The cognitive and neural mechanisms involved in motor imagery of speech

Gwijde Gerfen Lidewij Maegherman
Speech Hearing and Phonetic Sciences

Thesis submitted for the degree of Doctor of Philosophy
University College London (UCL)

December 2020

Primary Supervisor: Dr. Patti Adank
Secondary Supervisor: Dr. Joseph T. Devlin

Declaration

I, Gwijde Gerfen Lidewij Maegherman, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Gwijde Gerfen Lidewij Maegherman

Abstract

Inner speech is a common phenomenon that influences motivation, problem-solving and self-awareness. Motor imagery of speech refers to the simulation of speech that gives rise to the experience of inner speech. Substantial evidence exists that several cortical areas are recruited in general motor imagery processes, including visual and speech motor imagery, but the evidence for primary motor cortex involvement is less clear. One influential model proposes that motor cortex is recruited during speech motor imagery, while another prominent model suggests motor cortex is bypassed. This thesis presents six experiments that explore the role of motor cortex in speech motor imagery. Experiments 1-3 build on established visual motor imagery tasks and expand these tasks to the speech motor imagery domain for the first time, using behavioural (experiments 1 and 2) and neuroimaging methods (experiment 3). Experiment 4 uses transcranial magnetic stimulation to explore motor cortex recruitment during a speech imagery condition, relative to a motor execution and baseline condition in hand and lip muscles. Experiments 5 and 6 use transcranial magnetic stimulation to explore speech motor imagery in tongue muscles relative to a hearing and a baseline condition. The results show that recruitment of motor cortex during speech motor imagery is modulated depending on task demands: simple speech stimuli do not recruit motor cortex, while complex speech stimuli are more likely to do so. The results have consequences specifically for models that always or never implicate motor cortex: it appears that complex stimuli require more active simulation than simple stimuli. In turn, the results suggest that complex inner speech experiences are linked to motor cortex recruitment. These findings have important ramifications for atypical populations whose inner speech experience may be impaired, such as those who experience auditory verbal hallucinations, or those with autism spectrum disorder.

Impact statement

The thesis presents an investigation into motor imagery of speech and investigates this topic through a variety of research methods, from behavioural research to neuroimaging and neurostimulatory methods. Several new tasks were developed throughout the thesis for use with these research methods, all of which may aid researchers exploring similar topics. Specifically, the novel auditory imagery task in experiments 1-3 was thoroughly documented, investigated and evaluated. Since a visual motor imagery task was used as a basis for the development of the auditory imagery task, the results extend to motor imagery processes generally and further add to this existing literature. As such, the first chapter provides a holistic view on the differences and similarities of the neural underpinnings of different types of motor imagery. The novel auditory task in experiments 5 & 6 developed a new task paradigm that can be used in future research, but also uses a tongue motor cortex stimulation method, which has rarely been used, and so adds valuable methodological guidance to the literature. A comprehensive overview of tongue motor cortex stimulation research is provided in chapter 5, aiding research in this field. In addition, the time course of motor imagery of speech was researched in detail and provides future research with useful metrics, expectations and starting points.

The experimental results shed light on a key issue in the motor imagery literature: the recruitment of motor cortex as part of a speech forward model. In doing so it informs our understanding of inner speech processes. Importantly, this has various clinical implications, specifically with regard to our understanding of the phenomenology of auditory verbal hallucinations, and the limitations that impaired inner speech may set for those with autism spectrum disorder, or those suffering from stroke-related impairments. This is especially important as inner speech is crucial for self-motivation, self-monitoring and self-awareness; those suffering from inner speech impairments may feel the consequent psychological effects of being unable to perform these self-directed cognitive behaviours.

The results also have implications from a technological standpoint. Computer-brain interaction has long been a goal not only to help those unable to communicate effectively (e.g., patients with locked-in syndrome or those recovering from illness that prevents muscle use/speech) but also those who are able to talk but see ‘silent communication’ as the logical next step in technological advancement in communication research. Several efforts, both academic and industry-level, are underway to integrate brain-based telecommunication into our daily lives, and the results in this thesis help focus the efforts on those areas that are known to be involved in motor imagery of speech, or speech simulation.

Table of Contents

List of figures

List of tables

List of abbreviations

3T	3 Tesla
A1	Primary Auditory cortex
Ag/AgCl	Silver chloride
AIC	Akaike Information Criterion
ALE	Activation Likelihood Estimation
aMT	active Motor Threshold
ANOVA	Analysis of Variance
ASD	Autism Spectrum Disorder
AUC	Area-Under-the-Curve
AVH	Auditory Verbal Hallucination
BA	Brodmann Area
BIC	Bayes Information Criterion
BOLD	Blood-Oxygen-Level-Dependent
CI	Confidence Interval
CNS	Central Nervous System
DLPFC	Dorsolateral Prefrontal Cortex
EMG	Electromyography
EPI	Echo-Planar Imaging
FDI	First Dorsal Interosseous
fMRI	functional Magnetic Resonance Imaging
FOV	Field of view
FWE	Family-Wise Error
FEF	Frontal Eye Field
HRF	Haemodynamic Response Function
IFG	Inferior Frontal Gyrus
IPL	Inferior Parietal Lobule
LME	Linear Mixed Effects
M	Mean
M1	Primary Motor cortex
MD	Mean Difference
MEG	Magnetoencephalography
MEP	Motor-Evoked Potential
MFG	Middle Frontal Gyrus
mm	millimetres
mV	millivolt
ms	millisecond
μV	microvolt
MNI	Montreal Neurological Institute
MOCA	Montreal Cognitive Assessment
MTG	Middle Temporal Gyrus

OO	Orbicularis Oris
P2P	Peak-to-peak
PET	Positron Emission Tomography
PMC	Premotor Cortex
PTA	Pure-Tone Average
rMT	resting Motor Threshold
RT	Reaction Time
s	second
S1	Primary Sensory cortex
SD	Standard Deviation
SEM	Standard Error of Mean
SMA(pre-/-proper)	Supplementary Motor Area
SMG	Supramarginal Gyrus
SPL	Sound Pressure Level/Superior Parietal Lobe
SPM	Statistical Parametric Mapping
SSBE	Standard Southern British English
STG/STS	Superior Temporal Gyrus/Sulcus
TD	Typically-developing/-ed
TE	Echo time
TMS	Transcranial Magnetic Stimulation
TPj	Temporo-Parietal junction
TR	Repetition time
UCL	University College London
UREC	UCL Research Ethics Committee
VDM	Voxel Displacement Map
VISQ	Variety of Inner Speech Questionnaire

Acknowledgements

In the first instance, I would like to thank Patti for all her guidance throughout the past four years. Thank you for the countless hours of discussion and for letting me do this PhD on my own terms. Thank you for the guidance on funding, academic politics, and stories personal and academic that put me at ease when I felt overwhelmed. Thank you for pulling me back into reality the countless times I suggested yet another study using yet another research technique – in the end I think not doing that EEG study was for the best! Thanks also to Joe for being there when I got lost in my own projects and analyses, and for helping me to take a step back when it was most needed. *Nil volentibus arduum*, as they say.

Thanks to the entire Speech on the Brain lab in its many incarnations. In loose chronological order I want to thank Dan and Helen for being the absolute best scientists to have in senior lab positions. Dan, you may not realise how invaluable those many sound boarding sessions were to me, but I can tell you none of this thesis would exist without the hours of discussion and trying out of new things in the lab. To Helen for showing me what academic professionalism looks like paired with a healthy dose of caring for lab members and students. Thanks to you both for forming the core of what I hope to achieve.

Thanks to Shego, Max, and Andrew who indulged me by letting me chatter for hours, sometimes on relevant topics but more often irrelevant ones. Again, the amount of understanding gained through talking with you is immeasurable. Thank you for letting me try any number of new paradigms on you, and for help with understanding my own designs and statistics when I needed it. To Tony and Han for helping me understand where all this research led in the final year of working through everything. Thank you, friends, for your patience. Thanks also to all PhD students with whom I have shared room 326, including but not limited to Anna, Julie, Anqi, Shiran, Katharina, Clara (and Clara), Magda, Ana, Elise, Bryony, Yue and many more – sorry for being so distracting!

Thanks to Adam, Julie and Faidra for showing me how critical it is to think about where your academic journey leads. To all the project students who put their faith in me and

decided I would be an acceptable co-supervisor. A particular thanks to Adriana for being a like-minded and driven project student.

I want to thank friends young and old for being supportive over the years and for the many hours of discussion, light-hearted or otherwise. Thank you to Mark and Harrison for sticking with me all these years: it's been a long time since BISB and Moszkva Bisztró but as the past fades, so we forge our future. Thanks to Otto and Sára for always being there when the need was greatest and for pushing me onwards – real friends are few and far between. Thank you to Karolina, Simon, Eszter, and Costanza for taking my mind off things when this whole project got a little too frightening, as was too often the case. Thank you Christelle, Vincent, Chloé, Amélie and Laura for being so supportive in the final stretch – *vous étiez vraiment formidables ces dernières années!*

I want to thank my parents, Edris and Koen, *ons mama en onze pa*, for their unwavering support through school and university years. Without you I never would have finished, let alone started – thank you for always showing your interest in my work. My love for language stems from you, thank you for passing it on. To Enide, thank you for being the sister I needed, always immensely helpful yet keeping me firmly grounded. Thanks to Moni and Marc, and Laura-Johanna, for making me feel so welcome in your homes. *Mik hoog…*

I also want to thank all my participants over the years. I have lost count but there are easily over 300 of you who were (mostly) happy to be subjected to electrodes, zaps and scans. Thank you to all the staff at Chandler house – academic, administrative and essential – for making it feel like a true second home. Thank you also to Riikka and David for taking the time to read this thesis – I hope you found some enjoyment in its contents. Thank you to Waitrose, for providing over a thousand free(-ish) coffees – I needed them.

Finally, the greatest thank you to Renate for letting me show my fears and uncertainties without having to hold back, and for being there to comfort me in times of doubt. For helping me cope with stress, anger, and impostor syndrome. For letting me talk for hours on end about what I can and cannot do, and for steadfastly believing in me. For sticking with me through lockdown, and for giving me all the hugs I needed. Thank you.

"Speech, originally, was the device whereby [mankind] learned, imperfectly, to transmit the thoughts and emotions of [its] mind. By setting up arbitrary sounds and combinations of sounds to represent certain mental nuances, [they] developed a method of communication - but one which in its clumsiness and thick-thumbed inadequacy degenerated all the delicacy of the mind into gross and guttural signalling."

Isaac Asimov, Second Foundation

1 Chapter One

1.1 Literature review

1.1.1 Introduction to the literature review

The history of inner speech is one that combines the histories of philosophy, neuroscience, and language. For centuries, philosophers such as Aristotle and Locke have pondered the relation between pure thought and inner speech (Aristotle, ca. 350 B.C.E/1928; Locke, 1690). Any discussion of inner speech is interwoven with the concepts of thought, the neurological foundation of communication with the self, and its linguistic nature which is thought to be functional rather than epiphenomenal (Fernyhough, 2017; Jeannerod, 1994; Kosslyn, 2005). The epigraph (above, p.24), taken from the novel 'Second Foundation' (Asimov, 1953, p. 10) illustrates the belief that speech is an impoverished method of conveying mental states and thoughts, limited by our personal ability to convert thoughts into a sound-stream. This view persists to some level today, underscored by the popularity of telecommunication and current research in computer-brain interaction to interface directly with speech brain areas without the need for overt articulation (Anumanchipalli et al., 2019).

An abundance of literature exists for the process of speech production and perception, and several popular models have been created on the basis of behavioural and neurological experiments (Hickok, 2012; Rauschecker & Scott, 2009). Less attention has been devoted to the topic of speech production 'in the mind', and whether such self-directed, silent speech retains the functional and cortical elements of overt speech production. This lack of attention exists despite clear evidence that shows the importance of self-directed (silent) speech not only from a developmental aspect (Sokolov, 1972; Vygotsky, 1986), but also its continued importance in self-awareness, self-motivation and self-correction (Langland-Hassan & Vicente, 2018) and its role in the life experience of atypically developed populations, such as populations with autism spectrum disorder (ASD) and auditory verbal hallucinations (AVH) (Alderson-Day & Fernyhough, 2015).

This thesis has the overarching aim to investigate the neural architecture that leads to motor imagery of speech, and more specifically to investigate the role of motor cortex

(M1). While M1 has an evident role in overt motor production, its inclusion in models for motor imagery of speech is still debated (Lœvenbruck et al., 2018; Tian & Poeppel, 2012).

The literature review begins with a description of speech production with specific regard for the motor control models that allow for fluent speech and rapid error correction (1.2). A discussion of motor imagery of speech follows, detailing its phenomenological properties and its underlying processes (as well as those processes it likely does not involve) (1.3). I then consider how motor imagery of speech can make use of the same motor control pathways as speech production and posit the possible neurophysiological correlates involved. (1.4). Finally, a brief discussion shows why this research is timely and important (1.5).

1.2 Speech production

As stated by Hickok (2012), speech production has primarily been studied along two paths which seldom converge: motor control and psycholinguistics. This review will focus on motor production of speech on the assumption that there exists some lemma-level representation which the speaker wishes to communicate (see Indefrey & Levelt, 2000; Levelt et al., 1999). The motor production model described here is an amalgam of various models, such as the influential Directions Into Velocities of Articulators model (Guenther et al., 2006; Tourville & Guenther, 2011) as well as others, e.g., the Hierarchical State Feedback Control model, see Hickok, (2012), Lœvenbruck et al., (2018), Tian & Poeppel, (2013). While many of the model's components and their uses are widely agreed upon, others are more contested. The discussion here represents a fundamentally modern view on speech production.

1.2.1 A neurocognitive model of speech production

Speech production is comprised of two parallel pathways: a motor pathway and a somatosensory pathway. These pathways employ a network of cortical areas to precisely plan speech production. The execution of these speech production plans culminates in an audible pressure wave produced by a complex coordination of muscles. Several cortical areas play an important role in early stages of speech production, such as left posterior superior temporal gyrus and sulcus (STG/STS), left posterior middle temporal gyrus (MTG), which aid in lexical retrieval and the auditory specification of a desired auditory state. This auditory specification is sent to the temporo-parietal junction (TPj) where the desired auditory state is transformed into auditory goals, which are themselves sent to left inferior frontal gyrus (IFG) and ventral premotor cortex (PMC) where goal-directed motor programmes are specified for articulatory planning. This specification process has also implicated the pre-supplementary motor area (pre-SMA) in higher-level response selection, and SMA-proper as well as dorsolateral prefrontal cortex (DLPFC) in controlling motor output. These programmes are then forwarded to primary motor cortex (M1) where they are integrated for execution (P. Tremblay et al., 2015; P. Tremblay & Gracco, 2009).

In a parallel processing stream, a similar process occurs for the somatosensory pathway: following lexical retrieval in the MTG, somatosensory specification is performed in the anterior supramarginal gyrus (SMG) and primary somatosensory cortex (S1). Similar to the auditory specification of the desired auditory state in the motor stream, the somatosensory specification of the desired somatosensory state is sent to the cerebellum for transformation from desired state into somatosensory goals. From the cerebellum, these somatosensory goals are sent to M1, where they are integrated with the auditory motor programme for execution by way of motor commands. Already at the level of M1, articulator-specific activation can be distinguished through the somatotopic organisation of M1 areas (Penfield & Rasmussen, 1950).

Throughout and after the cortical operations, the motor commands' motor signals make their way to subcortical structures such as the thalamus and basal ganglia (through the cortico-striatal-thalamic loop) and cerebellum (through the cortico-cerebellar-thalamic loop). Past these structures the signals are conveyed through the pons and medulla oblongata, after which the cranial nerves carry the signals towards the muscle of the face and tongue, larynx, and the spinal cord itself carrying relevant signals to the respiratory muscles (P. Tremblay et al., 2015).

The combination of the auditory-motor and somatosensory streams provides a solution to the problem facing any motor action process, namely the integration of near-immediate feedback and error correction during said motor action (here speech). For a better understanding of these issues, it is necessary to first consider general motor architecture models, and to relate these models to speech production.

1.2.2 Why speech production requires a feedforward motor architecture

It is evident from the literature that auditory feedback plays a crucial role in error correction, as evidenced in delayed feedback (Lee, 1950; Takaso et al., 2010), feedback in noise (Lombard, 1911) and amplified feedback studies (Chang-Yit et al., 1975). It is equally evident that somatosensory feedback from the articulators is also extremely important in speech production (Abbs et al., 1984; Kelso et al., 1984). More importantly, somatosensory feedback is crucial and distinct from auditory feedback, as shown through the intelligible speech of deaf speakers and feedback perturbation studies (S. Tremblay et al., 2003). Although these two types of feedback provide important contributions to fluent speech production, they are also insufficient for this task. Complications such as neuronal delay, transformation speed and feedback noise (Houde & Nagarajan, 2011; Nozari et al., 2011) reveal that these two types of overt feedback cannot by themselves form the entire speech production feedback architecture. For instance, these types of feedback cannot explain why so few errors are made in fluent speech production, as such feedback is inherently too slow and imprecise. As a result of these observed limitations, it has been suggested that an internal monitor is present which can make use of speech production plans to predict where errors may be made before they are physically generated and fix them. Early suggestions of what form this monitor might take were not necessarily based on motor control but were in some way based on the production of speech (Levelt, 1983). Regardless of its form, this internal monitor must be highly detailed and work at speed (Haggard et al., 1999).

1.2.3 Internal forward model

In their 1970 paper, Conant and Ashby suggest that in order for a regulator (such as the brain) to be maximally efficient and successful, it must be isomorphically related to the system it represents. That is, if the brain as a regulator of any process – e.g., motor execution – is to be useful for its purpose, it must build a model of the system that is to be regulated, and this model must have a one-to-one relation to the system it attempts

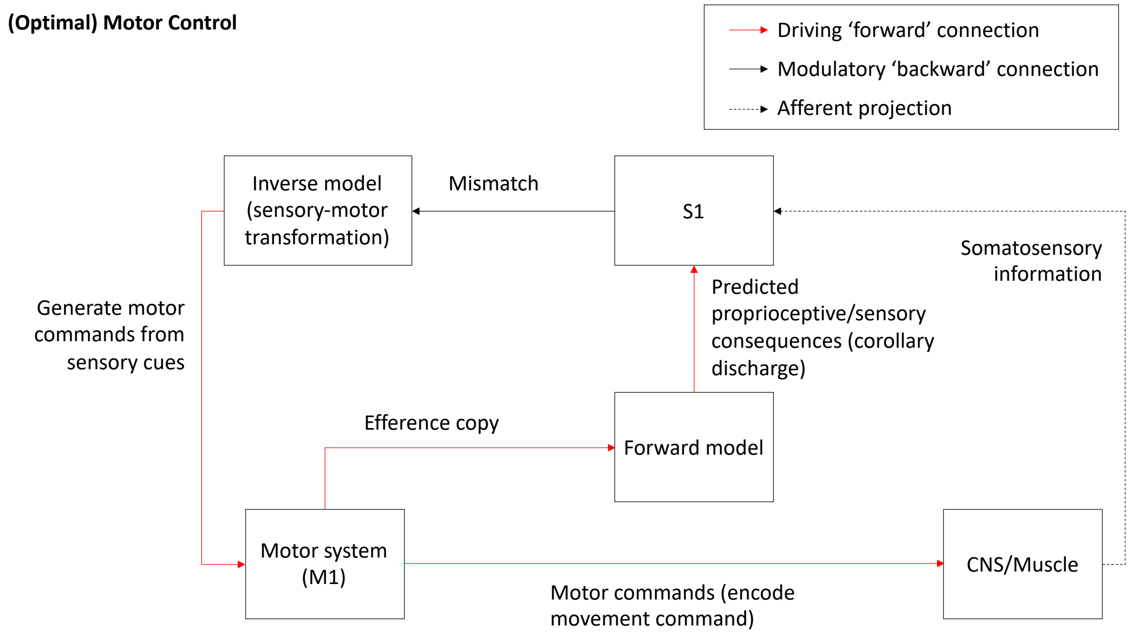


Figure 1.1: Optimal motor control schema (adapted from Adams, Shipp and Friston, 2012). See also Figure 1.4. for the extended version proposed by Tian and Poeppel (2012).

to regulate. In motor control, an example of such a model would be a model for the central nervous system as proposed by Miall and Wolpert (1996). In their framework, the central nervous system internally simulates the motor system, mimicking natural behavioural processes. This internal model is capable of using the present state of the motor system to predict the next state of the motor system on the basis of a motor command. Such a system would be invaluable in motor planning, general motor control and motor learning (Frith, 1992; Miall & Wolpert, 1996; Wolpert et al., 1995; Wolpert & Kawato, 1998), and constitutes the internal monitor previously referred to.

To avoid terminological confusion, I will briefly describe how different types of models interrelate. The definitions used here are heavily based on those found in Kawato (1999). An internal model is a neural mechanism that can mimic the input/output characteristics, or their inverses, of the motor apparatus while a forward (internal) model uses efference copies of motor commands to predict sensory consequences (also referred to as corollary discharge). An inverse (internal) model can calculate necessary feedforward motor commands (usually conceptualised as corrective) from desired trajectory information (see Figure 1.1 for an example).

When combined, these models form a powerful apparatus capable of predicting consequences from motor actions and calculating required changes based on these predictions. As initially proposed by Miall and Wolpert (1996) the good regulator is instantiated as a forward model (see Figure 1.1): when motor commands are sent to the motor system and the sensory system, the motor commands are also sent to a forward model via an efference copy containing these motor commands, which is processed in parallel (von Holst & Mittelstaedt, 1950). This forward model can then be used to estimate the expected sensory feedback as a result of the motor commands sent to the muscles, known as corollary discharge (Li et al., 2020; Sperry, 1950; Wolpert & Kawato, 1998). Should the estimated feedback differ from the sensory feedback, a sensory discrepancy is found and a mismatch (error) signal can be passed on to the inverse model, which transforms the sensory discrepancies into corrective motor commands. The efferent motor commands can then be modified to instruct the muscles to perform corrective actions. Given that this manner of overt motor control is posited to be active for each type of motor action, it is theoretically attractive to assume speech production is included. It is also parsimonious to hypothesise that the inner workings of such a system might contribute to similar processes, specifically with regard to the imagining of speech actions (i.e., the synthesis of covert speech actions experienced as inner speech). A more in-depth discussion of such a system is presented in section 1.2.4.

