & Schutter, D. J. L. G. (2021). Electrode montage-dependent intracranial variability in electric fields induced by cerebellar transcranial direct current stimulation. *Scientific Reports*, *11*(1), 22183.

Klaus, J., & Schutter, D. J. (2018). Non-invasive brain stimulation to investigate language production in healthy speakers: A meta-analysis. *Brain and Cognition*, *123*, 10–22.

Kleiner, M., Brainard, D. H., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*, 1–16.

Kubota, K., Negishi, T., & Masegi, T. (1975). Topological distribution of muscle spindles in the human tongue and its significance in proprioception. *The Bulletin of Tokyo Medical and Dental University*, *22*(3), 235–242.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13).

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, *29*(1), 98–104.

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. Cengage learning.

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.

Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., & Ostry, D. J. (2014a). Plasticity in the Human Speech Motor System Drives Changes in Speech Perception. *Journal of Neuroscience*, *34*(31), 10339–10346.

Lametti, D. R., Krol, S. A., Shiller, D. M., & Ostry, D. J. (2014b). Brief Periods of Auditory Perceptual Training Can Determine the Sensory Targets of Speech Motor Learning. *Psychological Science*, *25*(7), 1325–1336.

Lametti, D. R., Oostwoud Wijdenes, L., Bonaiuto, J., Bestmann, S., & Rothwell, J. C. (2016). Cerebellar tDCS dissociates the timing of perceptual decisions from perceptual change in speech. *Journal of Neurophysiology*, *116*(5), 2023–2032.

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

Lametti, D. R., Smith, H. J., Watkins, K. E., & Shiller, D. M. (2018b). Robust Sensorimotor Learning during Variable Sentence-Level Speech. *Current Biology*, *28*(19), 3106–3113.e2.

Landau, W. M., Bishop, G. H., & Clare, M. H. (1964). Analysis of the form and distribution of evoked cortical potentials under the influence of polarizing currents. *Journal of Neurophysiology*, *27*(5), 788–813.

Lane, H., & Tranel, B. (1971). The Lombard Sign and the Role of Hearing in Speech. *Journal of Speech and Hearing Research*, *14*(4), 677–709.

Lenth, R., Buerkner, P., Giné-Vázquez, I. H., Jung, M., & Love, J. M. (2023). Emmeans: Estimated Marginal Means, aka Least-Squares Means.

Li, L. M., Uehara, K., & Hanakawa, T. (2015). The contribution of interindividual factors to variability of response in transcranial direct current stimulation studies. *Frontiers in Cellular Neuroscience*, *9*.

Li, W., Zhuang, J., Guo, Z., Jones, J. A., Xu, Z., & Liu, H. (2019). Cerebellar contribution to auditory feedback control of speech production: Evidence from patients with spinocerebellar ataxia. *Human Brain Mapping*, *40*(16), 4748–4758.

Lichtheim, L. (1885). On aphasia. In *Broca's Region* (pp. 318–347). Oxford University Press.

Lindblom, B., Lubker, J., & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics*, *7*(2), 147–161.

Liu, P., Chen, Z., Larson, C. R., Huang, D., & Liu, H. (2010). Auditory feedback control of voice fundamental frequency in school children. *The Journal of the Acoustical Society of America*, *128*(3), 1306–1312.

Lombard, É. (1911). Le signe de l'élévation de la voix. *Annales des Maladies de L'Oreille et du Larynx.*, *37*, 101–119.

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.

MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., & Munhall, K. G. (2012). Children's Development of Self-Regulation in Speech Production. *Current Biology*, *22*(2), 113–117.

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*, *129*(2), 955–965.

Macmillan, N., Hautus, M., & Creelman, C. (2021). *Detection Theory: A User's Guide*. Routledge.

Marangolo, P., Fiori, V., Caltagirone, C., Pisano, F., & Priori, A. (2018). Transcranial Cerebellar Direct Current Stimulation Enhances Verb Generation but Not Verb Naming in Poststroke Aphasia. *Journal of Cognitive Neuroscience*, *30*(2), 188–199.