1.2.4 Predictive coding, or: conventional motor versus active inference

In speech as in general motor control, different approaches exist towards the feedforward and feedback system, and how changes are rapidly effected throughout this

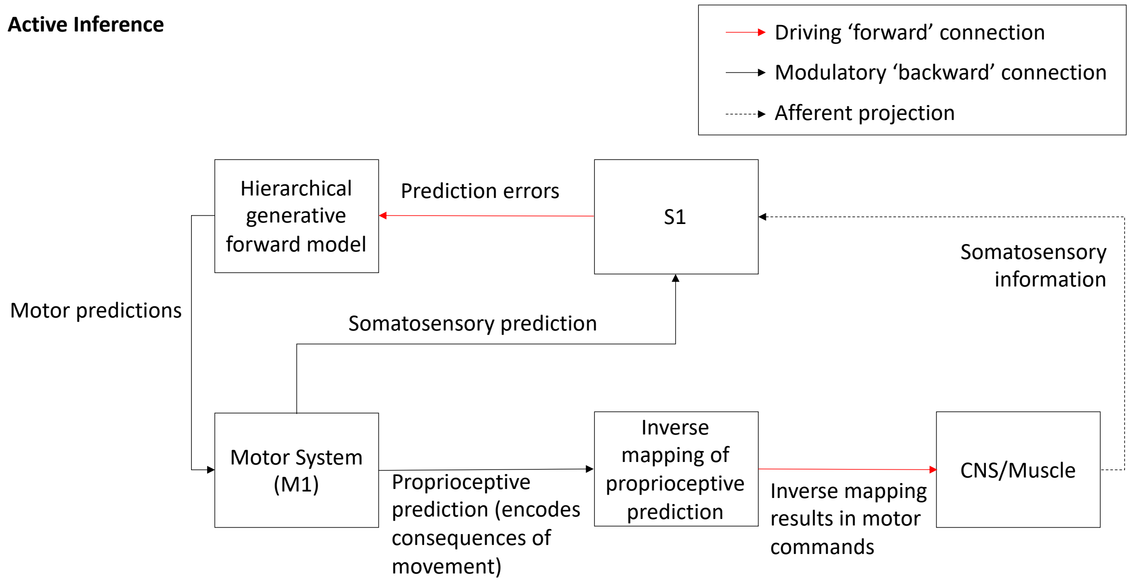


Figure 1.2: Active inference model (adapted from Adams, Shipp and Friston, 2012)

system. The conventional motor control model has already been described above and can be seen in Figure 1.1. An alternative model following the theory of predictive coding (Adams et al., 2012; Friston, 2003) posits that not motor commands (or efference copies) but proprioceptive and somatosensory predictions are sent through the network (Figure 1.2). While the motor control model relies on state estimates and state changes to modify efferent motor commands, the predictive control system uses a hierarchical generative forward model taking proprioceptive and somatosensory prediction errors as input, and outputs proprioceptive predictions that are improvements on a previous prediction (since prediction errors have been taken into account). Instead of an inverse model comparing the estimated and actual state of the motor system and sending corrective motor commands, the generative forward model uses error signals from lower levels to improve the predictions at the higher level, which in turn minimises the error in the system, and in this way 'predicts' the causes of sensory input (Rao & Ballard, 1999). In effect, predictive coding instantiates a percept on the basis of repeated top-down predictions and bottom-up prediction errors. Both models describe speech production systems effectively, and it is not the aim of this thesis to differentiate between the two approaches. In the remainder of the thesis the optimal motor control model (Figure 1.1) will be followed for simplicity.

The next section focuses on what is meant by motor imagery of speech, and how the discussion on forward models in speech production is relevant to motor imagery of speech.

1.3 Motor imagery of speech

1.3.1 Motor imagery of speech as thought

Due to the long-standing interest in the topic of inner speech (see Langland-Hassan & Vicente, 2018), the nomenclature of various phenomena and processes requires some clarification. The work presented here falls under the scope of, at its most specific, the processes behind motor imagery of speech – itself a subset of a number of variously defined processes.

For those who regularly experience inner speech – as much as 82.5% of the population (McCarthy-Jones & Fernyhough, 2011) – the term 'thinking' is synonymous with inner speech (Alderson-Day & Fernyhough, 2015) to such a degree that inner speech is considered a *sine qua non* for consciousness (Baars, 2003). However, a crucial question concerns the limits of this inner speech, both in descriptive and research terms. As noted by Alderson-Day and Fernyhough (2015), an effort to view all of the literature concerning inner speech required the use of no less than 14 different search phrases: *inner speech, private speech, self-talk, covert speech, silent speech, verbal thinking, verbal mediation, inner monologue, inner dialogue, inner voice, voice imagery, speech imagery, and auditory verbal imagery,* and the phrase that aligns most closely with the work presented here, *articulatory imagery* (p. 932). In this section, I will define the specific level of cognitive processing this work focuses on and delineate how this set of processes differs from other, potentially confounding sets of processes. Though the focus is heavily on speech, the initial stages are equally relevant to more general motor actions that have consequences in any modality, be it auditory, visual, or other.

The first distinction relates to the phenomenology of speech of external (world) origin, and speech of internal (mind/brain) origin. Speech that is externally generated (i.e., is audible), both perceived and produced, is fundamental to our personal cognitive development and our relation to the world around us, as shown through the distinct viewpoints of Vygotsky and Piaget. While Piaget believed that egocentric speech precedes social speech, following the development of logical verbal thinking, Vygotsky instead proposes that social speech precedes egocentric speech, which itself precedes inner speech (Vygotsky, 1934/1986). Whichever schema one chooses to follow, I will argue that the concept of cognition is closely related to that of inner speech.

The internal/external distinction follows Vogt et al., (2013), in which the authors propose that action observation and motor imagery involve a range of similar sensory and motor representational processes. The objects, actions or action consequences perceived (percepts, see also Marr, 1985) or produced (productions) range from externally generated, i.e., in the world, to internally generated, i.e., in the mind. Regarding internally generated percepts, it is important to clarify that not all thinking is linguistic in nature. Thought may have a representation in a specific modality, but it need not.

Conscious thought, however, can be said to have some representational, perceptual aspect, certainly where speech is concerned (Baars, 2003; Martínez-Manrique & Vicente, 2015; McCarthy-Jones & Fernyhough, 2011), though it need not be a fully-formed auditory segment. As there may be interplay between linguistically formulated thought, visual imagery, and perhaps even gustatory or olfactory percepts, this subset of processes is referred to as abstract thought (see Figure 1.3).

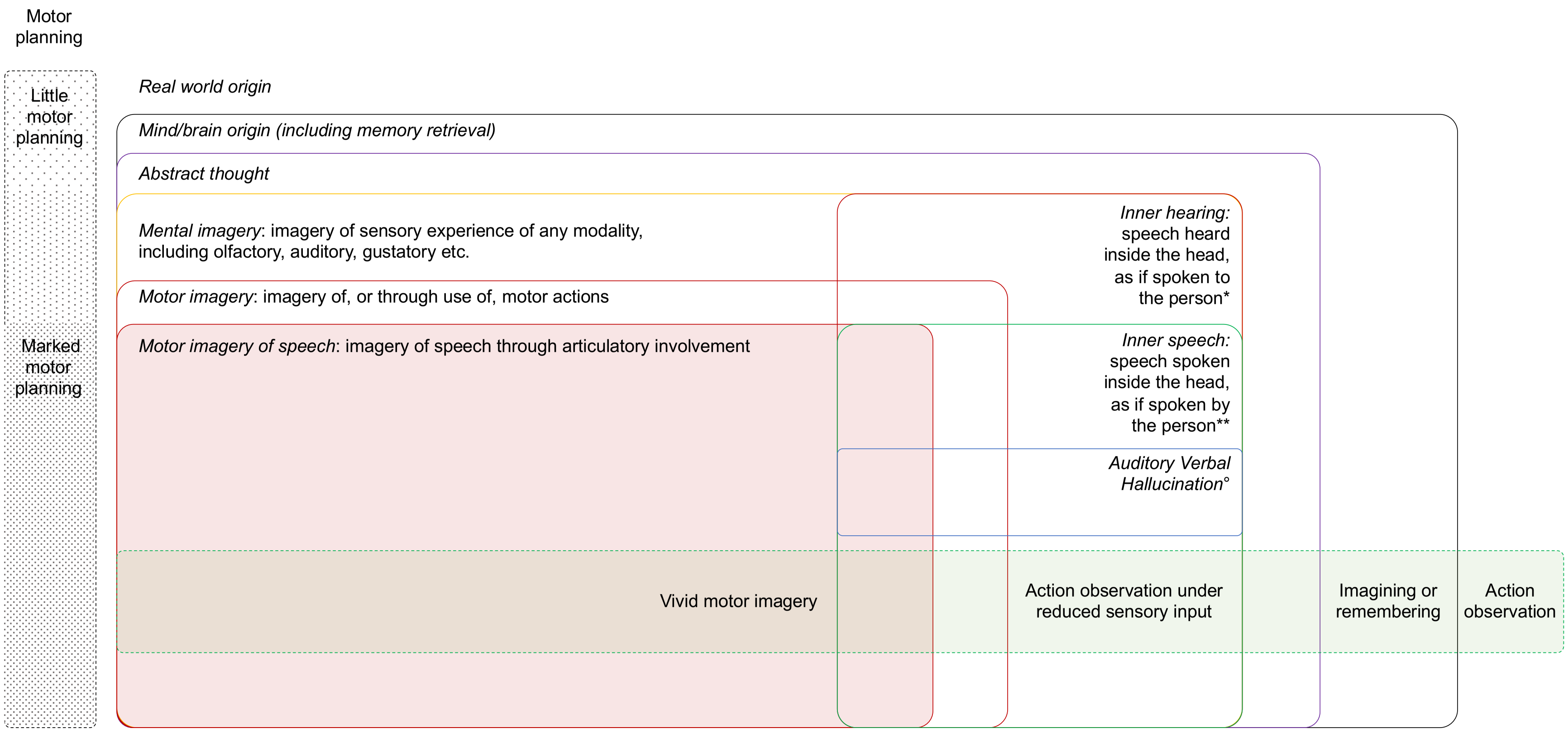


Figure 1.3: A diagrammatic exploration of the sets of processes involved in motor imagery of speech.
Red shading indicates the specific processes addressed in this work. Green shading follows a continuum proposed by Vogt et al. from action observation to vivid motor imagery. Motor planning shading indicates the level of (articulatory) motor planning at play at each level of processing. * By not integrating inner hearing with inner speech necessarily, I allow for inner hearing to be the result of memory retrieval ONLY, requiring no level of motor imagery of speech. However, inner speech is a subset of inner hearing as inner speech is always 'heard' – one cannot perform inner speech without also hearing the 'output'. **Auditory speech can be 'perceived' without the person being aware that they are performing inner speech, but inner speech is always accompanied by inner hearing. ° This model allows only for the view that AVH necessarily recruits both inner speech and inner hearing mechanisms (on the basis that the inner speech awareness is disrupted). It does not allow for AVH to be the result of simple memory retrieval, but also does not insist on precise motor imagery, allowing for impoverished levels of articulation to be modified by process other than precise motor imagery of speech (e.g., semantic / syntactic priming) and produce an 'articulated' AVH as a result.

Within the region of abstract thought those elements that make up volitional imagery can be sequestered: conscious, wilful mental imagery, defined by Kosslyn, Ganis, et al., (2001; p. 635) as occurring "[…]when perceptual information is accessed from memory, giving rise to the experience of 'seeing with the mind's eye', 'hearing with the mind's ear' and so on. […] Mental images need not result only from the recall of previous perceived objects or events; they can also be created by combining and modifying stored perceptual information in novel ways". Key in this definition is the sense of agency of the experiencer: the experiencer is also the wilful creator of the mental images (referred to as mental percepts rather than 'images' so as not to limit the definition to the visual modality, as in Farah, 1989). Tian et al., (2016) define mental imagery as "an internally generated quasi-perceptual experience" (p. 1), again highlighting the agency of the generator. In this sense one can, for example, generate a visual percept of a landscape vista based on the reactivation of memories related to the topic at hand: my visual percept may include a wood cabin by a lake whereas yours may not. At this level, the concepts of inner hearing and inner speech appear: these processes give rise to the mental perception of hearing speech, the 'hearing with the mind's ear'. The two processes are shown as distinct (following Hurlburt et al., 2013) however, since inner speech results in inner hearing, but inner hearing is not necessarily precipitated by inner speech – inner hearing can also be the result of memory retrieval. Both processes occur within mental imagery.

Motor imagery is the specific subset of processes of mental imagery that involves the imagined use of effectors (muscles) in a similar way to motor preparation and motor planning processes (Jeannerod, 1994). Jeannerod proposed a functional equivalence between motor imagery and motor preparation on the basis of the effects that motor imagery has on motor learning, the physiological relation between motor imagery and motor preparation, and the similarity of neural structures involved in both processes. For instance, generating a vista as described above may involve mental imagery processes, but it would not necessarily involve the use of motor imagery; visual imagery (via memory retrieval) would suffice. If the imagery included body motion, e.g., a person chopping wood, motor imagery could be involved. Jeannerod (1994) proposed that such imagery would additionally be constrained by the same kinetic and temporal motor rules as motor planning for overt action. Parsons (1994) discussed this as 'mental simulation of one's action', and Alderson-Day and Fernyhough (2015) discussed motor simulation as part of 'embodied simulation', referring to accounts by Bergen (2012) and by extension, Barsalou (2008) in which the authors propose that cognition does not arise through amodal symbol systems in the brain (Pylyshyn, 1984), but rather that "modal simulations, bodily states and situation action underlie cognition" (Barsalou, 2008, p. 617).

It follows from the above that the processes of motor imagery of speech are themselves a subset of processes of motor imagery, specifically those processes that involve the speech motor architecture. The definition of motor imagery of speech maintained throughout this thesis is *motor imagery processes that involve articulatory motor processing, inclusive of* (due to issues of nomenclature throughout the past few decades) *most theories of motor imagery, mental imagery of movement and simulation of movement.*

Defining motor imagery of speech as above does not invalidate questions with regard to how different levels of inner speech and even motor imagery of speech may be accomplished. As noted by Fernyhough (2004) and more recently Grandchamp et al., (2019), there is good reason to suppose that even using the above definition of motor imagery of speech, there are differences between expanded (i.e., fully phonologically-featured) and condensed (i.e., less-than-fully phonologically-featured) inner speech. Expanded inner speech can feature 'articulated' sentences complete with pragmatic features such as sarcastic tone, as opposed to condensed inner speech which may be little more than a barely 'articulated' word, as stated by Oppenheim and Dell (2010). This is in line with the Vygotskian thesis which proposes social, dialogic speech occurring first, the subsequent transformation towards private, monologic speech, and finally the internalisation and abbreviation of speech (Alderson-Day & Fernyhough, 2015; Geva et al., 2011). Additionally, the concept of 'verbal mind wandering' should be dissociated from motor imagery of speech. There are mental imagery processes that give rise to 'fully-articulated' inner speech that, while volitional, lack the same level of agency as that of verbally recounting a list in one's head with the specific purpose of remembering it. Such a process is more likely to use the 'phonological loop' as proposed by Baddeley and Hitch (1974) in the context of working memory. To clarify, the term 'motor imagery of speech' as used in this work should be considered to represent volitional, agentive, and expanded articulatory motor processes.

1.3.2 Motor imagery of speech as motor control and simulation

Moulton and Kosslyn (2009) provide an expanded theory of Jeannerod (1994, 2001) related to inner speech by considering the concept of an 'emulative simulation' or emulation. In contrast to an 'instrumental simulation', an emulation imitates not just the stages of the event being simulated but also the algorithmic processing that occurs during the transformation of successive states. In their view, an emulation of a dialogue involves predicting dialogue. The authors claim that such emulations can be thought of as 'second-order' simulations – imitating not just the content of the dialogue, but also the processes that change the content. As the authors state, imagery and simulation are similar in several ways: (a) they both constitute epistemic devices, i.e., they make knowledge available or generate knowledge as output; (b) simulation and imagery can be considered to be functionally similar, and are inherently predictive; (c) simulation relies on perceptual representations, and while the authors focus on visual imagery and simulation, this can be discussed in terms of speech imagery and simulation as well.

The definition as given in 1.3.1 above states that motor imagery of speech in some form involves the use of articulatory motor processing. What is not stated explicitly, is that this articulatory motor processing is the same as that expected in motor execution of speech (i.e., performing such movements as required for overt speech). This ambiguity is intentional. The level of phonetic and articulatory specification is a matter of contention, intra- and inter-personally: one person may consistently specify inner speech articulation more than another person, or a person may vary how specified their inner utterances are depending on the situation or need (Langland-Hassan & Vicente, 2018; Oppenheim & Dell, 2008). One of the most important questions regarding motor imagery is to which degree its processes are the same as processes motor execution (Decety, 1996; Hardwick et al., 2018; Jeannerod, 2001). This question also remains for motor imagery of speech specifically (Dell & Oppenheim, 2015; Perrone-Bertolotti et al., 2014). Of specific interest is the potential role of M1 in motor imagery of speech.

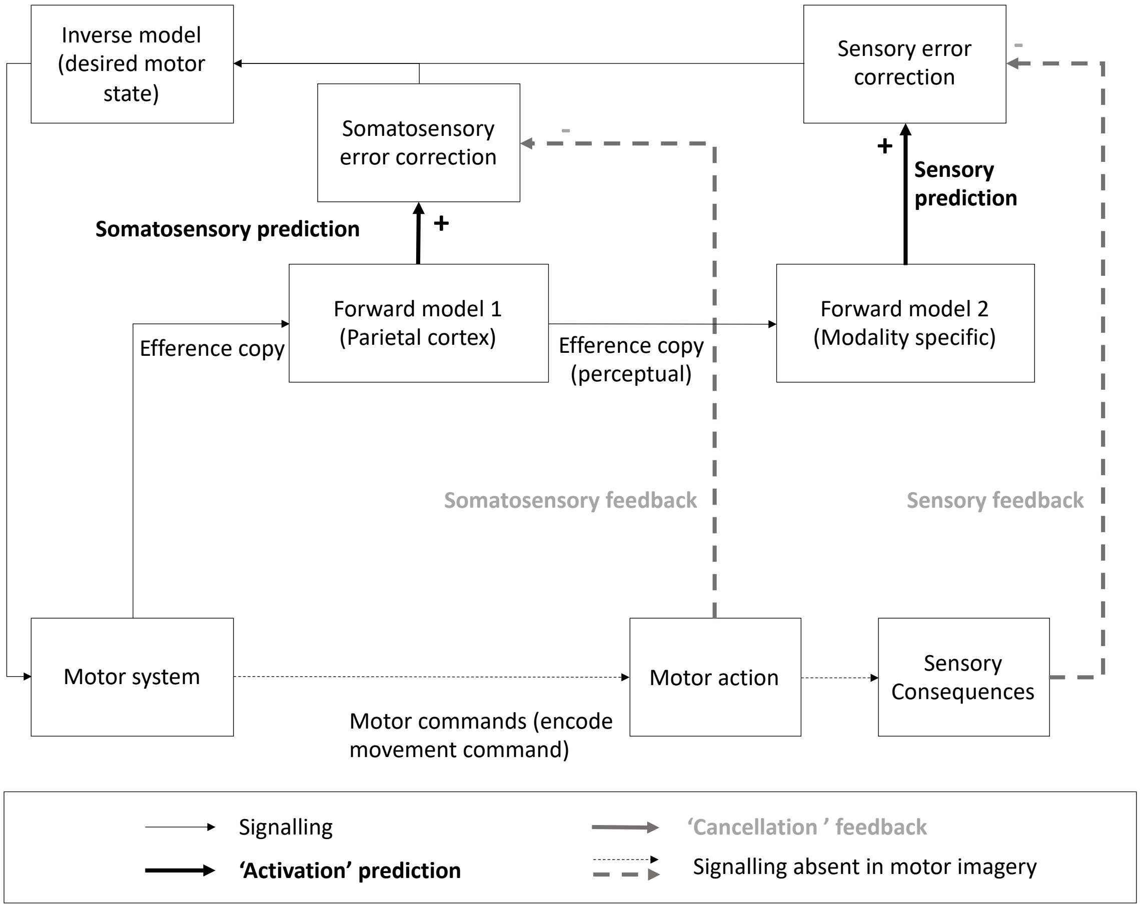


Figure 1.4: Model of speech processing and mental imagery of speech adapted from Tian and Poeppel (2012), itself an expanded version of the motor control account in Figure 1.1.

As discussed in section 1.2.3, the existence of an internal forward model during speech production is supported by literature from a number of different vantage points in speech production. Since this model exists for speech production, these schemas have also been used to model the interactions of systems and modules that are thought to make up motor simulation in motor imagery of speech. One theory is presented by Tian and Poeppel (2012), whose model for imagery of speech is shown in Figure 1.4. Here, imagery of speech contains many of the same components and processes as overt speech production. The motor system sends motor signals to both the CNS and articulators, as well as an efference copy to the first forward model in the parietal cortex, which creates a somatosensory prediction on the basis of this efference copy. The motor signals to the articulators result in action, which itself has sensory consequences – similarly, the simulation-estimation in the first forward model results in a perceptual efference copy interpreted by the modality-specific (here auditory) second forward model, which computes its own sensory prediction. In overt speech, the somatosensory feedback from motor action cancels out those elements of the somatosensory prediction that are the same; that is, if the prediction and feedback match, all signals are cancelled out and no errors are found. Similarly, the sensory feedback is compared to the sensory prediction and if all signals cancel out no error correction is required or carried out. If there is a mismatch at either correction node, corrections are sent to the inverse model which represents the desired motor state, which is updated and the appropriate signalling to the motor system effected. The Tian and Poeppel model suggests that speech imagery is the result of the efference copy pathway described in the absence of the motor execution pathway. Efference copies are sent to forward models 1 and 2 sequentially, so that the prediction of somatosensory and sensory nature still takes place, but with no feedback cancellation phase. The prediction itself constitutes the inner speech percept. In a follow-up paper (Tian & Poeppel, 2013) the authors clarify that if covert execution is the goal, M1 is bypassed and a simulation of motor plans is instead performed in parietal cortex. While M1's role in overt speech production is undisputed, in this model M1 has no role in motor imagery of speech.

A potential issue for the Tian and Poeppel model exists in the use of motor efference copies as the driving force behind inner speech. What constitutes an action in motor imagery is unclear: in the above description one could argue that overt speech is the action, and inner speech the by-product of the processes keeping overt speech error-free and motor control smooth and rapid. However, there is good reason to suppose that (wilful and volitional) inner speech itself constitutes an action, in which case it may itself need an efference copy: this is also supported by the finding that inner speech exhibits certain types of error correction, but not others (Dell & Oppenheim, 2015; Oppenheim & Dell, 2010), and that inner speech exhibits auditory attenuation in similar ways to overt speech (Agnew et al., 2013; Scott, 2013). These findings suggest that inner speech shows evidence of corollary discharge, and so that a monitor of inner speech exists at some level. Additionally the account by Tian and Poeppel (2012) suggests that sense of agency requires a separate 'higher function' source monitor (ibid, p. 7).

A solution to these action-constituting and agency questions can be found in a proposal from the literature on auditory verbal hallucination (AVH). The AVH literature (see Jones & Fernyhough, 2007 and Seal et al., 2004) suggests that inner speech itself constitutes a type of action, which again implies the use of traditional motor control

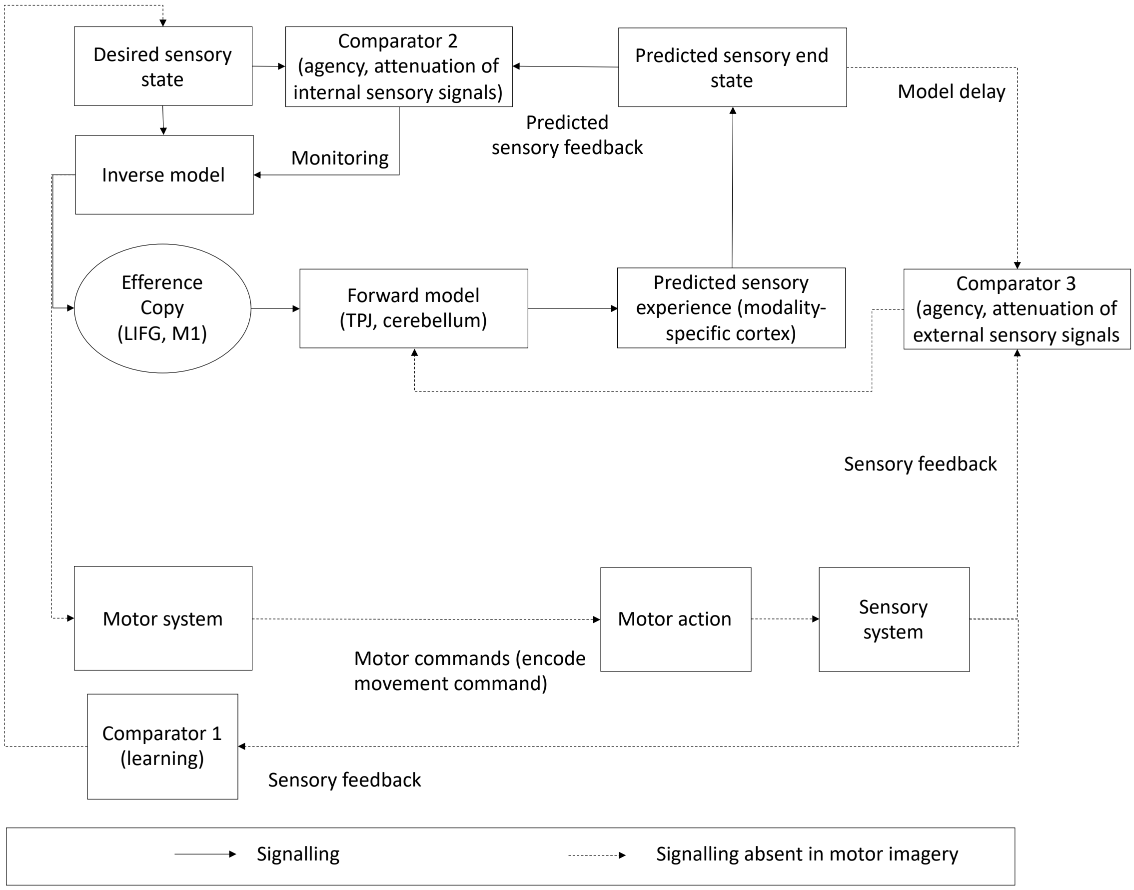


Figure 1.5: Model of the predictive control account of inner speech, adapted from Lœvenbruck et al, 2018. This model is itself an expansion of the model in Tian & Poeppel (2012), which can be seen in Figure 1.3, as well as an expansion on Jones & Fernyhough (2007). Note that this adaptation is simplified: for full details see Lœvenbruck et al., 2018.

theory. Here, the inner speech is not merely the result of the efference copy, but the product of its own efference signal, meaning that inner speech itself has an efference copy. Where the efference signal in Tian and Poeppel (2012) produces overt speech and the efference copy produces a simulation of said overt speech, the efference signal in Jones and Fernyhough (2007) produces inner speech while the efference copy produces a simulation of said inner speech. In the model of AVH, an error related to the efference copy (of the inner speech action) results in a lack of predictions at some level. This in turn causes either an emotion of other-authorship (mismatch between inner speech and lack of predicted sensory consequences) or the lack of self-authorship entirely (no awareness of predicted sensory consequences means no awareness of action performance by the agent).

A further approach can be found in Lœvenbruck et al., (2018) (see Figure 1.5), which presents a variant of the predictive control account found in Blakemore et al., (1998) as adapted specifically for inner speech (see also Pickering & Garrod, 2014 for use of forward models at the earliest stages of speech production). Here, the question of agency is taken into account more prominently, as is the concern around monitoring, by assigning a comparator (see Figure 1.5, 'Comparator 2') that compares signalling from the predicted sensory end state (effected through the efference copy route) and the desired sensory state. The comparator passes on corrections to the inverse model when necessary and, since it is aware of the predicted sensory feedback, is able to account for attenuation of internal sensory signals, such as inner speech slips and auditory attenuation as alluded to previously. In comparison to the Tian and Poeppel model in Figure 1.4, the Lœvenbruck model is only marginally more complex (with regard to imagery of speech) through its inclusion of a comparator responsible for agency and signal attenuation. However, the Lœvenbruck model is subtly different in that it implicates not just premotor cortices and left IFG, but also refers to M1 as active in the efference copy route (as opposed to only in the overt execution route, cf. Tian and Poeppel).

As previously noted, the Tian and Poeppel (2012) model proposes no specific function for M1 in motor imagery of speech, while the Lœvenbruck et al., (2018) model suggests an important function for M1, namely motor programme integration. This difference in responsibilities across the two models may be due to Lœvenbruck's specific declaration that their account relates to expanded, overt, wilful speech. However, in Tian and Poeppel (2014, p. 355), the authors go so far as to say that "[…] as long as there is no overt sound, our goal of an internally induced auditory representation from a motor efference copy is valid. Potential subvocal movement is irrelevant to the interpretation […]", thereby seemingly claiming that even finely-specified articulatory movement in speech imagery does not involve M1. However, there is good experimental evidence to incorporate M1 as a functional region in motor imagery of speech, specifically from the surface electromyography (EMG) literature. McGuigan and Dollins (1989) found increases in lip EMG activation when participants silently read the bilabial letter 'P' but not when reading 'T', and vice versa for tongue EMG. Additionally, research into AVH using EMG has shown increased activity in *orbicularis oris inferior* during AVH, providing support for an account incorporating M1 such as the Lœvenbruck model in inner speech (Rapin et al., 2013). Other support comes from studies that show highly phoneme-specific, motor-like articulatory requirements for motor imagery of speech, for instance the finding that categorical perception of overtly heard consonants can be influenced by imagery of specific consonants (Scott, 2013). The finding that inner speech errors exists at the feature level even when they are not articulated overtly (Corley et al., 2011; Oppenheim & Dell, 2010) provides additional support.

Note that there is a significant degree of overlap between speech perception/production models, motor imagery of speech models and verbal monitoring accounts. In fact, it is difficult to discuss any of these processes without discussing the others. Verbal self-monitoring is thought to involve very much the same forward model architecture as discussed above – one example is the aforementioned model by Pickering and Garrod (2013, 2014), which as noted by Gauvin and Hartsuiker (2020) uses an internal model in two ways. Firstly, production is guided by prediction of semantic, syntactic and phonetic levels of the speech act (created and compared to the percept sequentially in that order much like the Tian and Poeppel approach). Secondly, the listener uses a 'prediction-by-simulation' approach to predict upcoming semantic, syntactic and phonetic forms via their own speech production, thereby connecting self-monitoring with other-monitoring, and speech production with speech perception. A debate continues surrounding the exact framework and to what degree the three processes can be linked (see Nozari, 2020; Roelofs, 2020b, 2020a).

1.4 Neurophysiological correlates of motor imagery of speech

In this section, the Tian & Poeppel (2012) and Lœvenbruck et al., (2018) models are evaluated with respect to the neuroimaging and neurostimulatory literature. Of specific interest to this thesis is (a) evidence for and against the involvement of M1, and (b) the role of inhibitory connections in the model. I briefly describe results concerning motor imagery generally, after which I discuss speech motor imagery specifically.

1.4.1 Neuroimaging of motor imagery

Several review papers have been published on the subject of neuroimaging and neurostimulation of motor imagery (for the most recent and most relevant, see Hardwick et al., 2018, but also Hétu et al., 2013 and Munzert et al., 2009), which allows for a succinct but informed discussion.

Hardwick et al., (2018) present an activation likelihood estimation (ALE) meta-analysis comprised of no less than 303 motor imagery, as well as 595 action observation, and 142 motor execution studies. Results from the motor imagery studies revealed a cortical network including the supplementary motor areas (SMA) bilaterally and the left-lateralised cingulate and putamen, bilateral inferior and middle superior parietal cortex, basal ganglia, cerebellar regions (lobule VI), and left dorsolateral prefrontal cortex (DLPFC). In contrast, motor execution showed activation in sensorimotor cortices spanning M1 and S1 bilaterally (though showing a larger cluster in the left hemisphere), premotor cortices in bilateral SMA including cingulate cortex as well as clusters in bilateral ventral premotor cortex, as well as regions in the inferior parietal lobule. Subcortical activation was found in bilateral thalamus and cerebellum (lobule VI).

On the basis of this ALE meta-analysis, the authors identify several similarities and differences between the cortical networks for motor execution and motor imagery. Similarities included activation in premotor and parietal cortices, which is in line with expectations of a forward-model account which implicates premotor regions in motor planning, and parietal cortex in somatosensory integration in both motor execution and motor imagery conditions.

One key finding was that only motor imagery task results consistently showed activation in DLPFC, when motor execution did not (see also Deiber et al., 1998 for a PET study showing similar results). The authors offer two alternative explanations for this finding: the involvement of DLPFC in executive function during action preparation (Mars & Grol, 2007), or alternatively, its role in motor inhibition (Rogasch et al., 2015). Conversely, only motor execution showed activation in M1 and S1, and only motor execution consistently activated M1. This finding is in agreement with Hétu et al., (2013), though both papers remark on the fact that activation in M1 during motor imagery may be too transient, or below the imaging threshold (for instance, due to rapid inhibition from connected areas, such as SMA or DLPFC as part of overt movement cancellation). At face value, the results from the above meta-analyses support the hypothesis that M1 is not involved in motor imagery.

The neurostimulation literature shows a different picture: specifically, studies using TMS have shown that during motor imagery, effector-specific areas of M1 show enhanced motor-evoked potentials (MEPs) relative to a baseline task (Date et al., 2015; Guillot et al., 2012; Munzert et al., 2009). Enhanced excitability was found during motor imagery of wrist (Yahagi et al., 1996), hand (Eisenegger et al., 2007; Fadiga et al., 1998; Kumru et al., 2008; Léonard & Tremblay, 2007) and leg (F. Tremblay et al., 2001). These results therefore support the hypothesis that M1 is involved in motor imagery.

Studies investigating motor imagery and motor execution often also investigate action observation since action observation is thought to engage covert action imitation (Heyes, 2011; Iacoboni, 2005), an idea fuelled in part by the discovery of mirror neurons (Gallese et al., 1996; Rizzolatti et al., 2001). While Hardwick et al., (2018) find no evidence of M1 involvement in action observation, a number of studies exist in the TMS literature that show enhanced excitability during action observation (e.g., D'Ausilio et al., 2009; Fadiga et al., 2002; Watkins et al., 2003). This discrepancy bolsters the argument that current neuroimaging methods may not be precise enough to observe M1 activation in either action observation or motor imagery conditions.

The discrepancy between TMS and fMRI findings has been discussed at length, specifically with regard to speech muscles: Möttönen & Watkins (2012) describe several explanations for the differences in findings between the two methods. A first point refers to the somatotopical and anatomical differences between participants, differences which are more apparent and more easily accounted for in TMS analysis compared to fMRI analysis, which uses a group-level, MNI- (or previously Talairach-) standardised brain space. The use of this standardised space is sensible, but individual representations of articulators may not overlap precisely and activation therein may be lost during standardisation. A second point suggests that the haemodynamic responses as measured in fMRI do not reflect the same underlying neural mechanics as those measured by the TMS – MEP method. Specifically, the haemodynamic response function relates to changes in the ratio of oxygenated blood as necessitated by increased neuronal activity, whereas MEP amplitude is related to changes in membrane potential and transsynaptic efficacy as well as temporal and spatial summation of inhibitory and excitatory connections in a targeted motor neuron pool (Möttönen & Watkins, 2012; Rösler, 2001; Rösler & Magistris, 2012). As the authors note, the neural mechanisms underlying motor excitability changes may not relate directly to energy consumption changes in a way that would show an altered haemodynamic response (Möttönen & Watkins, 2012).

1.4.2 Neuroimaging of motor imagery of speech

Neuroimaging research into motor imagery of speech, although limited, returns a similar pattern to that of general motor imagery. A meta-analysis by Perrone-Bertolotti et al., (2014) discusses a number of studies showing that certain language areas, such as inferior frontal gyrus (IFG) and inferior parietal lobule (IPL), are active in both overt and inner speech. Overt speech results in increased activation in motor and premotor regions relative to inner speech, an observation the authors note is supportive of the hypothesis that inner speech processing is similar to overt speech processing with the addition of blocking movement processes. However, the authors also note that inner speech activated certain regions more relative to overt speech, such as bilateral MTG, right cingulate gyrus, bilateral inferior parietal lobe, left dorsal frontal cortex, middle frontal gyrus (MFG) and right cerebellum. The authors acknowledge the possibility that some of this activation may be related to processing of inhibition of overt responses as opposed to covert response production.

It is unclear how different types of speech imagery tasks affect the results from a neuroimaging meta-analysis like Perrone-Bertolotti et al., (2014). Tasks range from cognitively simple tasks like word repetition to more complex tasks like verb and sentence generation. One example which uses tasks involving different levels of linguistic processes is Huang et al., (2001), which uses a letter-naming task as well as an animal-name-generation task, both in an inner speech and overt speech condition. The authors found that both silent and overt speech conditions led to activation of SMA, cingulate gyrus, SMG, MTG and insula (as well as occipital lobe since stimuli were visually presented). However, similar to general motor imagery tasks, M1 (specifically, speech-related areas of mouth, lip and tongue) was activated relative to baseline bilaterally only during both overt speech tasks, but not during the inner speech tasks. Conversely, certain areas thought to be involved in inhibition of speech, such as middle frontal gyrus, are found to be more active in the inner speech condition than in the overt speech condition (see also Basho et al., 2007). Generally, findings of studies incorporating inner speech conditions show that a shared neural substrate exists between overt execution of speech and inner speech, specifically with regard to motor planning, but there is little evidence that M1 is involved in motor imagery of speech.

The discussion on speech imagery agrees with the discussion on general imagery, however while general imagery literature does include neurostimulation studies to counteract the potentially rapid inhibition of M1 (rendering any activation difficult to detect in fMRI paradigms), no such studies exist for speech motor imagery. Several studies investigate the excitability of articulatory M1 during action observation and find increased excitability relative to a baseline condition (Fadiga et al., 2002; Möttönen & Watkins, 2012; Neef et al., 2011; Nuttall et al., 2017; Sato et al., 2010). Therefore, as no study so far has extended neurostimulation research into motor imagery of speech, it is necessary to conduct neurostimulation studies that focus on the role of motor areas in motor imagery of speech to clarify their roles in inner speech and forward models of speech.

1.5 The importance of speech motor imagery research

Inner speech is a key aspect of self-motivation (Baars, 2003), problem-solving (Perrone-Bertolotti et al., 2014; Sokolov, 1972), executive function and working memory (Baddeley & Hitch, 1974), and other common daily life tasks (Alderson-Day & Fernyhough, 2015; Grandchamp et al., 2019). As described in Morin (2009) and Perrone-Bertolotti et al., (2014), inner speech serves as a bridging cognitive device that aids in various social constructions, e.g., an internal reproduction of social mechanisms allows for perspective-taking, self-observation through psychological distancing from the self, and use as a problem-solving device. Its importance is consequently felt most strongly in those cases where pathologies cause a reduced ability to use inner speech. In autism spectrum disorder (ASD) research, the link between inner speech and theory of mind processes such as perspective-taking and self-observation, is often made (Baron-Cohen et al., 1985; Holland & Low, 2010). For instance, inner speech impairment has consistently been correlated with increased non-verbal relative to verbal reasoning in children with ASD (Lidstone et al., 2009). As Alderson-Day and Fernyhough (2015) note, there is likely to be a strong link between delayed or disrupted inner speech development and cognitively challenging processes, such as theory of mind and executive functioning.

Outside of the developmental impact of impaired inner speech, inner speech has also been put forward as the prime cognitive process behind mediation, self-consciousness (Siegrist, 1995), and self-awareness (Stamenov, 2003). Stamenov notes that a prominent symptom of schizotypal disorder is the hearing of an inner voice that is not self-attributed, also known as an auditory verbal hallucination (AVH). As has been alluded to throughout the introduction, AVH is a prime example of how the integration of internal forward models in speech processing, and the clarification of motor efference copies and their use in motor imagery of speech, can shed light on speech pathologies (Heinks-Maldonado et al., 2006).

Finally, there is a sizeable literature on rehabilitation of motor skills after stroke through motor imagery and action observation (de Vries & Mulder, 2007; Mulder, 2007) but it remains unclear whether motor imagery after stroke shows positive effects on brain plasticity with respect to M1 function (Sharma et al., 2006). A similar question exists specifically for the use of speech motor imagery in post-stroke speech impairment rehabilitation (Page & Harnish, 2012). Behavioural, neuroimaging and neurostimulatory research on M1 during speech motor imagery may therefore have an important impact on rehabilitative care.

1.6 Outline of the thesis

This thesis aims to establish whether M1 is actively involved in speech motor imagery, and thereby disambiguate between some of the prevalent models in the inner speech literature, such as Tian and Poeppel (2012) and Lœvenbruck et al., (2018). In this way, the aim relates to the role of M1 in the internal forward model that gives rise to the experience of inner speech. This question is investigated using behavioural, neuroimaging and neurostimulatory task designs. Throughout the thesis, the following major aims are investigated:

Aim No. 1:	To replicate motor imagery tasks of visual actions, and to replicate these tasks for auditory actions, on the basis of behavioural paradigms (chapters 2 and 3);
Aim No. 2:	To investigate motor imagery tasks using novel paradigms, specifically suited to speech and language (chapters 2-5);
Aim No. 3:	To investigate the role of motor cortex in motor imagery of speech, through neuroimaging and neurostimulation (chapter 3 - 5);
Aim No. 4:	To investigate the relationship between inner speech experience and motor cortex activation (chapters 2-5).

The experiments are presented as follows: two experiments focused purely on behavioural measures, such as reaction time (RT) and accuracy, one experiment focusing on RT and accuracy while also measuring whole-brain cortical activation through fMRI, and finally two experiments focusing on motor cortex specifically, employing TMS.

In chapter two, I present the results of three experiments that make use of a well-established visual imagery task as well as a novel auditory imagery task. These tasks are evaluated in two behavioural experiments: one lab-based and one delivered on-line using the Gorilla research platform. Replication forms a key part of this chapter: experiment 1 replicates previous findings from other research groups, and experiment 2 replicates our own findings for both tasks. Behavioural results (reaction time and accuracy) are used to show that imagery processes occur during the visual and auditory tasks. This chapter primarily addresses aims 1-2.

In chapter three, the visual and auditory imagery tasks are critically evaluated in a combined behavioural - fMRI experiment. This chapter replicates experiment 2 while also investigating changes in cortical activation during the visual and auditory tasks relative to a baseline condition. This chapter shows the first insight into global brain activation, and specifically motor cortex use (or lack thereof) during the imagery tasks. This chapter addresses all four aims.

In chapter four, I present the results of a second auditory imagery task that used TMS to index the excitability of lip M1 and hand M1 in three conditions: motor execution, motor imagery, and a baseline condition. This neurostimulatory method is used to build on the findings from chapter three and focuses specifically on lip motor cortex activation in an auditory motor imagery task. Crucially, this chapter presents an experiment in which lip and hand motor cortex use are investigated separately but using comparable tasks. This yields insight into effector-specific activation of motor cortex during motor imagery. This chapter addresses aims 2-4.

In chapter five, I present the results of a follow-up auditory imagery task which used TMS to index the excitability of tongue M1 during motor imagery, action observation, and a baseline condition. Neurostimulation is again used to explore motor cortex with a focus on the tongue, this time using a task that is even more closely associated with speech-specific motor programmes. This chapter addresses aims 2-4.

In chapter six, I present a discussion of all experiments and how the findings relate to the questions raised in chapter one, as well as a discussion on the limitations and implications of this body of work. Finally, suggestions on future research directions are presented.

2 Chapter Two

2.1 Introduction

2.1.1 General introduction

As stated in chapter one, there are various hypotheses regarding which processing pathways and neural underpinnings motor imagery has in common with motor execution (Decety, 1996; Guillot et al., 2012; Jeannerod, 2001). Some have argued that the assumption of speech motor imagery requiring a distinct processing pathway to any other type of motor imagery is not parsimonious (Lœvenbruck et al., 2018; Tian and Poeppel, 2012). It is generally presupposed that general motor imagery and speech motor imagery are likely to make use of similar processes, and that speech is not, from a motor standpoint, special (Jones and Fernyhough, 2007; Seal et al., 2004).

Motor imagery has been investigated using tasks which can be in one of two categories: explicit motor imagery tasks, in which participants are actively told to imagine performing a task (e.g., Kosslyn et al., 1998), and implicit motor imagery tasks, in which participants are required to perform motor imagery to complete the task, but are not provided with this strategy by the researcher (e.g., Date et al., 2015). An explicit task has the advantage of making use of the desired strategy of motor imagery, at the expense of the participant's knowledge that this is what is being investigated. An implicit task has the advantage that participants are unaware that motor imagery is the factor of interest in the experiment. An important disadvantage of using implicit tasks, however, is that the strategy used to complete the task cannot be established to be motor imagery with certainty. Despite this disadvantage I chose to focus on an implicit motor imagery task, as implicit tasks have been successful in previous studies (Eisenegger et al., 2007; Parsons, 1994; Parsons et al., 1995; Wexler et al., 1998). To alleviate the issue of not knowing whether motor imagery was used as a strategy, a short self-report questionnaire was also included.

Several studies have successfully made use of hand imagery tasks to investigate implicit motor imagery processing. One prominent task that is implicit and has consistently been shown to involve a motor imagery strategy is the Shepard-Metzler task (Shepard and Metzler, 1971), which has been extensively studied using behavioural and neuroimaging approaches (Cona et al., 2017; Richter et al., 2000). In a typical Shepard-Metzler task, participants are shown two images of objects side-by-side. Each object is made up of ten solid cubes attached to one another via one face of each cube only, resulting in a snake- or arm-like structure (see Figure 2.1). The purpose of the task is for the participant to decide whether the object they see on the left is the same object as the object on the right. The task is made difficult through rotation of one image relative to the other image by as much as 180° (in steps of 20°) on the vertical or depth axis. The results show that there is a clear linear trend of increasing reaction time as degree of rotation increases, suggesting that some form of mental rotation takes place during the decision-making process (Eisenegger et al., 2007; Kosslyn et al., 1998; Richter et al., 2000; Shepard & Metzler, 1971). The mental rotation strategy is also supported by reports from participants, which state that they performed mental rotation of one of the objects to judge whether the objects were the same or different. As with other studies from the embodiment literature (Decety et al., 1989; Oishi et al., 2000) these participant reports combined with temporal similarities between motor imagery and motor execution point to the functional equivalence hypothesis (Jeannerod, 2001). Functional equivalence in turn suggests the existence of a forward model that is active in both motor execution and motor imagery (Tian & Poeppel, 2010; see section 1.3.2).