Martin, C. D., Niziolek, C. A., Duñabeitia, J. A., Perez, A., Hernandez, D., Carreiras, M., & Houde, J. F. (2018). Online Adaptation to Altered Auditory Feedback Is Predicted by Auditory Acuity and Not by Domain-General Executive Control Resources. *Frontiers in Human Neuroscience*, *12*.

Medeiros, L. F., de Souza, I. C. C., Vidor, L. P., de Souza, A., Deitos, A., Volz, M. S., Fregni, F., Caumo, W., & Torres, I. L. S. (2012). Neurobiological Effects of Transcranial Direct Current Stimulation: A Review. *Frontiers in Psychiatry*, *3*, 110.

Miall, R. C., Weir, D. J., Wolpert, D. M., & Stein, J. F. (1993). Is the Cerebellum a Smith Predictor? *Journal of Motor Behavior*, *25*(3), 203–216.

Miall, R., & Wolpert, D. (1996). Forward Models for Physiological Motor Control. *Neural Networks*, *9*(8), 1265–1279.

Minarik, T., Berger, B., Althaus, L., Bader, V., Biebl, B., Brotzeller, F., Fusban, T., Hegemann, J., Jesteadt, L., Kalweit, L., Leitner, M., Linke, F., Nabielska, N.,

Reiter, T., Schmitt, D., Spraetz, A., & Sauseng, P. (2016). The Importance of Sample Size for Reproducibility of tDCS Effects. *Frontiers in Human Neuroscience*, *10*.

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*, *130*(5), 2978–2986.

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*, *133*(5), 2993–3003.

Moein, N., Mohamadi, R., Rostami, R., Nitsche, M., Zomorrodi, R., & Ostadi, A. (2022). Investigation of the effect of delayed auditory feedback and transcranial direct current stimulation (DAF-tDCS) treatment for the enhancement of speech fluency in adults who stutter: A randomized controlled trial. *Journal of Fluency Disorders*, *72*, 105907.

Mollaei, F., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2016). Sensorimotor control of vocal pitch and formant frequencies in Parkinson's disease. *Brain Research*, *1646*, 269–277.

Mollaei, F., Shiller, D. M., & Gracco, V. L. (2013). Sensorimotor adaptation of speech in Parkinson's disease. *Movement Disorders*, *28*(12), 1668–1674.

Moser, D., Fridriksson, J., Bonilha, L., Healy, E. W., Baylis, G., Baker, J. M., & Rorden, C. (2009). Neural recruitment for the production of native and novel speech sounds. *NeuroImage*, *46*(2), 549–557.

Mountcastle, V. B. (1995). The Parietal System and Some Higher Brain Functions. *Cerebral Cortex*, *5*(5), 377–390.

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*, *125*(1), 384–390.

Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*, *11*(10), 1217–1222.

Nault, D. R., & Munhall, K. G. (2020). Individual variability in auditory feedback processing: Responses to real-time formant perturbations and their relation to perceptual acuity. *The Journal of the Acoustical Society of America*, *148*(6), 3709–3721.

Neilson, M. D., & Neilson, P. D. (1987). Speech motor control and stuttering: A computational model of adaptive sensory-motor processing. *Speech Communication*, *6*(4), 325–333.

Ning, L.-H., Shih, C., & Loucks, T. M. (2014). Mandarin tone learning in L2 adults: A test of perceptual and sensorimotor contributions. *Speech Communication*, *63-64*, 55–69.

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of Physiology*, *527*(3), 633–639.

Niziolek, C. A., & Guenther, F. H. (2013). Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. *Journal of Neuroscience*, *33*(29), 12090–12098.

Oldrati, V., & Schutter, D. J. L. G. (2017). Targeting the Human Cerebellum with Transcranial Direct Current Stimulation to Modulate Behavior: A Meta-Analysis. *The Cerebellum*, *17*(2), 228–236.

Panouillères, M. T. N., Joundi, R. A., Brittain, J.-S., & Jenkinson, N. (2015). Reversing motor adaptation deficits in the ageing brain using non-invasive stimulation. *The Journal of Physiology*, *593*(16), 3645–3655.