Figure removed to avoid copyright infringement.

See Kosslyn, S. M., Digirolamo, G. J., Thompson, W. L., & Alpert, N. M. (1998). Mental rotation of objects versus hands: Neural mechanisms revealed by positron emission tomography. Psychophysiology, 35(2), p153 (figure 1) for a relevant figure.

Figure 2.1: Shepard-Metzler stimuli (top) and Parsons stimuli (bottom) adapted from Kosslyn et al., (1998). Stimuli on the left show stimuli in the 'different' condition, stimuli on the right show stimuli in the 'same' condition. Note that the classic hand laterality task from Parsons (1994) does not require showing two figures of hand as the question pertains to whether the hand is a left or right hand. The Shepard-Metzler task however does require two figures to compare whether the cubes are stacked in the same way or not.

A follow-up experiment using similar stimuli attempted to differentiate between the possible motor strategies used: Kosslyn, Ganis, et al., (2001) used the same paradigm as described above with the addition of a pre-task. In this pre-task, participants were shown a physical 3D model of a Shepard-Metzler shape and either asked to move it with their hand (endogenous rotation), or to move it using a motor (exogenous rotation). Participants were told that during the task, they should imagine rotation in the same way as they had just been shown. Participants then performed the same-different task while undergoing positron-emission tomography (PET). The results showed that those participants who had rotated the stimulus by hand in the pre-task showed activation in the M1 hand area, whereas those that had used the motor did not. This difference in activation pointed to M1 use in hand motor imagery, though it should be noted that the task had now become explicit rather than remaining implicit, which is an issue not addressed by the authors.

A second prominent task was developed and tested by Parsons (1994). This task also used implicit motor imagery and has been studied extensively (Cona et al., 2017; Date et al., 2015; Hyde et al., 2017; Parsons et al., 1995; Vingerhoets et al., 2002). In this task, participants were shown an image of a hand, and the participant had to decide whether the hand shown was a left or a right hand (left-right judgement) by pressing a foot switch under their left or right foot. The hand stimuli, similar to the Shepard-Metzler shapes, were rotated on various axes, at 30° angles (see Figure 2.1). The results from a number of variations on the task showed that an upward trend exists here as well: from 0° (hands pointing upwards) those orientations in which hands were rotated medially (hands pointing toward the midline) showed a minor but notable increase as rotation angle increases, which continued at an even greater rate when the hands are rotated laterally (hands pointing away from the midline). Kosslyn et al., (1998) performed a comparison between the Shepard-Metzler task and the Parsons hand task again incorporating PET, finding that only the Parsons hand task consistently produced activation in motor and premotor regions, while both tasks showed activation in the parietal lobe. This result suggested that mental rotation of hands, as opposed to mental rotation of objects, is more likely to engage motor planning regions. While both tasks can claim to make use of motor imagery, only the hand task showed activation of motor planning regions during motor imagery.

Another study by Hyde et al., (2017) used single-pulse TMS to record MEPs from hand motor cortex during a version of the hand rotation task. Several changes were introduced in their version, most notably the use of eye tracking to record left-right judgements (by tracking eyes to a pre-defined region on the left and right side of the screen) and the use of 45° rotation angles, reducing the number of rotations per hand from 12 to 8. The results showed that the hand rotation task resulted in increased MEP amplitudes in a subset of participants who claimed to have used a motor imagery strategy. These results showed once again that the hand region of motor cortex is involved in motor imagery of hand rotation, and also show that motor imagery is measurable using TMS and MEP facilitation methods.

Overall, the hand rotation task has consistently shown robust responses pointing towards active M1 involvement in motor imagery regardless of response modality and number of rotation stimuli, and has been used in several studies using neuroimaging and neurostimulation. Additionally, unlike the Shepard-Metzler task, it does not require for two stimuli to be shown simultaneously, a design element that might prove difficult when transposed to an auditory task. The Parsons left-right judgement hand task (hereafter referred to as the visual task) was selected as the logical basis from which to build an auditory task, primarily so that only one stimulus is presented per trial.

A novel auditory task was designed as follows. The auditory task was required to contain a judgement task, as the forced-choice element of the task is what induces the implicit motor imagery. The task should invoke motor imagery as a strategy in making a choice. As the auditory task should not at this stage be linguistic, but rather purely motoric in nature, the stimuli needed to be short and non-meaningful. The lack of semantic content was important with respect to reducing the chances of confounding linguistic processes, such as priming. Since consonant stimuli are subject to categorical perception (i.e., a consonant with a voice-onset time straddling the boundary between a voiced and unvoiced consonant is always interpreted as belonging to either category) much more so than vowels (Altmann et al., 2014), vowels were chosen as the object of the auditory judgement task.

In the visual task, the participant is asked to make a choice that is objectively correct or incorrect: a hand stimulus must always be either a left or right hand. A similar objectively correct choice was required for the auditory task: in the auditory task, a vowel could be either rounded (produced with rounded lips) or unrounded (produced with spread/unrounded lips). Using the lip as a key component had the advantage that follow-up EMG or TMS-MEP experiments could target the active lip muscles responsible for this phonetic feature (Watkins et al., 2003). Lastly, the stimuli were required to vary on a scale of difficulty, similar to a medial-lateral rotation continuum, in which some vowels would be easier to classify, and others would be more difficult to classify. Here, it was possible to use the elements of closeness to native vowels: stimuli were produced on a continuum ranging from (near)native vowels to non-native vowels and back towards the native vowel range. The two tasks were engineered to contain the same number of stimuli, on a similar difficulty scale, with the expectation that similar motor imagery strategies would be used across tasks.

It is important to note that the native/non-nativeness of vowels is an informative, intentional factor. In the same way that medial rotations for hands are seen more often by people in the first person and are therefore expected to be easier to imagine due to a greater degree of familiarity (Parsons, 1994), native vowels are expected to be easier to imagine as well, allowing for a hypothesis related to expected difficulty per stimulus. If the difference between the lateral and medial orientations is influenced by how often participants are exposed to stimuli in life, the same holds true for native and non-native auditory stimuli. Specifically, the difficulty of processing a motor plan is related to the exposure to a stimulus (e.g., a common hand position is more easily identified as left/right than an uncommon hand position. This logic can be extended to stimuli that are commonly heard and produced, such as native vowels, and uncommonly heard and produced, such as non-native vowels). It should also be noted that the distinction rounded/unrounded is less intuitive than left/right hand. That is, participants are generally aware of a left-right directional distinction, but not aware of vowel classification features such as roundedness. Piloting showed that when participants were made aware of the rounded/unrounded distinction, they did not find it difficult to judge native stimuli as rounded or unrounded, similar to how medial stimuli (as expected) were classified as left or right with ease. It is not initially considered an inherently more difficult decision-making task.

2.1.2 Research aims

(1) Experiment 1:
   (a) To replicate the visual hand task as a basis for speech motor imagery research. This aim examines the following hypotheses:
      i. Lateral rotations are expected to result in longer reaction times than medial rotations due to longer motor trajectories;
      ii. No differences between left and right hand stimuli reaction times are expected as there is no inherent difficulty difference between the stimuli types;
      iii. Medial rotations are expected to be judged with significantly lower accuracy than lateral rotations, but accuracy is also expected to be very high (> 90%) (Hyde et al., 2013);
   (b) To create a task similar to the visual task employing auditory stimuli, leading to motor imagery of the articulators (lip movement). This aim examines the following hypotheses:
      i. Non-native vowels are expected to result in longer reaction times than native vowels, as it takes longer to compute the motor plans required for motor imagery of lesser-known articulations;
      ii. No differences between rounded and unrounded vowels are expected, as there is no inherent difficulty difference between the stimuli types;
      iii. Accuracy is expected to be similarly affected, so a small but significant difference is expected for the auditory task;

(2) Experiment 2:
   (a) To test whether the task developed in experiment 1 can be adapted to be (i) delivered online, (ii) with fewer stimuli, while retaining the motor imagery effects. The aim examines the following hypotheses:

i. Both the visual and auditory task show reaction time effects from experiment 1 remain when fewer stimuli are used;
ii. Both the visual and auditory task show accuracy effects from experiment 1 remain when fewer stimuli are used;

Additionally, the following hypotheses relating to the VISQ and potential learning effects are investigated throughout:

(3) It is possible that RT and accuracy scores are correlated with VISQ scores, although it is not certain which scores specifically would be relevant. The hypothesis holds that one of the VISQ scores may correlate with RT and accuracy scores, whereas the null hypothesis assumes that there is no relation between VISQ scores and behavioural measures;
(4) Learning or habituation is expected to take place during the tasks, but there is no expectation that the rate at which learning or habituation occurs differs between the visual and auditory tasks.

The results for these additional hypotheses (3-4) can be found in Appendix A.

2.2 Experiment 1

This experiment was the first attempt to replicate the study by Parsons, (1994) and test a speech motor imagery task that follows a similar design.

2.2.1 Methods

2.2.1.1 *Participants*

Forty-four participants took part in experiment 1 (23F, 21M, mean age = 21y 9m, *SD* = 2y, 6m, age range = 18-26). All participants were recruited via the university psychology subject pool (SONA) and all were speakers of Standard Southern British English (SSBE). Handedness was established using the Edinburgh Handedness Inventory (Oldfield, 1971): all participants were right-handed. All participants had normal or corrected-to-normal vision. Participants reported no history of neurological or psychiatric disorder. A Pure-Tone Average (PTA) audiometry test was performed in accordance with guidelines published by the British Society of Audiology (British Society of Audiology, 2012). Experiments were undertaken with informed verbal and written consent of each participant, following guidelines set out by the University College London Research Ethics Committee (UREC #0599/001). All research was carried out in accordance with the Declaration of Helsinki.

2.2.1.2 *Materials*

The hand stimuli were created using the Blender 3D computer graphics program (blender.org, 2015) and made use of a freely available model which was modified for use in the lab (https://3dhaupt.com/3d-model-anatomy-rigged-hands-low-poly-vr-ar-game-ready-blender/ ([accessed May 2017]). The hand stimuli consisted of an image of the model left hand which was rotated at eight angles: 0° (fingers pointing upward), 45°, 90°, 135°, 180° (fingers pointing down), 225°, 270° and 315° (see Figure 2.2). In addition, the original model was vertically mirrored to produce the right-hand stimulus, which

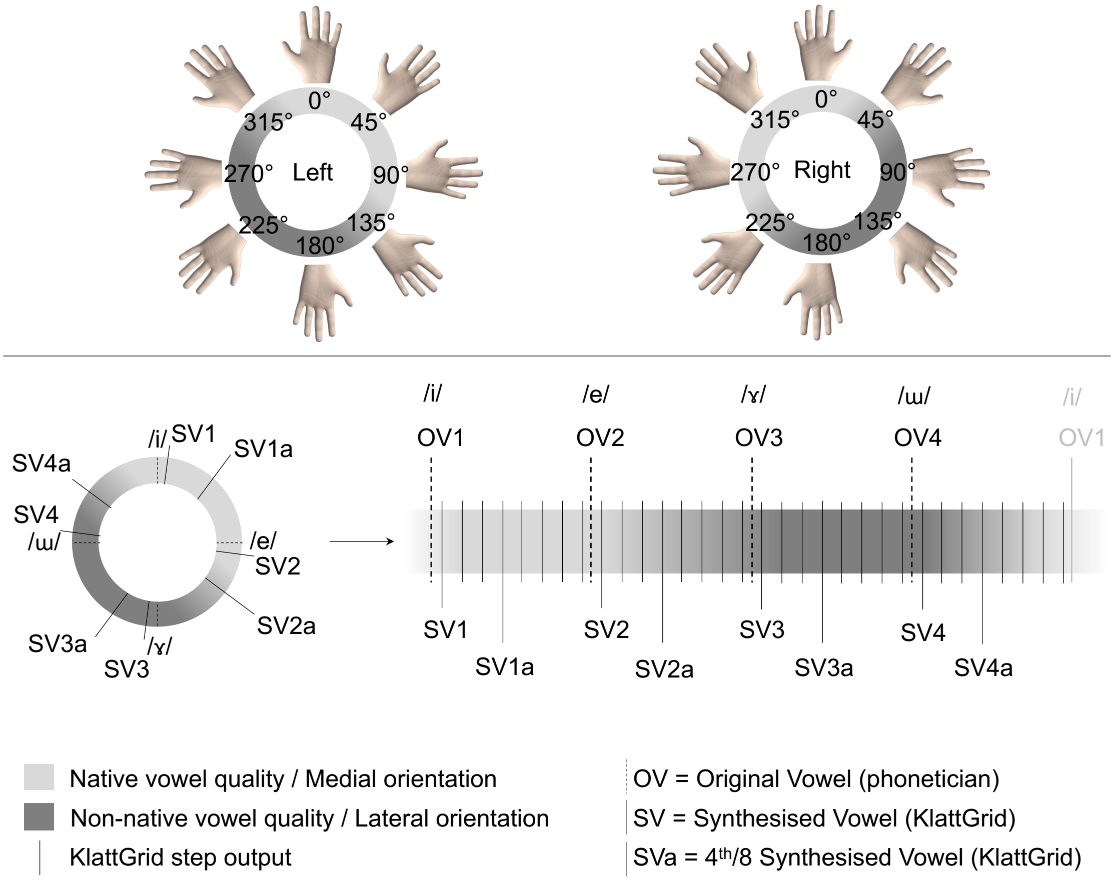


Figure 2.2: Hand and vowel stimuli at various rotations as used in experiments 1 – 3. The coloured rings show the similarities for the stimuli in the abstract, with lighter colours classified as easy and darker colours classified as hard stimuli. The coloured ring 'unfurls' to the band showing the original vowels, synthesised vowels and the 8-step KlattGrid continuum Praat provided. An additional continuum was created for the rounded vowels, but is not shown.

was also rotated at the same eight angles. Both stimuli used a palm view rather than a back-of-hand view, as palm view has been associated with increases in motor cortical areas – suggesting a motor imagery strategy – while the back-of-hand view has been associated more with increases in occipital areas – suggesting an image rotation strategy (Zapparoli et al., 2014). For the auditory stimuli, 25 tokens of each target vowel (/i/, /e/, /ɤ/, /ɯ/, /o/, /ɔ/, /ɵ/, /ʉ/) were recorded in a sound-proof anechoic chamber by a native 24y female SSBE speaker. Selection and modification of original vowel stimuli (see Figure 2.2) was performed using Praat (Boersma & Weenink, 2006). The elements for the stimulus contexts (/s/ and /zə/ were similarly identified from a number of tokens and extracted for later use. Once each vowel token had been extracted, all elements were normalised to 70dB SPL using the open-source Praat Vocal Toolkit (Corretge, 2012). Two vowels were moderately pitch-corrected (< 5Hz) to ensure all vowels retained the same pitch (233Hz).

In-house Praat scripts were then used to create KlattGrid-synthesised vowel stimuli with a pass Hann filter (8000Hz) applied. This procedure used two of the original vowels, e.g., /i/ and /e/, as input to create a stepwise continuum between the two original vowel tokens, yielding 8 synthesised vowels in equal steps. As shown in Figure 2.2, each of these continua had a starting vowel, such as SV1, which was a synthesised version of the original vowel token, e.g., /i/, as well as an end vowel, such as SV2, which was a synthesised version of the original vowel token /e/. Between SV1 and SV2, an additional synthesised vowel token was taken from the continuum at step number 4, in this example continuum SV1a, which was slightly closer in quality to the starting vowel than the end vowel (since there are 2 steps between step 1 and 4, but 3 steps between step 4 and 8). As a result, for the purposes of analysis, if SV1 was classified as native, so was SV1a. If a synthesised vowel was classified as non-native, as was the case for /ɤ/ (SV3), then the synthesised vowel (SV3a) was also classified as non-native.

Pilot testing (*n* = 5) with SSBE speakers using only the IPA KlattGrid vowel(e.g., SV1, SV2, etc.) and recorded vowels showed that SSBE speakers could not tell whether the vowels presented were synthetic or original productions, which showed the synthesised vowels were of sufficient quality to be used in testing. The vowel stimuli were further embedded in the consonantal frame (e.g., /si:zə/) as repetition of single vowels is likely to create confusion after continuous repeated exposure (Strange et al., 1983; Strange & Jenkins, 2013; personal communication, Bronwen Evans, May 2017). Contexts and vowels were concatenated with an overlap of 15ms.

The stimuli were displayed on a 21.5" monitor while participants sat ~70cm away from the monitor. Auditory stimuli were presented through Sennheiser HD25-SP II headphones at 70dB SPL. Two foot switches were used to collect response data similar to the setup in Parsons (1994) and Date et al., (2015). Foot switches were used rather than hand button presses to eliminate interference between motor planning in hand due to motor imagery and motor planning in hand due to pressing a button. Reaction time was measured as time from stimulus onset (in both auditory and visual conditions) to foot switch press. A webcam (unbranded) outputting .mov files at a resolution of 720 x 480 pixels was used to view participants live, but not to record them.

2.2.1.3 Procedure

Participants were welcomed and the study was explained. Verbal and written consent was obtained. Participants positioned their hand on the edge of a table, with the back-of-hand side up so that they could not see their palms. The participant positioned their left foot on the left foot switch and their right foot on the right foot switch. The experimenter briefly explained the two tasks in the study. For the visual task, participants were told they would see a left or right hand on the screen, and that they should respond whether the hand was a left or right hand by using the foot switches. For the auditory task, participants were told they would hear a non-word with one strong vowel, and that they should respond whether the vowel was rounded or unrounded. As this distinction is not commonly known (though intuitively understood) an example of an unrounded and rounded vowel was given prior to the practice task, and the experiment did not continue until participants were able to paraphrase the difference between the two categories correctly. Participants were made aware that they would be informed when a new task would begin, as the instructions would always be displayed prior to a new task. They were also made aware that the on-screen display showed not only the stimulus, but also the choice that each foot switch represented, e.g., pressing the left foot switch represented a choice for left hand, and the right foot switch represented a choice for the right hand (see Figure 2.3). For the auditory task the assignment was counterbalanced across participants (remaining the same per participant). Counterbalancing was not implemented for the visual task as it was deemed too cognitively tasking to ask participants to press the right button when a left hand was presented (and vice versa) to the extent that it might influence reaction time measurements. Participants were told to answer as quickly and as accurately as possible

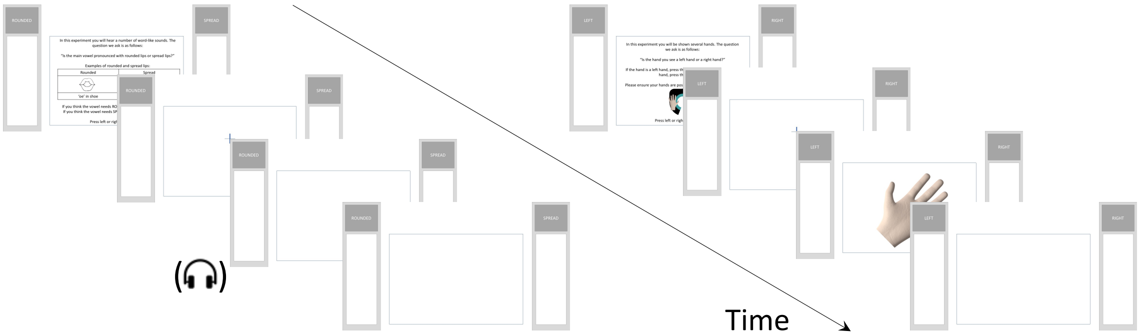


Figure 2.3: Single trial (including initial instructions) for the auditory task (left) and the visual task (right). This version was used in experiments 1 and 3 and modified slightly for on-line use for experiment 2 (borders removed).

but were not told how to make a judgement in either task (that is, no strategy was presented). Participants were made aware that the experimenter would leave the room, but that they would be monitored via a webcam and that they were free to leave at any time if they wished to stop the experiment. The participant was then shown the practice task which showed a subset of stimuli, and also indicated whether they had made the correct choice once they had pressed the foot switch. Once the practice task was completed, participants were told that the real task would not show them if they had made the correct or incorrect choice.

Per task, 192 stimuli were presented (4 blocks of 48 stimuli, with self-timed breaks between blocks), with the total stimuli per session coming to 384. Stimuli were counterbalanced across and within participants with regard to presentation and order of task. Each new block the task was switched between visual and auditory. It is important to note that, due to the visual stimulus being shown in the course of 1 frame and the auditory task being shown as time unfolds (starting with a sibilant before being followed by the vowel of import), a difference in reaction time between tasks is expected and cannot be taken as way to judge whether tasks are differentially difficult.

2.2.1.4 Participant assessments

General cognitive ability was assessed using the Montreal Cognitive Assessment (MOCA, Nasreddine et al., 2005). The test is a one-page 30-point test administered in 10 minutes, assessing short-term memory recall (5 points), visuospatial abilities (4 points), executive function (4 points), attention, concentration and working memory (5 points), language (6 points) and orientation to time and place (6 points). A score of <26 is indicative of mild cognitive impairment, and results in the participant's exclusion from data analysis. See Appendix B for the MOCA.

Levels and types of inner speech experience were assessed using the Varieties of Inner Speech Questionnaire (VISQ, McCarthy-Jones & Fernyhough, 2011). This assessment assesses the phenomenological properties of inner speech using an 18-question questionnaire, giving insight into four dimensions: dialogicality (monologue vs dialogue quality), condensed vs expanded speech quality, evaluative/motivational quality, and the degree to which inner speech includes voices other than the participant's own. This assessment was used only for correlational purposes – participants were not excluded for any reason related to the VISQ. See Appendix C for the VISQ.

Hearing ability and potential hearing loss was assessed using a standard Pure Tone Audiometry (PTA) assessment, using a clinically approved audiometer (Maico, MA41) Frequencies of 1000Hz, 2000Hz, 4000Hz, 8000Hz, 250Hz and 500Hz were tested (in that order) in each ear. Participants were excluded from any analysis if a volume of 25dB SPL was exceeded at any frequency, as this is symptomatic of mild hearing loss. See Appendix D for the PTA and handedness assessments. Participants were also asked general background questions to assess language background and medical background where relevant. See Appendix E for the background questions.

2.2.1.5 Analysis

Descriptive statistics were performed for the VISQ data to improve understanding of the participants' experience of inner speech. A first analysis involved a repeated-measures Analysis of Variance (ANOVA) to compare relative difficulty levels of tasks and difficulty levels within them by looking at reaction times (in milliseconds and further log-transformed to account for expected skew) and accuracy (% correct). Data analysis was performed on log-transformed data, but raw reaction times are reported in the text and figures as they are more meaningful and aid understanding. The focal analysis of the tasks was a repeated-measures ANOVA comparing the means of dependent variables reaction time and accuracy across the conditions Difficulty (easy vs hard), Hand (left vs right hand) in the case of the visual task and Shape (rounded vs unrounded lips) in the case of the auditory task, resulting in a 2 x 2 ANOVA with dependent variables reaction time and accuracy. How a stimulus was categorised depended on modality: in the visual task, a stimulus was difficult if it pointed laterally (i.e., hand turned away from the midline) and easy if the hand was turned medially (towards the midline) as lateral rotations require more complex trajectory planning involving multiple muscle groups (Parsons, 1994). In the auditory task, a vowel was hypothesised to be difficult if it was a non-native vowel, and easy if it conformed to a (near-)native vowel (see Figure 2.2): this corresponded to the similar assumption that motor planning for unknown vowels would require more complex trajectory planning than motor planning for well-known vowels. Hedges' *g* effect size statistic was used to enable comparisons of significant effects across experiments, due to the small number of participants (< 30) in experiment 3 (see Lakens, 2013).

An additional one-way ANOVA was run to check whether there was an effect of counterbalancing in the auditory task (whether the left foot switch entailed a choice for rounded lips or unrounded lips and vice versa for the right foot switch). Finally, a repeated-measures ANOVA was performed to investigate rates of learning by looking at reaction times and accuracy across blocks, resulting in a 2 x 4 ANOVA. The results of this last analysis can be consulted in Appendix A. Follow-up *t*-tests were performed where appropriate. All results were Bonferroni- and Greenhouse-Geisser-corrected where appropriate.

2.2.2 Results

2.2.2.1 VISQ

Participants rated each question on the VISQ on a scale of not applying to them (1 -3) or applying to them (4-6, see 2.2.1.4 and Appendix C). The closer a score comes to the maximum of 6, the more that type of inner speech applies, and the closer it comes to 1, the less that type of inner speech applies: for dialogic inner speech a higher score reflects

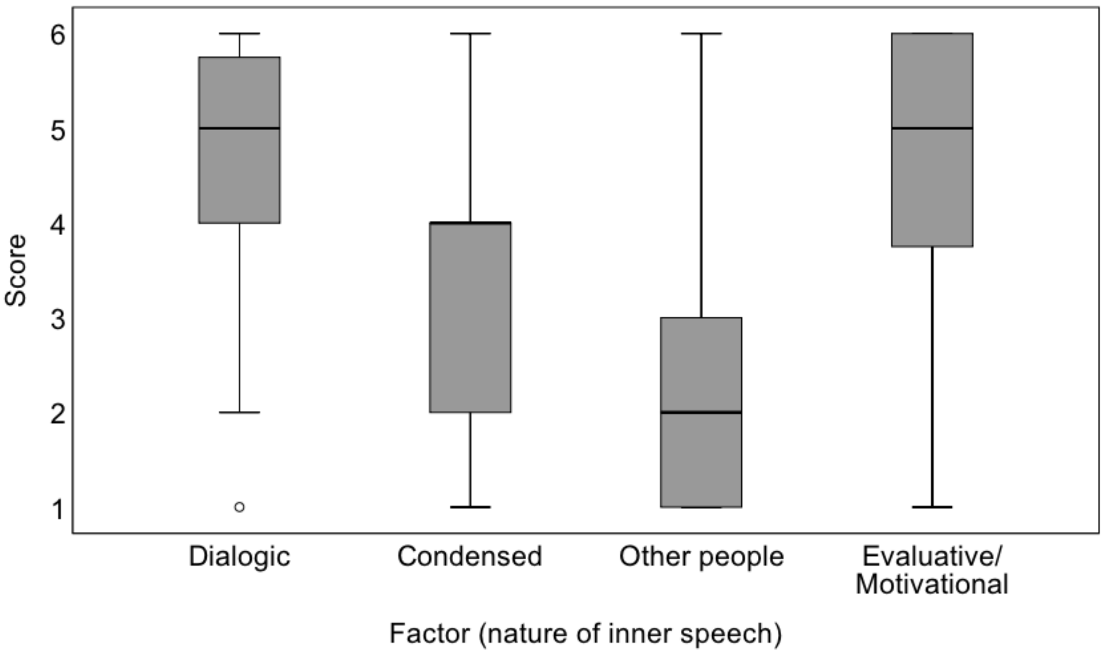


Figure 2.4: VISQ scores per type of inner speech for experiment 1. Scores are amalgamated across five questions that relate to each factor. See Appendix C for the questionnaire. (1 = Certainly does not apply, 6 = Certainly does apply). All plots show standard boxplots unless indicated otherwise.

a back-and-forth quality of inner speech (with oneself), while a lower score reflects a running monologue. In the case of condensed inner speech, a higher score means more condensed speech while a lower score means more fully-formed, expanded inner speech. However, a high score for other people in inner speech suggests the amount of inner speech that appears to involve other imagined speakers is high, but a low score does not suggest that the participant experiences only themselves talking to themselves – it merely means experiences involving other people in inner speech are not (or less) present. Finally, a high evaluative/motivational inner speech suggests a significant amount of inner speech revolves around evaluating or motivating oneself, while a low score suggests inner speech is more general and not specific to this situation. A score of 3.5 shows that a type of inner speech is neither particularly applicable nor inapplicable and suggests that neither extreme is prevalent.

Overall participants reported that evaluative/motivational speech and dialogic inner speech were the most common types of inner speech experienced (see Figure 2.4). Condensed inner speech showed medium to lower scores overall, suggesting that overall participants feel neither condensed nor expanded inner speech more than the other, though the boxplot shows expanded speech is somewhat more prevalent. A majority of participants reported that other people were rarely involved in their inner speech experience, however this does not suggest necessarily that hearing oneself speak is prevalent.

2.2.2.2 Overall statistics

Overall statistics showed that out of 16896 trials, 22 trials showed raw reaction time values that represented out-of-bounds values, i.e., where the participant did not respond within the allotted time window (10,000ms). A further 815 trials were excluded on the basis that they exceeded 3SDs from the mean reaction time. This yielded 16059 trials for analysis. No further trials were excluded on the basis of accuracy, since imagery may take place regardless of whether the answer provided is correct or not. As individual accuracy was sufficient for each participant (lowest overall accuracy = 83.03%) no participants were excluded.

2.2.2.2.1 Overall statistics: Reaction time

Mean raw reaction time across tasks was 1184ms (*SD* = 243ms). As expected, raw reaction times showed a slight leftward skew (characteristic of reaction time data), so the log10 was used to derive a dataset more closely following a normal distribution. Skew was reduced from 1.129 to 0.343 as a result. Further analysis was performed on log10 reaction time data. Data were aggregated by Task (auditory vs visual task) and Difficulty (hard vs easy) and a repeated-measures ANOVA was performed to ascertain whether the tasks and difficulty levels were comparable based on reaction time data.

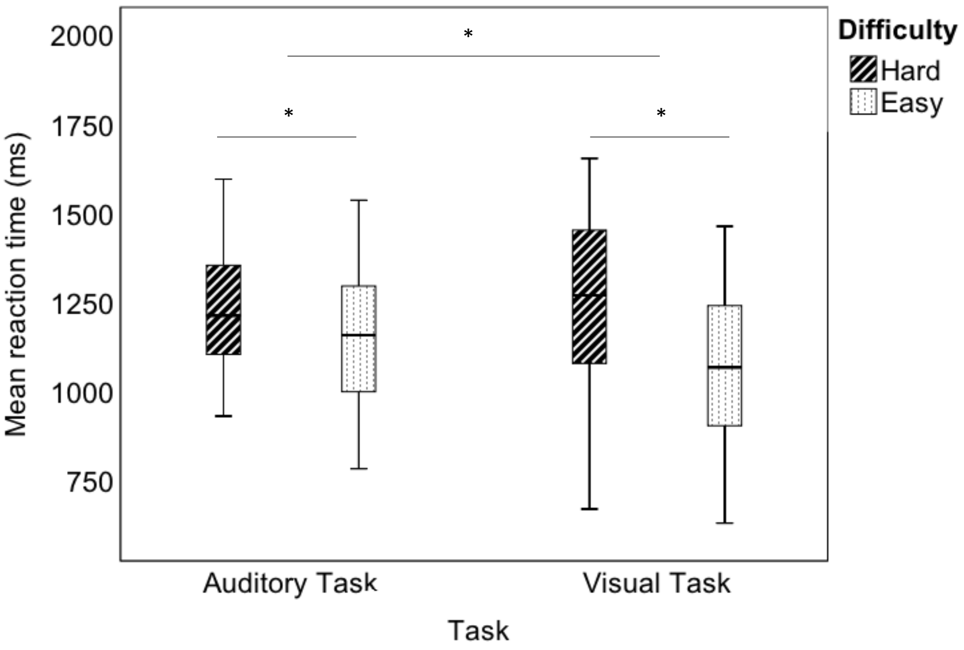


Figure 2.5: Main effects and interaction of Task and Difficulty for reaction time in experiment 1.
* = significant at p = .05. All plots show standard boxplots unless indicated otherwise.

There was a main effect of Task (F(1,43) = 6.07, p = .018, η_p^2 = .124, Hedges' g = 0.384), showing that participants were slightly but significantly faster in the visual task (M = 1151ms, SD = 223ms) compared to the auditory task (M = 1217ms, SD = 222ms). There was a main effect of Difficulty (F(1,43) = 167.29, p < .001, η_p^2 = .796, Hedges' g = 0.739), showing clearly that participants reacted faster to easy stimuli (M = 1112ms, SD = 189ms) than to hard stimuli (M = 1256ms, SD = 196ms). There was also an interaction effect of Task x Difficulty (F(1,42) = 19.92, p < .001, η_p^2 = .317), see Figure 2.5.

A follow-up paired-samples t-test was performed to investigate the interaction, which showed that the difference in reaction times between the auditory and the visual task was significant only for easy stimuli (MD = 114ms, SD = 15ms, t(43) = 3.93, p <.001, Hedges' g = -0.575), and not for hard stimuli (MD = 17ms, SD = 22ms, t(43) = 0.92, p = .365, Hedges' g = -0.150). These results suggest that hard stimuli resulted in slower responses than easy stimuli regardless of task, but that reaction times for hard stimuli were similar across tasks, while reaction times for easy stimuli were significantly faster in the visual task compared to the auditory task. The same analysis was run on the accuracy data.

2.2.2.2.2 Overall Statistics: Accuracy

Overall mean accuracy across tasks was 93.0% (*SD* = 25.5%). A repeated-measures ANOVA was performed to ascertain whether the tasks and difficulty levels were comparable on the basis of accuracy data.

There was a main effect of Task (*F*(1,43) = 38.89, $p < .001$, $\eta_p^2 = .475$, Hedges' *g* = -1.202), showing that participants were significantly more accurate in the visual task (*M* = 96.7%, *SD* = 3.3%) than in the auditory task (*M* = 89.3%, *SD* = 8.6%). There was also a main effect of Difficulty (*F*(1,43) = 72.86, $p < .001$, $\eta_p^2 = .628$, Hedges' *g* =-1.147), showing that easy stimuli showed greater accuracy (*M* = 96.4%, *SD* = 3.9%) than difficult stimuli (*M* = 89.6%, *SD* = 7.3%). There was also an interaction between Task x Difficulty (*F*(1,43) = 11.089, *p* = .002, $\eta_p^2 = .205$), see Figure 2.6.

A follow-up paired-samples *t*-test was performed to investigate the interaction, which showed a significant difference in the hard stimuli between the auditory and visual tasks (*MD* = -10.2%, *SD* = 11.1%, *t*(43) = -6.07, $p < .001$, Hedges' *g* = 1.128) and also in the easy stimuli between auditory and visual tasks (*MD* = -4.7%, *SD* = 7.8%, *t*(43) = -4.01, $p < .001$, Hedges' *g* = 0.805). There was also a difference between easy and hard stimuli in the auditory task (*MD* = 9.5%, *SD* = 9.3%, *t*(43) = 6.76, $p < .001$, Hedges' *g* = -0.965) as well as

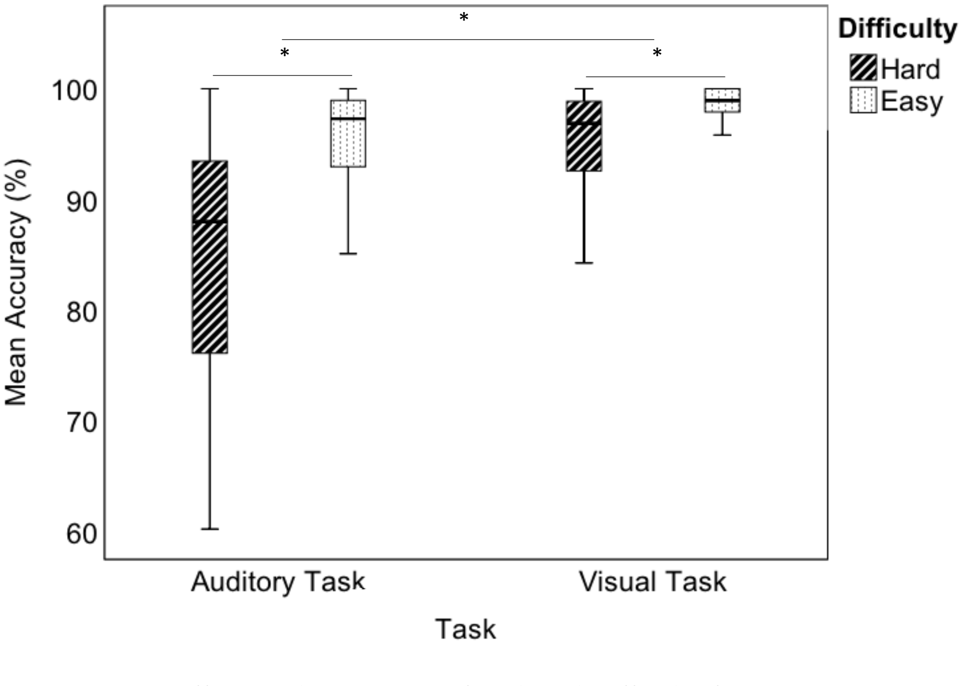


Figure 2.6: Main effects and interaction of Task and Difficulty for accuracy, experiment 1. * = significant at *p* = .05.

the visual task (*MD* = 4.0%, *SD* = 5.3%, $t(43) = 5.06$, $p<.001$, Hedges' $g = -0.924$). The significant interaction can therefore be attributed to the fact that the difference in accuracy between hard and easy stimuli is larger in the auditory task than in the visual task.

2.2.2.3 Visual task

2.2.2.3.1 Visual task: Reaction time

To investigate whether there were differences in reaction time in the visual task, a repeated-measures ANOVA with the factors Difficulty (hard vs easy) and Hand (left hand vs right hand) was run. The results showed that there was a significant main effect of Difficulty ($F(1,43) = 122.95$, $p < .001$, $\eta_p^2 = .741$, Hedge's $g = 0.946$), showing that participants' responses to easy (medial) stimuli (*M* = 1055ms, *SD* = 213ms) were significantly shorter than responses to hard (lateral) stimuli (*M* = 1248ms, *SD* = 251ms). However, there was no main effect of Hand ($F(1,43) < 0.01$, $p = .963$, $\eta_p^2 <.001$). Similarly, there was no interaction Difficulty x Hand ($F(1,43) = 0.16$, $p = .692$, $\eta_p^2 = .004$), see Figure 2.7. These results suggest that the rotation of the stimulus (lateral vs medial) was the main and only factor that results in increases in reaction times.

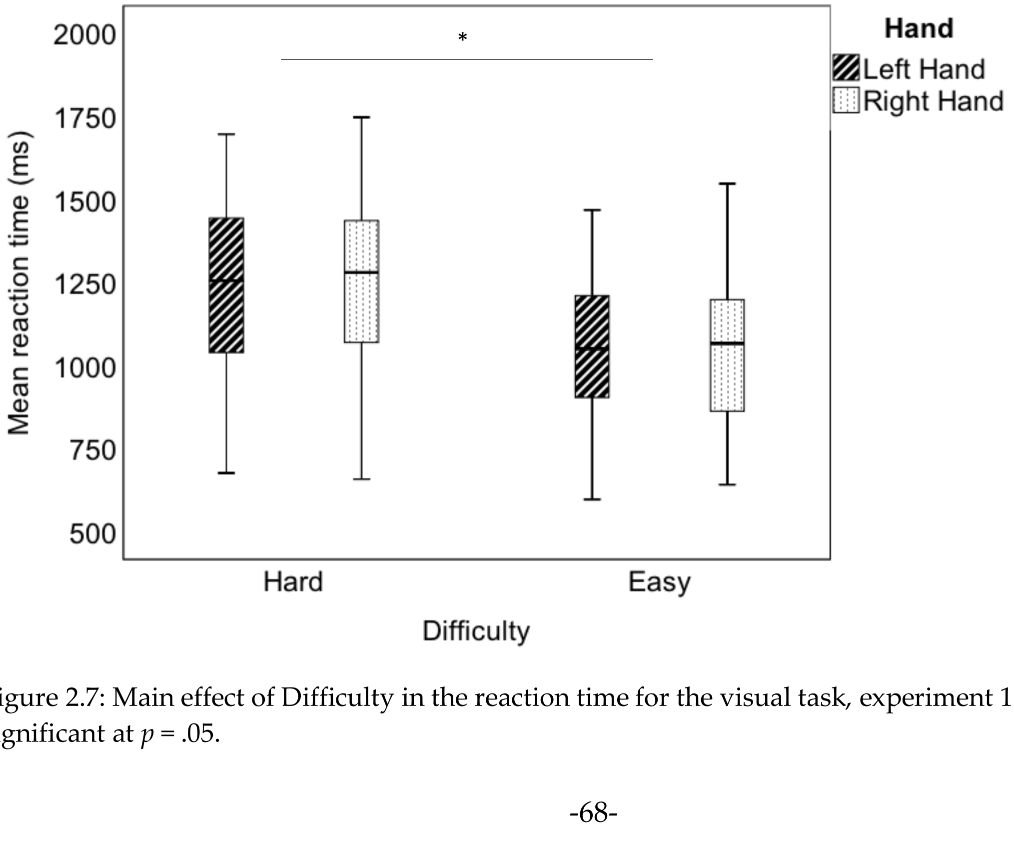


Figure 2.7: Main effect of Difficulty in the reaction time for the visual task, experiment 1. * = significant at $p = .05$.

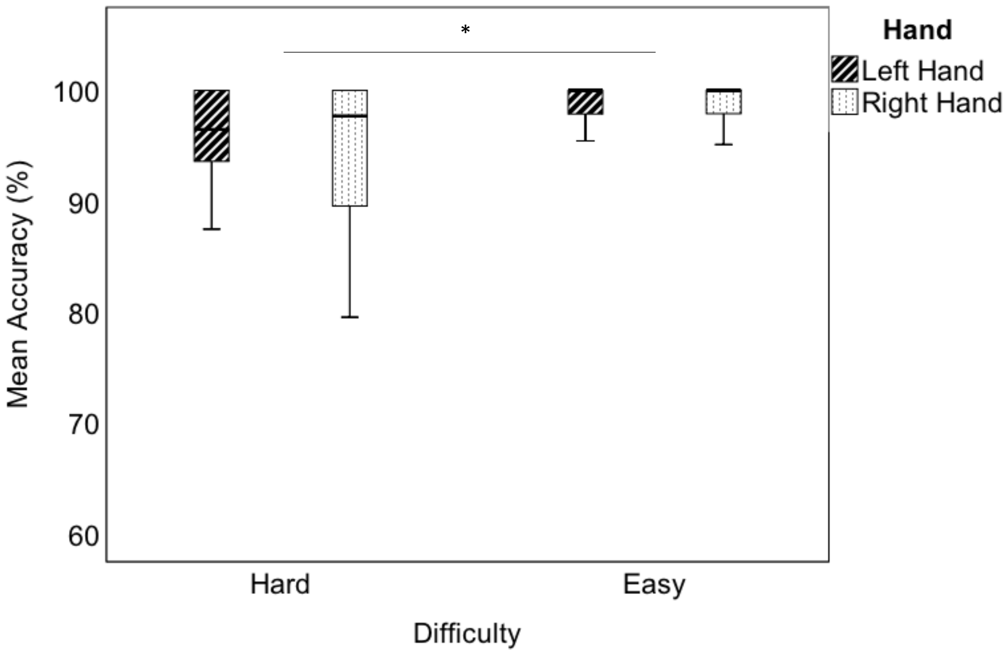


Figure 2.8: Main effect of Difficulty in accuracy for the visual task, experiment 1. * = significant at $p = .05$.

2.2.2.3.2 Visual task: Accuracy

To investigate whether there were differences in accuracy in the visual task, a repeated-measures ANOVA with the factors Difficulty (hard vs easy) and Hand (left hand vs right hand) was run. As with the reaction time data, there was a significant main effect of Difficulty ($F(1,43) = 25.68$, $p < .001$, $\eta_p^2 = .374$, Hedges' $g = -0.917$), which was due to slight but significantly higher accuracy for easy stimuli ($M = 98.7\%$, $SD = 1.3\%$) than for hard stimuli ($M = 94.7\%$, $SD = 5.9\%$). However, there was no effect of Hand ($F(1,43) = .68$, $p = .415$, $\eta_p^2 = .016$) and no interaction effect Difficulty x Hand ($F(1,43) = .602$, $p = .442$, $\eta_p^2 = .014$), see Figure 2.8. Note that the extremely high accuracy for the easy stimuli amounted to a ceiling effect. These results confirmed those from the reaction time data, showing that the main factor responsible for differences in reaction time was whether the stimulus is rotated laterally or medially.

2.2.2.4 *Auditory task*

As the auditory task featured counterbalancing across participants, a one-way ANOVA was conducted to check whether there was a difference between groups which had the rounded and unrounded choice assigned to the left and right foot switch, and vice versa. The ANOVA showed that there was no difference in reaction time between groups ($F(1,42) = 0.13$, $p = .716$, $\eta_p^2 = .003$), with those who used the former configuration responding no faster ($M = 1211$ms, $SD = 229$ms) than those who used the latter configuration ($M = 1245$ms, $SD = 365$ms).

2.2.2.4.1 Auditory task: Reaction time

To investigate whether there were differences in reaction time in the auditory task, a repeated-measures ANOVA with the factors Difficulty (hard vs easy) and Shape (rounded vs unrounded vowel) was run. There was a significant main effect of Difficulty ($F(1,43) = 48.88$, $p < .001$, $\eta_p^2 = .532$, Hedges' $g = 0.454$) which was due to significantly longer reaction times for hard stimuli ($M = 1270$ms, $SD = 236$ms) than for easy stimuli ($M = 1171$ms, $SD = 231$ms), see Figure 2.9. Surprisingly there was also a significant main effect for Shape ($F(1,43)=11.25$, $p = .002$, $\eta_p^2 = .207$, Hedges' $g = 0.335$) as a result of rounded stimuli eliciting significantly shorter reaction times ($M = 1179$ms, $SD = 222$ms) than

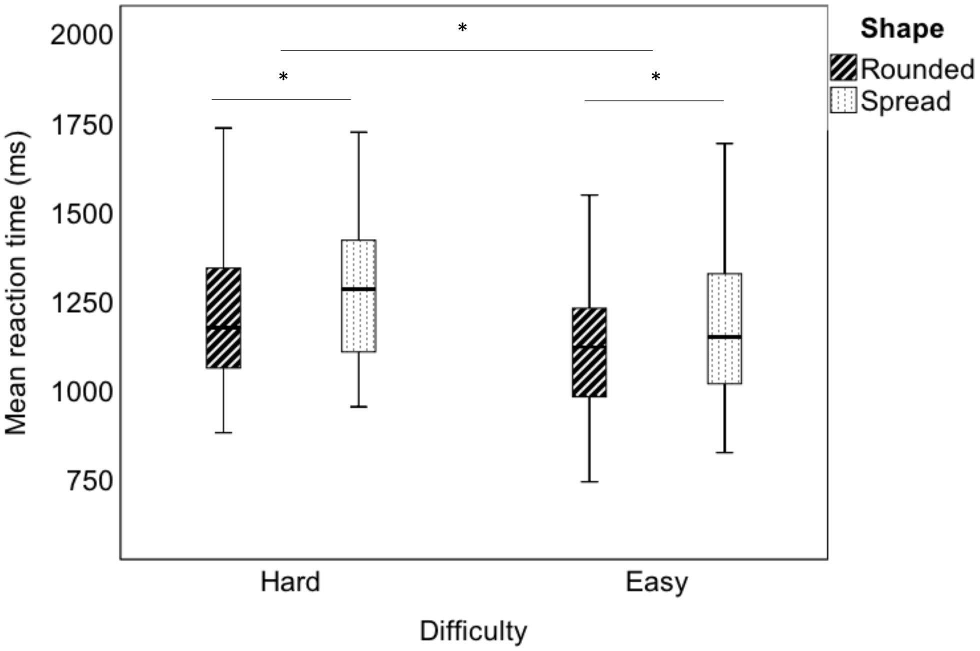


Figure 2.9: Main effect of Difficulty in reaction time for the auditory task, experiment 1. * = significant at $p = .05$.