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. *The Journal of Neuroscience*, *37*(38), 9249–9258.

Parrell, B., Bae, M., Naber, C., Kim, O. A., Niziolek, C. A., & McDougle, S. D. (2026, January). Audiomotor prediction errors drive speech adaptation even in the absence of overt movement.

Parrell, B., Lammert, A. C., Ciccarelli, G., & Quatieri, T. F. (2019a). Current models of speech motor control: A control-theoretic overview of architectures and properties. *The Journal of the Acoustical Society of America*, *145*(3), 1456–1481.

Parrell, B., & Niziolek, C. A. (2021). Increased speech contrast induced by sensorimotor adaptation to a nonuniform auditory perturbation. *Journal of Neurophysiology*, *125*(2), 638–647.

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019b). The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLOS Computational Biology*, *15*(9), e1007321.

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*, *70*(2), 453–469.

Paus, T. (2001). Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nature Reviews Neuroscience*, *2*(6), 417–424.

Paus, T., Marrett, S., Worsley, K., & Evans, A. (1996). Imaging Motor-to-Sensory Discharges in the Human Brain: An Experimental Tool for the Assessment of Functional Connectivity. *NeuroImage*, *4*(2), 78–86.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Peng, D., Lin, Q., Chang, Y., Jones, J. A., Jia, G., Chen, X., Liu, P., & Liu, H. (2021). A Causal Role of the Cerebellum in Auditory Feedback Control of Vocal Production. *Cerebellum*, *20*(4), 584–595.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., & Zandipour, M. (2000). A theory of speech motor

control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, *28*(3), 233–272.

Picard, N., & Strick, P. L. (1996). Motor Areas of the Medial Wall: A Review of Their Location and Functional Activation. *Cerebral Cortex*, *6*(3), 342–353.

Pope, P. A., & Miall, R. C. (2012). Task-specific facilitation of cognition by cathodal transcranial direct current stimulation of the cerebellum. *Brain Stimulation*, *5*(2), 84–94.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–47.

Proske, U., & Gandevia, S. C. (2012). The Proprioceptive Senses: Their Roles in Signaling Body Shape, Body Position and Movement, and Muscle Force. *Physiological Reviews*, *92*(4), 1651–1697.

Purcell, D. W., & Munhall, K. G. (2006). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119*(4), 2288–2297.

Raichle, M. E. (2010). Two views of brain function. *Trends in Cognitive Sciences*, *14*(4), 180–190.

Ramnani, N. (2006). The primate cortico-cerebellar system: Anatomy and function. *Nat Rev Neurosci*, *7*(7), 511–22.

Ramos Nuñez, A. I., Yue, Q., Pasalar, S., & Martin, R. C. (2020). The role of left vs. right superior temporal gyrus in speech perception: An fMRI-guided TMS study. *Brain and Language*, *209*, 104838.

Rao, N., Gendron, R., Manning, T. F., & Ostry, D. J. (2026). Sensory basis of speech motor learning and memory. *Proceedings of the National Academy of Sciences*, *123*(17), e2525468123.

Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., Celnik, P. A., & Krakauer, J. W. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on

consolidation. *Proceedings of the National Academy of Sciences*, *106*(5), 1590–1595.

Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., & Ackermann, H. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology*, *64*(4), 700–706.

Rochet-Capellan, A., & Ostry, D. J. (2011). Simultaneous Acquisition of Multiple Auditory–Motor Transformations in Speech. *The Journal of Neuroscience*, *31*(7), 2657–2662.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, *90*, 449–468.

Saltzman, E., & Kelso, J. A. (1987). Skilled actions: A task-dynamic approach. *Psychological Review*, *94*(1), 84–106.

Saltzman, E. L., & Munhall, K. G. (1989). A Dynamical Approach to Gestural Patterning in Speech Production. *Ecological Psychology*, *1*(4), 333–382.

Sami, S., & Miall, R. C. (2013). Graph network analysis of immediate motor-learning induced changes in resting state BOLD. *Frontiers in Human Neuroscience*, *7*, 166.

Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E., & Belin, P. (2015). Dorsal and Ventral Pathways for Prosody. *Current Biology*, *25*(23), 3079–3085.

Sato, M., & Shiller, D. M. (2018). Auditory prediction during speaking and listening. *Brain and Language*, *187*, 92–103.

Saturnino, G. B., Antunes, A., & Thielscher, A. (2015). On the importance of electrode parameters for shaping electric field patterns generated by tDCS. *NeuroImage*, *120*, 25–35.

Schmahmann, J. D., Guell, X., Stoodley, C. J., & Halko, M. A. (2019). The Theory and Neuroscience of Cerebellar Cognition. *Annual Review of Neuroscience*, *42*(1), 337–364.

Schuerman, W. L., Meyer, A. S., & McQueen, J. M. (2017). Mapping the Speech Code: Cortical Responses Linking the Perception and Production of Vowels. *Frontiers in Human Neuroscience*, *11*.

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.

Sengupta, R., & Nasir, S. M. (2015). Redistribution of neural phase coherence reflects establishment of feedforward map in speech motor adaptation. *Journal of Neurophysiology*, *113*(7), 2471–2479.

Sengupta, R., & Nasir, S. M. (2016). The predictive roles of neural oscillations in speech motor adaptability. *Journal of Neurophysiology*, *115*(5), 2519–2528.

Sengupta, R., Shah, S., Gore, K., Loucks, T., & Nasir, S. M. (2016). Anomaly in neural phase coherence accompanies reduced sensorimotor integration in adults who stutter. *Neuropsychologia*, *93*, 242–250.

Shiller, D., Lametti, D., & Ostry, D. J. (2013). Auditory plasticity and sensorimotor learning in speech production. *Proceedings of Meetings on Acoustics*, *19*(1).

Shiller, D. M., & Rochon, M.-L. (2014). Auditory-perceptual learning improves speech motor adaptation in children. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1308–1315.

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, *125*(2), 1103–1113.

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.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., & Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, *106*(31), 13040–13045.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., & Woolrich, M. W. (2011). Network modelling methods for FMRI. *NeuroImage*, *54*(2), 875–891.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, *43*(6), 482–489.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A Scale for the Measurement of the Psychological Magnitude Pitch. *The Journal of the Acoustical Society of America*, *8*(3), 185–190.

Stevens, W. D., Buckner, R. L., & Schacter, D. L. (2009). Correlated Low-Frequency BOLD Fluctuations in the Resting Human Brain Are Modulated by Recent Experience in Category-Preferential Visual Regions. *Cerebral Cortex*, *20*(8), 1997–2006.

Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *NeuroImage*, *44*(2), 489–501.

Stuart, A., Kalinowski, J., Rastatter, M. P., & Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *The Journal of the Acoustical Society of America*, *111*(5), 2237–2241.

Tahmasian, M., Eickhoff, S. B., Giehl, K., Schwartz, F., Herz, D. M., Drzezga, A., van Eimeren, T., Laird, A. R., Fox, P. T., Khazaie, H., Zarei, M., Eggers, C., & Eickhoff, C. R. (2017). Resting-state functional reorganization in Parkinson's disease: An activation likelihood estimation meta-analysis. *Cortex*, *92*, 119–138.

Tang, D.-l. (2024). Using altered auditory feedback to study pitch compensation and adaptation in tonal language speakers. *Frontiers in Human Neuroscience*, *18*, 1364803.

Tang, D.-L., McDaniel, A., & Watkins, K. E. (2021). Disruption of speech motor adaptation with repetitive transcranial magnetic stimulation of the articulatory representation in primary motor cortex. *Cortex*, *145*, 115–130.

Terband, H., Maassen, B., Guenther, F. H., & Brumberg, J. (2009). Computational Neural Modeling of Speech Motor Control in Childhood Apraxia of Speech (CAS). *Journal of Speech, Language, and Hearing Research*, *52*(6), 1595–1609.