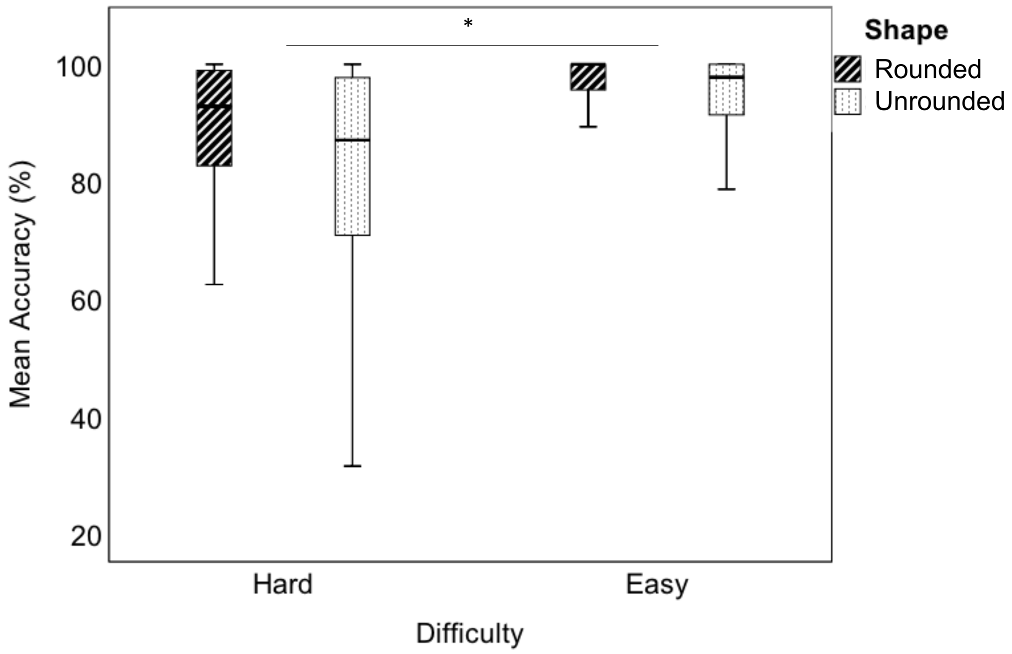


Figure 2.10: Main effect of Difficulty in accuracy for the auditory task, experiment 1. * = significant at p = .05.

unrounded stimuli (M = 1261ms, SD = 263ms). However, the Difficulty x Shape interaction was not significant ($F(1,43) = 0.05$, $p = .828$, $\eta_p^2 = .001$).

2.2.2.4.2 Auditory task: Accuracy

To investigate whether there were differences in accuracy in the visual task, a repeated-measures ANOVA with the factors Difficulty (hard vs easy) and Shape (rounded vs unrounded vowel) was run. As was the case for the reaction time data, there was a significant main effect of Difficulty ($F(1,43) = 48.03$, $p < .001$, $\eta_p^2 = .528$, Hedges' $g = -0.995$), since easy stimuli had higher accuracy (M = 93.9%, SD = 7.9%) than hard stimuli (M = 84.1%, SD = 11.3%). Unlike the reaction time data there was no main effect of Shape ($F(1,43) = 3.31$, $p = .076$, $\eta_p^2 = .071$). Once again there was no significant interaction Difficulty x Shape ($F(1,42) = 0.74$, $p = .396$, $\eta_p^2 = .017$), see Figure 2.10.

2.2.3 Discussion

Experiment 1 had two main aims: the first aim was to replicate the visual hand task found elsewhere in the literature as a basis for motor imagery research. The second aim was to create such a task that could be used in the auditory modality, involving lip motor imagery. The hypotheses and results are discussed here.

Hypotheses for this experiment were as follows (see below for further discussion):

- (a) To replicate the visual hand task as a basis for speech motor imagery research. This aim concerns the following hypotheses:
    - i. Lateral rotations were expected to result in higher reaction times than medial rotations due to longer motor trajectories. This hypothesis was confirmed;
    - ii. No differences between left and right hand stimuli reaction times were expected as there is no inherent difficulty difference between the stimuli types. This hypothesis was confirmed;
    - iii. Medial rotations are expected to be judged with significantly lower accuracy than lateral rotations, but accuracy is also expected to be very high (> 90%)(Hyde et al., 2013). This hypothesis was confirmed.

- (b) To create a task similar to the visual task employing auditory stimuli, leading to motor imagery of the articulators (lip movement). This aim covers the following hypotheses:
    - i. Non-native vowels are expected to result in higher reaction times than native vowels, as it takes longer to compute the motor plans required for motor imagery of lesser-known articulations. This hypothesis was confirmed;
    - ii. No differences between rounded and unrounded vowels are expected, as there is no inherent difficulty difference between the stimuli types. This hypothesis was not confirmed since responses to rounded stimuli showed shorter reaction times than unrounded stimuli;
    - iii. Accuracy is expected to be similarly affected, so a small but significant difference is expected for the auditory task. This hypothesis was confirmed.

The visual task as used here aligns closely to versions used by other researchers (Hyde et al., 2017; Parsons, 1994; Zapparoli et al., 2014) though was made in-house. The results from the visual task conform to those elsewhere and confirm all visual task hypotheses:

the main factor for increased reaction time is the lateral rotation of the hand, rather than any effect of whether the hand shown is a left or right hand. Similarly, accuracy was significantly higher for medial than for lateral rotations, although with very small differences between conditions. This suggests that the visual task itself performed as expected, and that participants followed similar strategies to participants in other studies. Indeed, as shown in Kosslyn, Thompson, et al., (2001) and Zapparoli et al.,(2014), participants engaged in this task showed increased activation in motor areas related to movement planning and movement. In previous literature, this difference is attributed to more complex motor trajectories for more difficult rotations, which can be seen both in execution and imagery conditions. As the results in experiment 1 concur with those in previous literature, the assumption of differences in motor imagery can be made here as well.

The auditory task was entirely novel. The hypotheses stated that trials with non-native vowels would show longer response times than trials with native vowels, and that there would be no differences between rounded and unrounded vowels. The results from the auditory task showed that the expected effect of difficulty is present: it was the case that participants spend more time considering whether a vowel is produced with unrounded or rounded lips if they are less familiar with the vowel. Similarly, their accuracy was reduced for vowels that are non-native and so for those vowels for which no motor trajectories exist. These results confirmed the hypotheses. The reasoning explaining the results in the visual task are equally valid here: more difficult motor planning trajectories, most likely due to physical constraints and an inability to quickly produce motor imagery of the required stimulus, result in longer reaction times.

Unlike the visual task, there was also a significant difference for the factor along which participants made their choice: rounded stimuli were categorised more quickly as such than unrounded vowels. These results contradict the hypothesis which stated that there would be no differences between rounded and unrounded vowels: there was no expectation that mouth shape would inherently result in a difference in reaction time, similar to how there was no expectation that reaction time to either side hand would be inherently different. Interestingly, this difference was only present in reaction time data, and not in the accuracy data. This may be an instance of a speed-accuracy trade-off, in which one category of stimuli was consistently more difficult to identify. This could also be due to an issue with the unnaturalness of the KlattGrid stimuli, compared to the relative naturalness of the hand stimulus, which apart from rotation, was identical regardless of trial. A follow-up experiment using more natural stimuli might be more informative in this respect, as it would help to see whether the issue is with KlattGrid stimuli or inherently with differences in mouth shape.

Using the visual task design to create the auditory task allowed for the implicit assumption that the tasks would be similar in several ways but different in others. One goal of creating the auditory task was to make it similarly difficult to the visual task. As can be seen in the overall statistics, participants were both faster and more accurate in the visual motor imagery task than in the auditory task, suggesting that the visual task is easier overall. However, as previously noted, in both tasks accuracy remained well above chance, with an average difference in accuracy of around 7.5% between tasks. Additionally, the rate of learning across tasks appeared to be different: participants improved more rapidly in the visual than in the auditory task. That is, the visual task shows faster reaction time relative to the auditory task without changes in accuracy relative to the auditory task. The auditory task remained quite difficult with consistently lower accuracy across blocks as well slower reaction times. The fact that the auditory task was significantly harder does not in and of itself pose an issue, but it is nevertheless useful to be aware of since participants always performed both tasks together.

Overall, experiment 1 appears to have been largely successful, with the aims largely met. Experiment 2 continued this paradigm with several changes to further improve the task.

2.3 Experiment 2

As briefly discussed previously (2.1.2), experiment 2 explored several changes to the paradigms used in experiment 1. Specifically, I wanted to investigate whether both the visual and auditory tasks were sufficiently robust to use fewer stimuli. This was in anticipation of a possible fMRI and TMS adaptation. Since both of these methods rely on averaging neural data, it was preferable to develop a design with a larger number of repetitions of few stimuli over a design with a large number of stimuli presented few times. Additionally, I was interested to see whether the behavioural findings would hold if testing occurred in a supervised environment, i.e., with a researcher present, compared to an unsupervised environment. If results are unchanged, the task could be used for large-scale (< 100 participants) on-line study of these tasks. As stated in the hypotheses (2.1.2.), no differences are expected due to either change. If results are unchanged from experiment 1, the shorter task would be considered sufficiently robust to be used in fMRI and TMS designs. One additional change was made not to use KlattGrid sounds, but instead use natural stimuli. This change was made to see whether the unexpected finding of differences in behavioural results could be due to unnatural-sounding synthesised vowels stimuli.

2.3.1 Methods

2.3.1.1 Participants

Thirty-six participants took part in experiment 2 (26F, 10M, mean age = 21y 10m, *SD* = 2y 6m, age range = 19-27, correct ages only). Due to an issue where some participants filled in their date of birth without changing the year (pre-set to the day the experiment took place), this age data should not be considered accurate – instead an assumption must be made that all participants were aged 18-30 as required for sign-up on the experiment webpage via the UCL subject pool. All participants were recruited via the university psychology subject pool (SONA) and all were speakers of Standard Southern British English (SSBE). Thirty-four participants were right-handed, while two were left-handed. All participants had normal or corrected-to-normal vision. Participants reported no history of neurological or psychiatric disorder. Experiments were undertaken with informed verbal and written consent of each participant, following guidelines set out by the University College London Research Ethics Committee (UREC #0599/001). All research was carried out in accordance with the Declaration of Helsinki.

2.3.1.2 Materials

The hand stimuli consisted of a subset of those used in experiment 1: an image of the model left hand which was rotated at four angles: 0° (fingers pointing upward), 90°, 180° (fingers pointing down) and 270°, yielding eight stimuli (four for left hand, four for right hand). In the case of the auditory stimuli, the KlattGrid vowels were not reused in an effort to investigate whether the effect of Shape found in experiment 1 might be due to the artificiality of the KlattGrid stimuli (see section 2.2.3). Instead, the original 25 tokens of each target vowel (/i/, /e/, /ɤ/, /ɯ/, /o/, /ɔ/, /ɵ/, /ʉ/) recorded in a sound-proof anechoic chamber by a native 24y female SSBE speaker were repurposed for experiment 2. As before, selection and modification of final vowel stimuli was performed using Praat (Boersma and Weenink, 2006). The elements for the stimulus contexts were similarly identified and extracted for later use. Once each vowel token had been extracted, all elements were normalised to 70dB SPL using the open-source Praat Vocal Toolkit (Corretge, 2012). Two vowels were moderately pitch-corrected (< 5Hz) to ensure all vowels retained the same pitch (233Hz). The vowel stimuli were embedded in a consonantal frame (e.g., /sozə/). Contexts and vowels were concatenated with an overlap of 15ms. In this way the process yielded eight natural-sounding non-words with one primary vowel (four rounded and four unrounded).

This experiment was created and hosted on the on-line Gorilla Experiment Builder (www.gorilla.sc, Anwyl-Irvine et al., 2018), referred to as ‘Gorilla’. Data was collected between 01 March 2019 and 01 May 2019. Participants were recruited through the UCL subject pool. As the task was performed locally, stimuli were displayed on screens with various sizes. Participants were requested to use headphones for the task, and although a Gorilla function to check whether headphones were in use was initially implemented, the feature was removed by Gorilla before the study was begun due to its reported inaccuracy. Participants used the left and right arrow buttons to respond, and the task could not be completed on tablets or phones; a laptop or desktop was required.

2.3.1.3 Procedure

Participants were greeted with a welcome screen and the study and information required from participants was explained. Written consent was obtained. Participants then filled out an information sheet that provided background information on age, language background, vision and hearing ability (as in experiment 1, see Appendix E). Participants were told that they would take part in two tasks, one visual and one auditory, and that each task would be explained in advance and a practice task provided. Instructions remained largely unchanged from experiment 1, save for the manner of responding: for the visual task, participants were told they would see a left or right hand on the screen, and that they should respond whether the hand was a left or right hand by using the arrow keys on their keyboard. For the auditory task, participants were told they would hear a non-word with one strong vowel, and that they should respond whether the vowel was rounded or unrounded. As this distinction is not commonly known (though intuitively understood) an example of an unrounded and rounded vowel was given prior to the practice task. Participants were made aware that they would be informed when a new task would begin, as the instructions would always be displayed prior to a new task. They were also made aware that the on-screen display showed not only the stimulus, but also the choice that each arrow key represented, e.g., pressing the left arrow key represented a choice for left hand, and the right arrow key represented a choice for the right hand. For the auditory task the assignment was again counterbalanced across participants (remaining the same per participant). Participants were asked to answer as quickly and as accurately as possible but were not provided with a strategy for making their choice. Prior to each task, each participant was shown the practice task which showed a subset of stimuli, and also indicated whether they had made the correct choice once they had pressed the arrow key. Once the practice task was completed, participants were told that the real task would not show them if they had made the correct or incorrect choice.

Per task, 48 stimuli were presented (3 blocks of 16 stimuli, with short breaks between blocks), with the total stimuli per session coming to 96. This represented a sharp decrease compared to experiment 1, with only 1/8th of total stimuli; however, since the number of stimuli themselves have been halved from 8 to 4, the number of stimuli shown is in fact 1/4th that of experiment 1. Stimuli were counterbalanced across and within participants with regard to presentation and order of task. Each task was finished before the next was begun, i.e., blocks were not interleaved as in experiment 1.

2.3.1.4 Analysis

The analysis for experiment 2 was identical to that of experiment 1 (see 2.2.1.5). The only difference here is that the MOCA was not performed since the testing occurred online.

2.3.2 Results

2.3.2.1 VISQ

Overall participants again reported that evaluative/motivational speech and dialogic inner speech were the most common types of inner speech experienced, though scores were slightly lower than in experiment 1. Condensed inner speech showed medium to lower scores overall, suggesting that participants overall feel a mixture of condensed and expanded inner speech (or neither type). Again, a large number of participants reported that other people were rarely involved in their inner speech experience, see Figure 2.11.

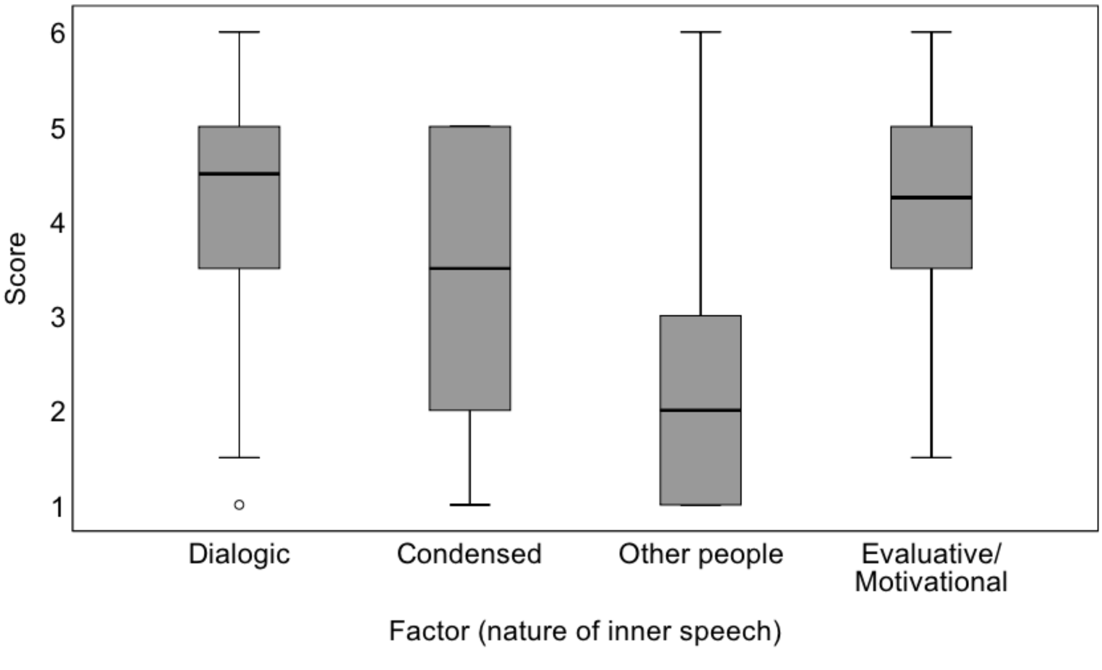


Figure 2.11: VISQ scores per type of inner speech for experiment 2. Scores are amalgamated across five questions that relate to each factor. See Appendix C for the questionnaire. (1 = Certainly does not apply, 6 = Certainly does apply).

2.3.2.2 *Overall statistics*

Overall statistics showed that out of 3456 trials, 30 trials had raw reaction time values that represented out-of-bounds values or exceeded 3SDs from the mean log10 reaction time. This yielded 3426 trials for analysis. One participant was found to have at-chance accuracy scores overall (55%), and upon investigation they appeared to have misunderstood the auditory task, in which they showed 15% accuracy. Their data was removed from the dataset for both tasks. No further individual trials were excluded on the basis of accuracy, since imagery may take place regardless of whether the answer provided is correct or not. One final participant was removed from the means dataset as their mean reaction times exceeded 3SDs from the sample mean. Due to this removal 34 participants and 3234 trials remained for analysis.

2.3.2.2.1 Overall Statistics: Reaction time

Mean raw reaction time across tasks was 1464ms (*SD* = 539ms). The log10 was used to derive a dataset more closely following a normal distribution. Skew was reduced from 2.022 (raw reaction time) to 0.459 (log10) as a result. Further analysis was performed on log10 reaction time data. Data were aggregated by Task (auditory vs visual task) and Difficulty (hard vs easy) and a repeated-measures ANOVA was performed to ascertain whether the tasks and difficulty levels were comparable based on reaction time data.

There was a main effect of Task ($F(1,33) = 14.18$, $p = .001$, $\eta_p^2 = .301$, Hedges' $g = 0.616$), showing that participants were significantly faster in the visual task (M = 1328ms, SD =

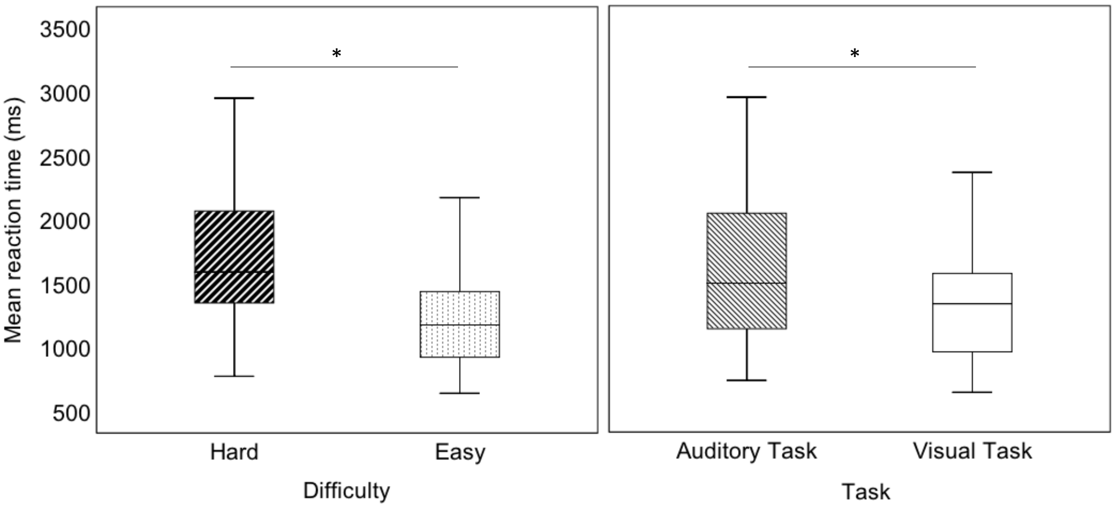


Figure 2.12: Main effect of Difficulty and Condition in reaction time across tasks, experiment 2. * = significant at $p = .05$.

366ms) compared to the auditory task (*M* = 1600ms, *SD* = 511ms). There was a main effect of Difficulty (*F*(1,33) = 116.29, $p < .001$, $\eta_p^2 = .779$, Hedges' *g* = 0.974), showing that participants reacted faster to easy stimuli (*M* = 1240ms, *SD* = 349ms) than to hard stimuli (*M* = 1688ms, *SD* = 458ms), see Figure 2.12. Unlike in experiment 1, there was no interaction effect of Task x Difficulty (*F*(1,33) < 0.01, $p = .765$, $\eta_p^2 = .338$) showing that differences across tasks per difficulty were similar.

2.3.2.2.2 Overall Statistics: Accuracy

Overall mean accuracy across tasks was 87.1% (*SD* = 16.2%). A repeated-measures ANOVA was performed to ascertain whether the tasks and difficulty levels were comparable on the basis of accuracy data.