Terband, H., van Brenk, F., & van Doornik-van der Zee, A. (2014). Auditory feedback perturbation in children with developmental speech sound disorders. *Journal of Communication Disorders*, *51*, 64–77.

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165.

Tourville, J. A., Cai, S., & Guenther, F. (2013). Exploring auditory-motor interactions in normal and disordered speech. *Proceedings of Meetings on Acoustics*, *19*(1), 060180.

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.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443.

van de Ven, V., Esposito, F., & Christoffels, I. K. (2009). Neural network of speech monitoring overlaps with overt speech production and comprehension networks: A sequential spatial and temporal ICA study. *NeuroImage*, *47*(4), 1982–1991.

van den Bunt, M. R., Groen, M. A., Ito, T., Francisco, A. A., Gracco, V. L., Pugh, K. R., & Verhoeven, L. (2017). Increased Response to Altered Auditory

Feedback in Dyslexia: A Weaker Sensorimotor Magnet Implied in the Phonological Deficit. *Journal of Speech, Language, and Hearing Research*, *60*(3), 654–667.

van den Bunt, M., Groen, M., van der Kleij, S. W., Noordenbos, M., Segers, E., Pugh, K., & Verhoeven, L. (2018). Deficient Response to Altered Auditory Feedback in Dyslexia. *Developmental Neuropsychology*, *43*(7), 622–641.

Vannorsdall, T. D., van Steenburgh, J. J., Schretlen, D. J., Jayatillake, R., Skolasky, R. L., & Gordon, B. (2016). Reproducibility of tDCS Results in a Randomized Trial. *Cognitive and Behavioral Neurology*, *29*(1), 11–17.

Vergallito, A., Feroldi, S., Pisoni, A., & Romero Lauro, L. J. (2022). Inter-Individual Variability in tDCS Effects: A Narrative Review on the Contribution of Stable, Variable, and Contextual Factors. *Brain Sciences*, *12*(5), 522.

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.

von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip. *Naturwissenschaften*, *37*(20), 464–476.

Weightman, M., Brittain, J.-S., Hall, A., Miall, R. C., & Jenkinson, N. (2022). Timing is everything: Event-related transcranial direct current stimulation improves motor adaptation. *Brain Stimulation*, *15*(3), 750–757.

Weightman, M., Lalji, N., Lin, C.-H. S., Galea, J. M., Jenkinson, N., & Miall, R. C. (2023). Short duration event related cerebellar TDCS enhances visuomotor adaptation. *Brain Stimulation*, *16*(2), 431–441.

Wernicke, C. (1969). The Symptom Complex of Aphasia. In R. S. Cohen & M. W. Wartofsky (Eds.), *Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968* (pp. 34–97). Springer Netherlands.

Westwood, S. J., Olson, A., Miall, R. C., Nappo, R., & Romani, C. (2017). Limits to tDCS effects in language: Failures to modulate word production in healthy participants with frontal or temporal tDCS. *Cortex*, *86*, 64–82.

Wiltshire, C. E., & Watkins, K. E. (2020). Failure of tDCS to modulate motor excitability and speech motor learning. *Neuropsychologia*, *146*, 107568.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338–347.

Wong, M. N., Chan, Y., Ng, M. L., & Zhu, F. F. (2019). Effects of transcranial direct current stimulation over the Broca's area on tongue twister production. *International Journal of Speech-Language Pathology*, *21*(2), 182–188.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *NeuroImage*, *14*(6), 1370–1386.

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS ONE*, *8*(7), e68910.

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*, *116*(2), 1168–1178.

Yuan, Q., Lametti, D. R., Williams, I., Zhu, H., & Watkins, K. E. (2025). Failure to replicate enhancement of speech adaptation using tDCS over motor cortex and cerebellum. *Cortex*, *192*, 152–164.

Zheng, Z. Z., Vicente-Grabovetsky, A., MacDonald, E. N., Munhall, K. G., Cusack, R., & Johnsrude, I. S. (2013). Multivoxel Patterns Reveal Functionally Differentiated Networks Underlying Auditory Feedback Processing of Speech. *The Journal of Neuroscience*, *33*(10), 4339–4348.