There was a main effect of Task (*F*(1,33) = 38.53, $p < .001$, $\eta_p^2 = .539$, Hedges' *g* = -1.215), which was due to participants being more accurate in the visual task (*M* = 92.3%, *SD* = 8.7%) than in the auditory task (*M* = 81.9%, *SD* = 8.2%). There was also a main effect of Difficulty (*F*(1,33) = 198.02, $p < .001$, $\eta_p^2 = .857$, Hedges' *g* = -2.425), showing that participants were more accurate for easy stimuli (*M* = 96.6%, *SD* = 4.7%) than for hard stimuli (*M* = 77.6%, *SD* = 9.9%), see Figure 2.13. There was also an interaction effect for Task x Difficulty (*F*(1,33) = 56.149, $p < .001$, $\eta_p^2 = .630$), see Figure 2.14.

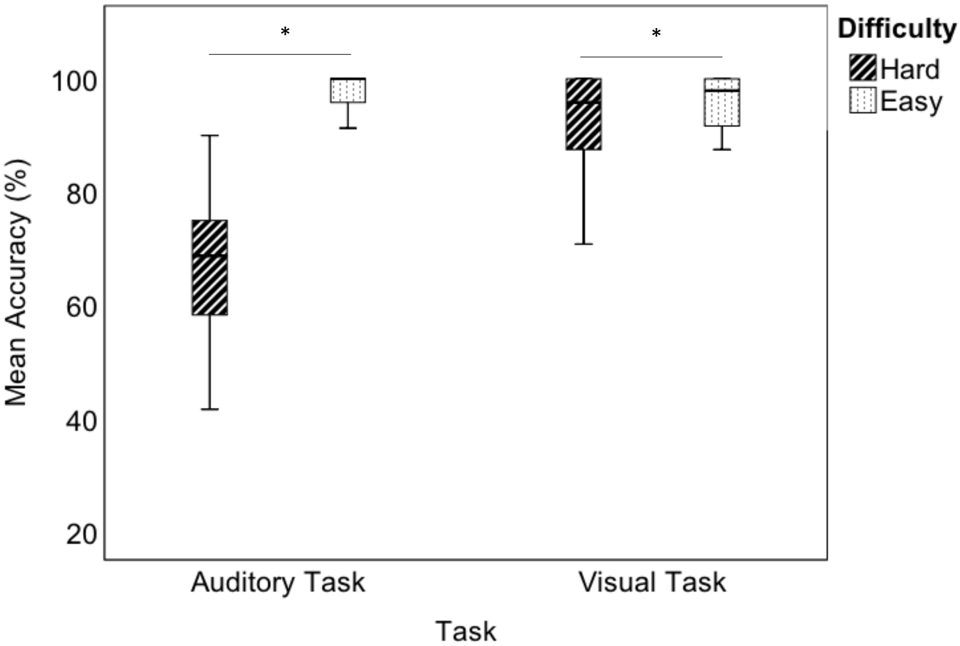


Figure 2.13: Main effect of Task in accuracy across task, experiment 2. * = significant at $p = .05$.

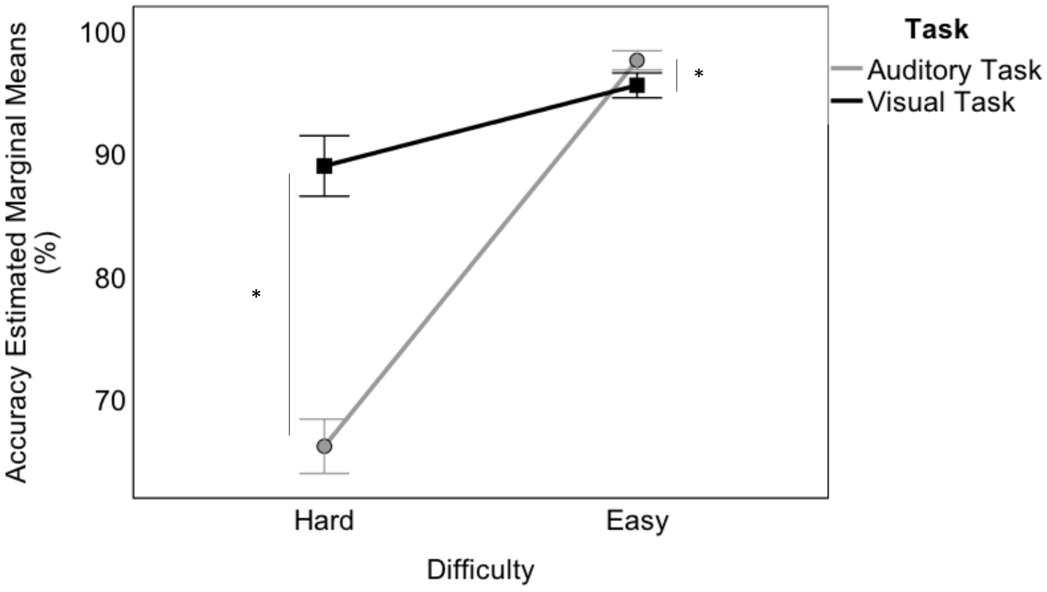


Figure 2.14: Interaction effect of Difficulty for accuracy across tasks, experiment 2. Error bars indicate +- 1SEM. * = significant at *p* = .05.

Follow-up paired-samples *t*-tests were performed to investigate this interaction. There was a large difference between the easy and difficult stimuli in the auditory task (*MD* = 31.5%, *SD* = 11.3%, *t*(33) = 16.31, *p* < .001, Hedges' *g* = -3.209), while there was a smaller difference between the easy and difficult stimuli in the visual task (*MD* = 6.6%, *SD* = 13.6%, *t*(33)= 2.82, *p* = .008, Hedges' *g* = -0.592). The more relevant difference lay in the difference between tasks: there was a large difference in the hard condition between the auditory and visual task (*MD* = -22.9%, *SD* = 18.9%, *t*(33) = -7.06, *p* <.001, Hedges' *g* = 1.654) whereas there was a much smaller difference for the easy stimuli (*MD* = 2.0%, *SD* = 4.8%, *t*(33) = 2.47, *p* = .019, Hedges' *g* = -0.381). While there was an effect of Difficulty in both tasks (harder stimuli are always less accurate) and there was also clearly an effect of Task (auditory task is less accurate overall), the difference in accuracy scores also depended on a combination of these two factors, and it was clear from the interaction that relative to easy stimuli, hard stimuli in the auditory task were disproportionately harder, while hard stimuli in the visual task were only slightly (but still significantly) harder (see Figure 2.14).

2.3.2.3 Visual Task

2.3.2.3.1 Visual task: Reaction time

To investigate whether there were differences in reaction time in the visual task, a repeated-measures ANOVA with the factors Difficulty (hard vs easy) and Hand (left hand vs right hand) was run. The results showed that there was a significant main effect of Difficulty ($F(1,33) = 112.15$, $p < .001$, $\eta_p^2 = .773$, Hedges' $g = 0.962$), showing that participants' responses to hard (lateral) stimuli (M = 1537ms, SD = 453ms) were significantly longer than responses to easy (medial) stimuli (M = 1118ms, SD = 318ms). However, there was no main effect of Hand ($F(1,33) = 1.93$, $p = .174$, $\eta_p^2 = .055$). Similarly, there was no interaction Difficulty x Hand ($F(1,33) = 0.96$, $p = .336$, $\eta_p^2 = .028$). These results suggest that the rotation of the stimulus (lateral vs medial) was the only factor resulting in increases in reaction times (see Figure 2.15).

2.3.2.3.2 Visual task: Accuracy

To investigate whether there were differences in accuracy in the visual task, a repeated-measures ANOVA with the factors Difficulty (hard vs easy) and Hand (left hand vs right hand) was run. As with the reaction time data, there was a significant main effect of Difficulty ($F(1,33) = 7.93$, $p = .008$, $\eta_p^2 = .194$, Hedges' $g = -0.588$), which was due to higher accuracy for easy stimuli (M = 95.6%, SD = 5.8%) than for hard stimuli (M = 89.0%, SD = 14.6%). However, there was no effect of Hand ($F(1,33) = 0.29$, $p = .593$, $\eta_p^2 = .009$) and no interaction effect Difficulty x Hand ($F(1,33) = 0.47$, $p = .497$, $\eta_p^2 = .014$), see Figure 2.15. These results agree with those from the reaction time data, as well as experiment 1,

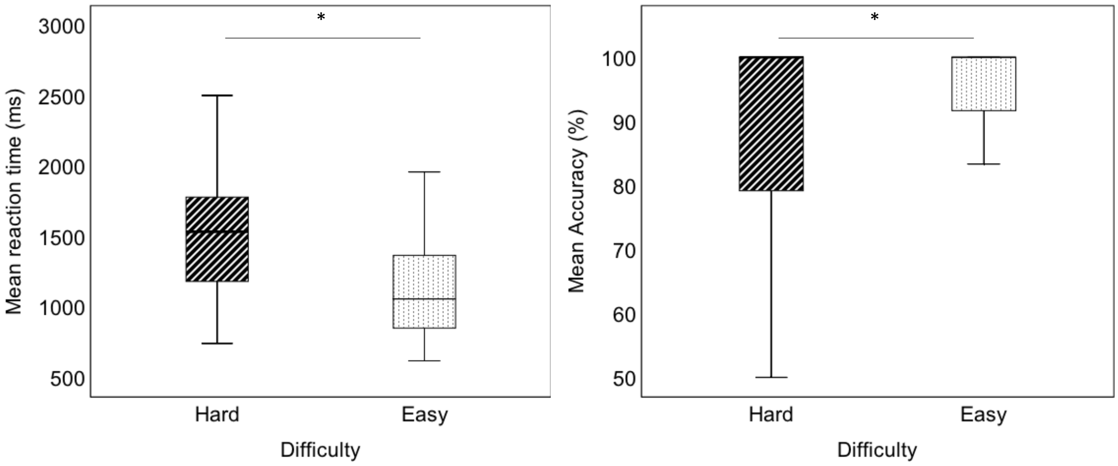


Figure 2.15: Main effects of Difficulty in reaction time and accuracy for the visual task, experiment 2. * = significant at $p = .05$.

showing that the main factor responsible for differences in reaction time is whether the stimulus is rotated laterally or medially.

2.3.2.4 Auditory Task

The auditory task featured counterbalancing across participants, so once again a one-way ANOVA was performed to check whether there was a difference between groups which had the rounded and unrounded choice assigned to the left and right arrow keys, and vice versa. The ANOVA showed that there was no difference of reaction time between groups ($F(1,32) = 0.04$, $p = .844$, $\eta_p^2 = .001$), with those who used the former configuration responding no faster ($M = 1603.9$ms, $SD = 497.8$ms) than those who used the latter configuration ($M = 1641.6$ms, $SD = 581.9$ms).

2.3.2.4.1 Auditory Task: Reaction time

To investigate whether there were differences in reaction time in the auditory task, a repeated-measures ANOVA with the factors Difficulty (hard vs easy) and Shape (rounded vs unrounded vowel) was run. There was a significant main effect of Difficulty ($F(1,33) = 55.08$, $p < .001$, $\eta_p^2 = .625$, Hedges' $g = 0.753$) which was due to significantly longer reaction times for hard stimuli ($M = 1842$ms, $SD = 624$ms) than for easy stimuli ($M = 1363$ms, $SD = 457$ms), see Figure 2.16. Unlike in experiment 1, there was no significant main effect of Shape ($F(1,33) = .092$, $p = .763$, $\eta_p^2 = .003$). Similar to experiment 1, there was no interaction of Difficulty x Shape ($F(1,33) = 1.08$, $p = .306$, $\eta_p^2 = .032$) This suggests that for the stimuli used in experiment 2, the main factor for differences in

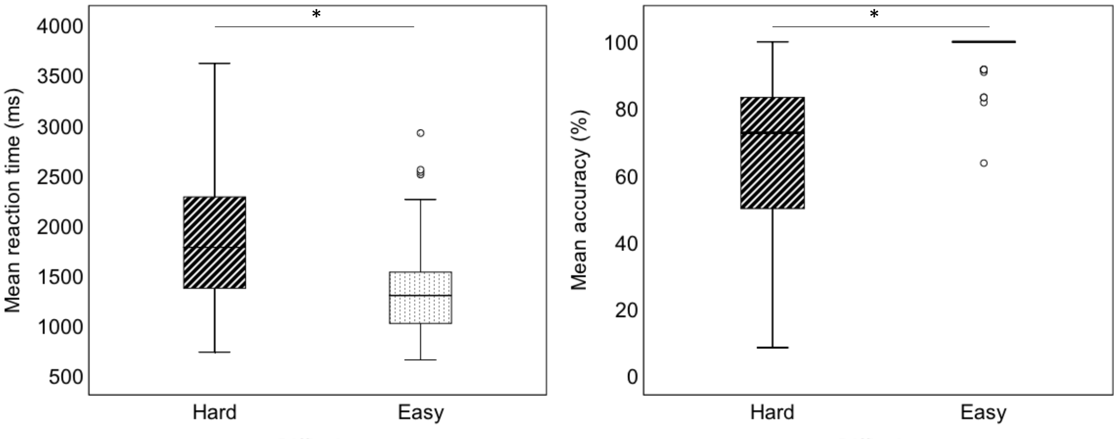


Figure 2.16: Main effect of Difficulty in reaction time and accuracy for the auditory task, experiment 2. * = significant at $p = .05$.

reaction time is the difficulty of the stimulus, as opposed to the shape used to produce the stimuli or some interplay between the two factors.

2.3.2.4.2 Auditory Task: Accuracy

To investigate whether there were differences in accuracy in the auditory task, a repeated-measures ANOVA with the factors Difficulty (hard vs easy) and Shape (rounded vs unrounded vowel) was run. There was a significant main effect of Difficulty ($F(1,33) = 269.61$, $p < .001$, $\eta_p^2 = .891$, Hedges' $g = -3.236$) due to an overall much greater accuracy for easy stimuli ($M = 97.6\%$, $SD = 4.7\%$) than for hard stimuli ($M = 66.0\%$, $SD = 12.8\%$). This pattern can also be seen in Figure 2.16, where accuracy for easy stimuli was so consistently high that those who did not achieve 100% were marked as outliers. There was also a main effect of Shape ($F(1,33) = 5.09$, $p = .031$, $\eta_p^2 = .134$, Hedges' $g = -0.603$) due to greater accuracy for rounded vowels ($M = 85.8\%$, $SD = 13.4\%$) than for unrounded vowels ($M = 77.8\%$, $SD = 12.8\%$). However, there was also an interaction Difficulty x Shape (F(1,33) = 8.22, $p = .007$, $\eta_p^2 = .199$) such that main effects are interpreted with some caution (see Figure 2.17).

Follow-up *t*-tests were performed to investigate this interaction. The *t*-tests highlighted the following results: for hard stimuli the unrounded vowels have significantly lower

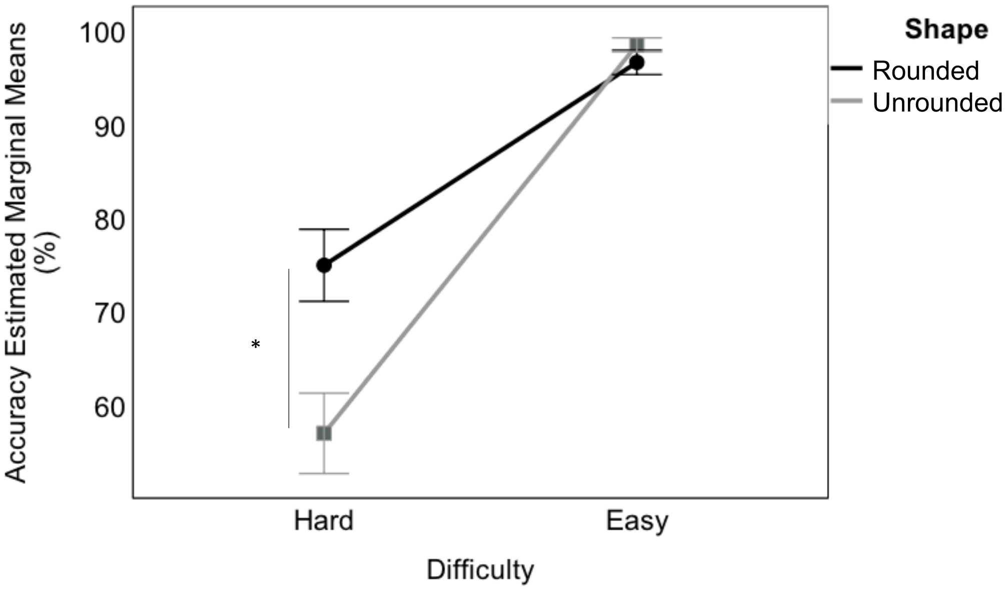


Figure 2.17: Interaction effect of Difficulty and Shape in the accuracy data for the auditory task in experiment 2. Error bars indicate +- 1SEM. * = significant at $p = .05$.

accuracy than rounded vowels (*MD* = -17.9%, *SD* = 40.1%, *t*(33) = 2.61, *p* = .013, Hedges' *g* = 0.746), while this difference is non-existent for the easy stimuli, where unrounded vowels and rounded vowels (*MD* = 1.9%, *SD* = 8.3%, *t*(33) = 1.30, *p* = .201, Hedges' *g* = -0.299) are both very accurate categories. It is likely that this difference in accuracy for the hard stimuli is also what drives the main effect of Shape, since that main effect is not particularly strong. These results indicate that accuracy is influenced by the interplay of difficulty and shape of the vowel stimuli: although hard stimuli always have significantly lower accuracy than easy stimuli, there is a difference in accuracy for hard stimuli such that rounded stimuli are judged more accurately than unrounded stimuli.

2.3.3 Discussion

Experiment 2 aimed to investigate the use of fewer stimuli to obtain similar results and effects as in experiment 1, as well as the potential to perform the task on-line as opposed to in laboratory conditions and under supervision. In addition, experiment 2 explored using natural vowel stimuli instead of synthesised vowel stimuli. Hypotheses for this experiment were as follows:

(a) To test whether the task developed in experiment 1 can be adapted to be (i) delivered online, (ii) with fewer stimuli, while retaining the motor imagery effects. The aim concerns the following hypotheses:
    i. Both the visual and auditory task show reaction time effects from experiment 1 remain when fewer stimuli are used;
    ii. Both the visual and auditory task show accuracy effects from experiment 1 remain when fewer stimuli are used;

The overall cross-task comparison found that the visual task consistently showed faster reaction times and higher accuracy compared to the auditory task. Regardless of task, there was a consistently faster reaction time and higher accuracy for the visual task than for the auditory task. These findings concur with those from experiment 1 and suggest that once again participants found the auditory task more difficult than the visual task. These results are not problematic *per se* but do suggest that a comparison should be made within task rather than across to learn more about the differences between easy and hard stimuli. Again, it should be noted that a difference in reaction time was to some degree expected due to the way in which stimuli are shown in each task – the dependent variable of import is accuracy. Across the board, reaction times were slower and standard deviations greater than in experiment 1, and this is likely due to on-line stimulus delivery. Compared to experiment 1, there is consistency with auditory reaction times consistently longer than visual reaction times, as expected.

The visual task showed results consistent with previous literature and experiment 1. As expected, reaction times were longer for difficult stimuli (lateral positions) than for easy stimuli (medial positions), a result that was also seen in the accuracy data, where easy stimuli showed greater accuracy than hard stimuli. These results remain irrespective of which hand showed the position, whether it was a left- or a right-hand stimulus image. Additionally, the lack of interaction effects shows that the rotation is the single factor that explains the differences between the rotation stimuli. As discussed in section 2.2.3, these results can be attributed to motor imagery of hand movement, as they have in the past. These results confirm that the visual task shows robust effects even when the number of stimuli is reduced to eight, and when the task is performed on-line without supervision.

The auditory task showed results that were somewhat different from experiment 1. As expected, reaction times were longer when stimuli were non-native (hard) compared to when stimuli were native (easy). Unlike in experiment 1, there was no evidence that the shape of the vowel played a role in the reaction time changes: whether the vowel was unrounded or rounded, the only factor that influenced reaction times was whether it was a known native vowel or an unknown non-native vowel. This result aligns more closely with the original hypothesis that the shape of the vowel should not play a role in reaction time changes, in the same way that whether a left- or right-hand stimulus image is shown should not play a role. In this sense, the task used in experiment 2 succeeded in providing clearer effects for the novel task, which the previous version with more stimuli had been unable to do.

However, the auditory task accuracy data showed a stimulus-specific issue. A problem can first be spotted in the overall statistics, which showed that the difference between hard and easy stimuli in the tasks is quite different: hard stimuli in the auditory task showed much lower accuracy than hard stimuli in the visual task, and this remains the case when considering them relative to each respective task's easy stimuli. Once again the largest factor by far that influenced accuracy was difficulty, with native stimuli consistently outperforming non-native stimuli. The accuracy data did show that there was an effect of shape, and upon further investigation it seems this main effect was driven by an interaction which clearly showed a large difference in accuracy between unrounded and rounded vowels, but only when the stimuli were non-native; the effect disappeared for native vowels, which were almost always judged accurately regardless of shape. In effect, the problem found in experiment 1 persists, albeit in a reduced fashion: it seems the task itself is simply much harder than the visual task, with differences between hard and easy stimuli quite large indeed.

With regard to learning, experiment 2 showed that learning did occur from the 1st block through to the 3rd block with regard to reaction time: that is, participants did become faster over time, but their accuracy did not improve (see Appendix A). Importantly, the interaction showed that very little learning took place in the case of the visual task, while learning did occur in the auditory task. This is unlike the results in experiment 1, where learning was evident in both tasks. The visual task reaction times, while stagnant, are longer in experiment 2 than in experiment 1, which may be a result of the on-line nature. Certainly, an argument could be made that there is less pressure to follow the instructions to be as quick and accurate as possible when there is no experimenter in the room with the participant. The same could be said for auditory stimuli, but as the analysis shows the task generally seems somewhat harder. In this case, habituation to task and stimuli (rather than actual learning) would also explain why the slope of the curve is steeper in the auditory task. The hypotheses state that learning or habituation is expected to take place at roughly the same rate between visual and auditory task, which is disproven here.

In conclusion, the tasks as used in experiment 2 (reduced number of stimuli & delivered on-line) showed both similarities and differences to the tasks as used in experiment 1. The visual task showed similar effects across the board, with the exception of learning which seemed to take place less in the on-line version, showing that the number of stimuli could be reduced. The auditory task showed similar effects as the previous iteration as well but also found a more complex interaction between the stimuli, with a specific difference for non-native unrounded stimuli. The auditory task results also suggest that the difference between easy and hard stimuli is larger than is the case in the visual task.

As previously discussed (2.1.1), a number of neuroimaging studies have shown the various cortical and subcortical areas active during motor imagery in the visual task. These activations are the foundation of the claim that motor planning processes and simulation processes occur during the visual task. Since the auditory task is novel, no neuroimaging studies have been performed to explore whether these same areas, or different areas, are active also during the auditory imagery task. While the results across tasks show the expected patterns (longer reaction times for stimuli with more complex motor simulation, shorter reaction time for simpler stimuli), there is no neural foundation for motor simulation in the auditory task. Experiment 3 investigated the visual and auditory task, with the goal to replicate previous findings for the visual motor imagery task as well as extend neural activation findings to the auditory imagery task.

Due to the inherent similarities of experiments in chapters 2 and 3, and to avoid repetition, a longer and more in-depth discussion encompassing experiments in both chapters is provided at the end of chapter 3. See this section (3.3) for limitations and future research directions

3 Chapter Three

3.1 Introduction

3.1.1 General introduction

The visual task described in experiments 1 and 2 has a long history of being used to investigate motor imagery behaviourally. In part this is due to the robust behavioural measures which have also been demonstrated in the previous sections, but it is also due to neuroimaging investigations of this task which have shown the use of specific motor-related brain areas. Parsons and colleagues (1995) used positron emission tomography (PET) to show that premotor cortices, parietal regions and cerebellar regions were active during the visual task, which the authors state points to a motor planning, somatosensory and sensorimotor network active during the task. The authors state that this finding is indicative of a motor imagery strategy. These experiments were repeated in different fashions by Kosslyn et al., (1998) and by Vingerhoets et al., (2002) by comparing it to the Shepard-Metzler task, showing similar results. Further studies can be found in Zapparoli et al., (2014), who themselves carry out the visual imagery task with specific attention to differences between left and right hands, and palm view and back view stimuli. In each case, the authors confirm activation in premotor cortices, parietal regions and cerebellum, and Zapparoli et al., (2014) showed that this network is more activated for palm view stimuli than back-of-hand stimuli. The authors reason that the back of the hand is seen more often from an egocentric view and therefore requires less active motor imagery.

As a new task, the brain regions involved in the auditory task have not been evaluated as being supportive or opposing of the hypothesis that similar motor simulation processes occur in the auditory task as in the visual task. While the behavioural effects discussed so far suggest that similar processes may be engaged in both tasks, there is no evidence regarding cortical activation during this task. Experiment 3 aims to investigate the cortical activation behind each task using functional magnetic resonance imaging (fMRI).

In order to better understand the network involved in motor imagery as opposed to simple visual and auditory processing, the fMRI design involved not only the visual and auditory imagery tasks, but also a simple visual and auditory reaction time task. Since both the simple and imagery tasks involved exposure to auditory and visual stimuli, low-level auditory and visual activation related to simple stimulus observation could be statistically accounted for when comparing activated regions. This allowed for an analysis of only those areas that were relevant for motor imagery – either per modality or across modalities. See the methods (3.2.1) for more detail.

3.1.2 Research aims

The aims and hypotheses for experiment 3 incorporate the behavioural aims and hypotheses from chapter 2 (here repeated for convenience, a-b) and expand on them using neuroimaging-specific aims and hypotheses (c) :

(a) To replicate results for the visual task from chapter 2. This aim examines the following hypotheses:
    i. Lateral rotations are expected to result in higher reaction times than medial rotations due to longer motor trajectories;
    ii. No differences between left and right hand stimuli reaction times are expected as there is no inherent difficulty difference between the stimuli types;
    iii. Medial rotations are expected to be judged with significantly lower accuracy than lateral rotations, but accuracy is also expected to be very high (> 90%) (Hyde et al., 2013);

(b) To replicate results for the auditory task from chapter 2. This aim examines the following hypotheses:
    i. Non-native vowels are expected to result in higher reaction times than native vowels, as it takes longer to compute the motor plans required for motor imagery of lesser-known articulations;
    ii. No differences between rounded and unrounded vowels are expected, as there is no inherent difficulty difference between the stimuli types;
    iii. Accuracy is expected to be similarly affected, so a small but significant difference is expected for the auditory task;

(c) To further test the robustness of the task from experiment 2 while also measuring brain activity using fMRI, to test whether visual and auditory tasks show neural activation that suggests a motor simulation strategy. This involves testing whether motor imagery pathways are shared regardless of modality, or whether they are modality-specific. The aim concerns the following hypotheses:
    1. Cortical areas related to motor imagery (M1,SMA, cerebellum) are expected to be more active in motor imagery conditions than in a baseline simple observation condition, regardless of modality – that is, a shared network for motor imagery exists which is active over and above a baseline simple observation condition network, regardless of modality;
    2. Cortical areas related to auditory stimuli (A1, STG) and visual stimuli (V1, fusiform gyrus) are expected to be more active in the auditory and visual conditions respectively – that is, a modality-specific differentiation exists over and above a baseline simple observation condition network, regardless of task;
    3. In addition to a shared motor imagery network, auditory imagery and visual motor imagery processes also show greater activation for each modality in higher-level modality-specific cortices, such as IFG for auditory and V4 for visual imagery conditions;

Note that the number of expected participants for this experiment was initially set at 24. Due to the coronavirus-disease-19 pandemic and the shutdown of testing laboratories across the university, the number of participants tested (17) was lower than expected.

3.2 Experiment 3

3.2.1 Methods

This experiment makes extensive use of fMRI. fMRI is a non-invasive neuroimaging technique that measures changes in cortical activity on the basis of changes in oxygenated blood flow to the brain. MRI makes use of the magnetisation of hydrogen protons in tissue, using a large bore magnetic field. A radiofrequency pulse excites some of these protons, generating transverse magnetisation. Turning off the radiofrequency pulse results in a relaxation phase, allowing hydrogen protons to return to their previous state (Stamatakis et al., 2017). The blood-oxygen-level-dependent (BOLD) response relies on the fact that deoxygenated blood causes interference in the transverse magnetization MR signal, while oxygenated blood leads to an improved signal (Friston et al., 2000; Ogawa et al., 1990). Local neuronal activity requires an increase in glucose and oxygen supply in order to support increased neuronal function (Arthurs & Boniface, 2002). The increase in the oxyhaemoglobin (diamagnetic) to deoxyhaemoglobin (paramagnetic) ratio due to an increased supply of oxygenated blood causes a net improved MR signal, indicating that those areas in which the signal is improved are relevant to ongoing processing. In this way, task-dependent processing in specific brain areas is correlated with greater signal (Huettel et al., 2014).

3.2.1.1 Participants

Seventeen participants took part in experiment 3 (7F, 10M, mean age = 23y 5m, *SD* = 3y 8m, age range = 18-31). All participants were speakers of Standard Southern British English (SSBE) with a minimum of high school education. Four participants were recruited via the university psychology subject pool (SONA), while 13 were recruited via an advertisement page hosted on www.callforparticipants.co.uk. Sixteen participants were right-handed, while one was left-handed. All participants had normal or corrected-to-normal vision. Participants reported no history of neurological or psychiatric disorder. Experiments were undertaken with informed verbal and written consent of each participant, following guidelines set out by the University College London Research Ethics Committee (UREC #0599/001). All research was carried out in accordance with the Declaration of Helsinki.

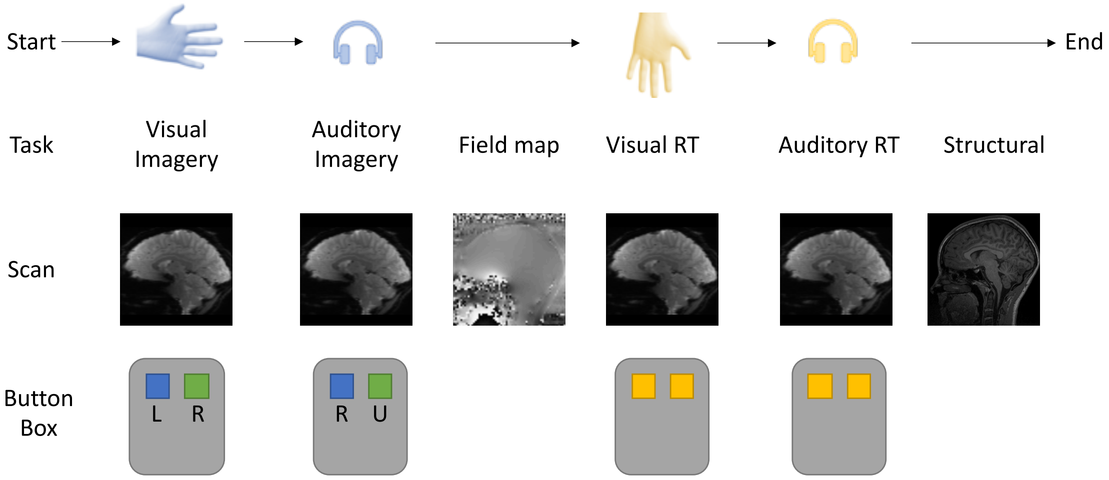


Figure 3.1: fMRI paradigm showing imagery tasks (blue) and simple reaction time tasks (yellow), relevant button boxes and scan types. The buttons were assigned either left/right, rounded/unrounded or non-choice, depending on the task. Tasks were counterbalanced across participants.

3.2.1.2 Materials

Materials were the same as those used for experiment 2. Stimuli were displayed via a projector onto a 13" in-bore screen in the fMRI scanner, and the participant's viewing distance was around ~20cm away from the screen (see Figure 3.1). Auditory stimuli were presented through Etymotic ER2 headphones at 70dB SPL. As foot switches were not an option in the scanner environment, a custom-made MRI-safe button box was used to collect response data instead of foot switches or arrow keys: one button was pressed by the right hand index finger for a left response, and another button was pressed by the right hand middle finger for a right response. This setup was the same regardless of handedness. An out-bore camera was used to check that participants did not use overt hand or lip movements during the task.

The fMRI design incorporated a simple visual and auditory task. Two reaction time tasks (one auditory observation task and one visual observation task) were created against which the imagery task could be compared in terms of brain activation in the factorial analysis of the data. The reaction time tasks made use of the same stimuli as the respective visual and auditory task and required the participants to press either button as soon as they saw the stimulus (visual task) or heard the stimulus (auditory task). In this way, cortical activation could be compared not only to an implicit rest baseline, but also to a task that required similar low-level visual and auditory processing, and similar motor activation for responses.

The fMRI experiment comprised 4 functional runs in an event-related design, a field map scan and a structural scan (see Figure 3.1). Each functional run consisted of 1 task with periods of resting baseline interspersed. Functional runs were counterbalanced across participants. The total time per functional run did not exceed 7 minutes, and stimuli were presented in 4 blocks of 8 stimuli, with a 15s rest period between blocks. Each functional run consisted of 32 trials in total, with reaction time and accuracy logged per trial for use in event-related analysis.

3.2.1.3 Procedure

Participants were welcomed at the scanning centre and the study was explained. Participants signed an informed consent form and the preparatory fMRI background screening forms. A practice task was performed outside the scanner which showed each stimulus at least once, and which also provided feedback as to whether their responses were correct. Participants were told they would not receive such feedback during the actual task, and that they should refrain from moving their hands or lips. Following this preparatory task, participants were taken through the procedure preparing them for the fMRI session during which all metal was removed. Participants were then taken into the scanner suite and instructed to lie in a comfortable position on the scanner bed. Once comfortable, participants were told the order of conditions (specific to them) and were made aware of the number of scanning runs. Earphones were inserted and visual checks performed to ensure auditory and visual stimuli were clear to the participant. Participants were given the button box and pressed the response buttons several times to check responses and familiarise participants with the setup. Participants completed two functional runs, after which a field map scan was acquired. Two more functional runs followed, with the structural scan as the last scan condition. The participant was thanked and invited to the follow-up session one week later, where they were in all cases paid for their time.

3.2.1.3.1 fMRI data acquisition

Functional MRI data was acquired using a 3T Siemens PRISMA scanner equipped with gradient-echo echo-planar imaging (EPI, flip angle 65°, TE = 35.20ms, TR = 1500ms, FOV = 212 x 212mm, slice thickness = 2mm, slices = 72, voxel size = 2 x 2 x 2mm) using a 32-channel head coil in an event-related design. Volumes were acquired in a multiband series (interleaved, ascending, multiband accel. factor 4) with a field map acquired for each participant (TE_{short} = 10.00ms, TE_{long} = 12.46ms, TR = 1020ms, phase encoding R/L, FOV = 192 x 192mm) at the end of each session. The number of volumes was dependent on reaction time of each participant but did not exceed 300. The first 4 volumes of each run were manually discarded to account for T1 longitudinal effects. A structural scan was also acquired (MPRAGE - GRAPPA, flip angle = 9°, TE = 2.98ms, TR = 2300ms, FOV = 256 x 256mm, slice thickness = 1mm, slices = 208, voxel size = 1 x 1 x 1mm) for use in unwarping and overlaying of activation maps per participant.

3.2.1.4 Analysis

The analysis of the behavioural data was largely identical to that in experiment 1 and 2 (see section 2.2.1.5): a primary 2 x 2 repeated-measures ANOVA with levels Task (auditory, visual) and Difficulty (easy, hard) followed by a secondary 2x 2 ANOVA with levels Difficulty (easy, hard) and Hand (left, right) or Shape (rounded, unrounded). Since the fMRI task included reaction time baseline tasks, an additional analysis was carried out on the reaction time tasks with regard to overall statistics (paired-samples *t*-test comparing visual and auditory reaction times) and learning across blocks (2 x 4 ANOVA). The results of this last analysis can be consulted in Appendix A.

3.2.1.4.1 fMRI data analysis

Functional MRI images were checked for abnormalities and converted from DICOM into NIFTI format using in-house scripts. Subsequent data analysis was performed using the software Statistical Parametric Mapping SPM 12 (SPM, Wellcome Trust, London) in Matlab R2019b (The Mathworks Inc., Massachusetts, USA). Phase and magnitude field map scans for each participant were converted into voxel displacement maps (VDMs) using the built-in FieldMap Toolbox (2.1). VDMs were subsequently used to realign and unwarp the functional EPI volumes (Ashburner et al., 2016; Hutton et al., 2002). Slice time correction was not performed as previous analysis has shown there to be little

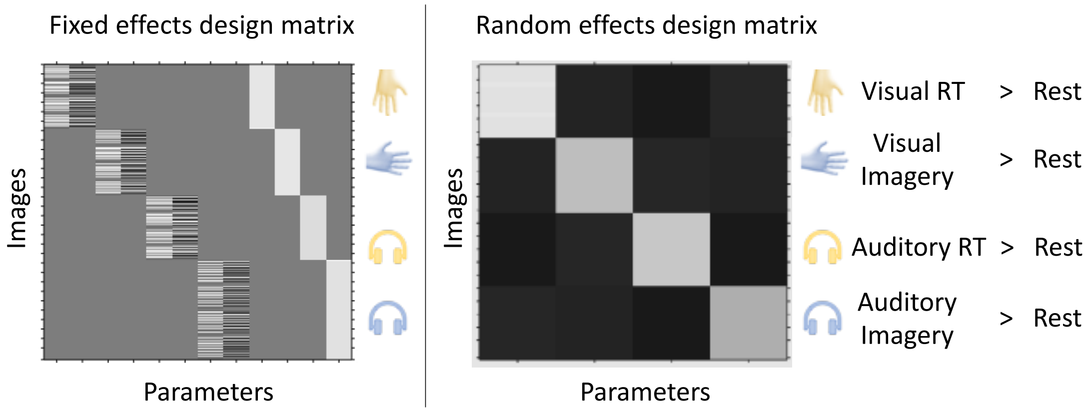


Figure 3.2: Design matrices for fixed and random effects analyses in experiment 3.

beneficial effect with TRs below 2s and is very complex with multiband acceleration in place (Parker et al., 2017; Poldrack et al., 2011). The realigned images were normalised into Montreal Neurological Institute (MNI) space using the built-in SPM Tissue Probability Map template and using the mean unwarped image to allow for analysis at the group level. EPI volumes were smoothed using an 8 x 8 x 8mm full-width-half-maximum Gaussian filter.

The haemodynamic response function was modelled as a canonical HRF + temporal derivatives, convolved with the trial onsets to generate explanatory variables used in the general linear model (Ashburner et al., 2016). High-pass filtering (1/128s) was employed in the fixed-effects analyses to reduce the influence of low-frequency noise. The fixed effects (per participant) analysis involved creating contrast vectors comparing activation for each condition to an implied rest baseline (e.g., visual reaction time > rest has the contrast vector [1 0 0 0], visual imagery > rest [0 1 0 0], auditory reaction time > rest [0 0 1 0] and auditory imagery > rest [0 0 0 1], temporal derivatives not included, see Figure 3.2). A separate fixed-effects analysis was carried out to obtain contrasts images for difficulty levels for the imagery tasks (e.g. visual imagery easy > rest, visual imagery hard > rest, auditory imagery easy > rest, auditory imagery hard) as well as laterality and roundedness for visual and auditory conditions respectively.

To compare effects of conditions across participants, random effects analyses using factorial 2 x 2 ANOVAs with factors Task (visual and auditory) and Response Type (imagery and reaction time) were implemented using SPM (see Figure 3.2). Again, a separate 2 x 2 ANOVA was carried out to investigate Difficulty (easy and hard), Hand (left and right) in the visual imagery condition and Shape (rounded vs unrounded) in the auditory imagery condition. At this lowest level, each participant had 8 responses per factor per level.

For each random-effects analysis, contrasts were inclusively masked with each condition > rest as well as related conditions where relevant to ensure that activations constituted main effects rather than interactions. These are indicated for each comparison in the results section below. Masks were used where relevant at the uncorrected level of $p < 0.05$. All results presented are significant at the level of $p < 0.05$ (FWE corrected). Activations were visualised in three-dimensional figures using BrainNet viewer (http://www.nitrc.org/projects/bnv/, Xia et al., 2013) while figures showing slices with parametric overlays were generated using the Multi-Image Analysis GUI (Mango, Lancaster and Martinez, 2007). Contrast images, maps and parametric maps were obtained through SPM.

3.2.2 Results

3.2.2.1 VISQ

In the case of this experiment, only a subset of 11 (out of 17) participants filled in the VISQ and MOCA as these questionnaires were generally filled in during a follow-up session related to a separate TMS session, which not all participants attended, either due to no longer wishing to take part in the TMS session, or due to coronavirus circumstances. The VISQ results are described here for completeness.

Overall, participants reported a high level of experience of dialogic inner speech and evaluative/motivational inner speech compared to other types of inner speech (see Figure 3.3). Generally, participants reported more expanded inner speech than condensed inner speech (if any). The scores for other people in inner speech were reported across the spectrum.

3.2.2.2 Overall (combined) statistics

Overall statistics for all tasks combined showed that out of 2176 trials, 41 trials had reaction time values that represented out-of-bounds values or exceeded 3SDs from the mean log10 reaction times, yielding 2135 trials for analysis. No further individual trials were excluded on the basis of accuracy, since imagery may take place regardless of whether the answer provided is correct or not.

3.2.2.2.1 Overall statistics: Reaction Time Tasks

Mean raw reaction time across reaction time tasks was 501ms (*SD* = 311ms). Mean raw reaction time for the visual reaction time task was 403ms (*SD* = 143ms) while the mean raw reaction time for the auditory reaction time task was 600ms (*SD* = 398ms). For further analysis the log10 reaction time data was used to counteract skew in the raw reaction time data (raw reaction time skewness: 2.484; log10 skewness: 1.279).

A paired-samples *t*-test was performed to compare the reaction times in the auditory reaction time task and the visual reaction time task. The *t*-test showed that the aforementioned difference between the two tasks was statistically significant (*t*(16) = 3.33, *p* = .004, Hedges' *g* = -0.672), see Figure 3.4.

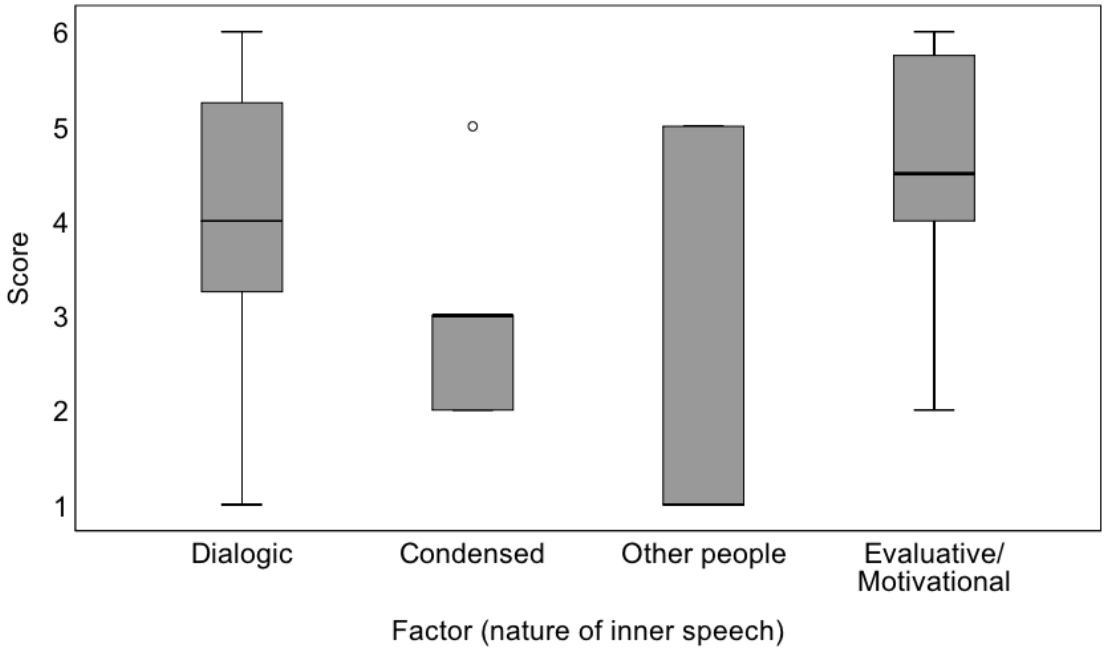


Figure 3.3: VISQ scores per type of inner speech for experiment 3. Scores are amalgamated across five questions that relate to each factor. See Appendix C for the questionnaire. (1 = Certainly does not apply, 6 = Certainly does apply).

3.2.2.2.2 Overall statistics: Imagery Tasks - Reaction Time

Data were aggregated by Task (auditory vs visual task) and Difficulty (hard vs easy) and a repeated-measures ANOVA was performed to ascertain whether the tasks and difficulty levels were comparable based on reaction time data, as for experiments 1 and 2.

Mean raw reaction time for the auditory task was 1437ms (*SD* = 351ms) while mean raw reaction time for the visual task was 1306ms (*SD* = 380ms). These results translate to similar mean log10 reaction times of 3.127 (*SD* = 0.093) for the auditory task and mean log10 reaction times of 3.073 (*SD* = 0.137) for the visual task. As opposed to the previous experiments, there was no main effect of Task ($F(1,16) = 2.26$, $p = .152$, $\eta_p^2 = 0.124$) since reaction times between the visual (*M* = 1307ms, *SD* = 356ms) and auditory task (*M* = 1438ms, *SD* = 274ms) were roughly similar, see Figure 3.4. However, as before there was a main effect of Difficulty ($F(1,16) = 32.77$, $p < .001$, $\eta_p^2 = .672$, Hedges' $g = 0.767$) as easy stimuli (*M* = 1267ms, *SD* = 241ms) showed shorter reaction times than hard stimuli (*M* = 1478ms, *SD* = 279ms, see Figure 3.5). There was no significant interaction Task x Difficulty ($F(1,16) = 0.38$, $p = .547$, $\eta_p^2 = .023$). These results suggest that the main factor influencing reaction time is difficulty of the stimulus.

3.2.2.2.3 Overall statistics: Imagery Tasks - Accuracy

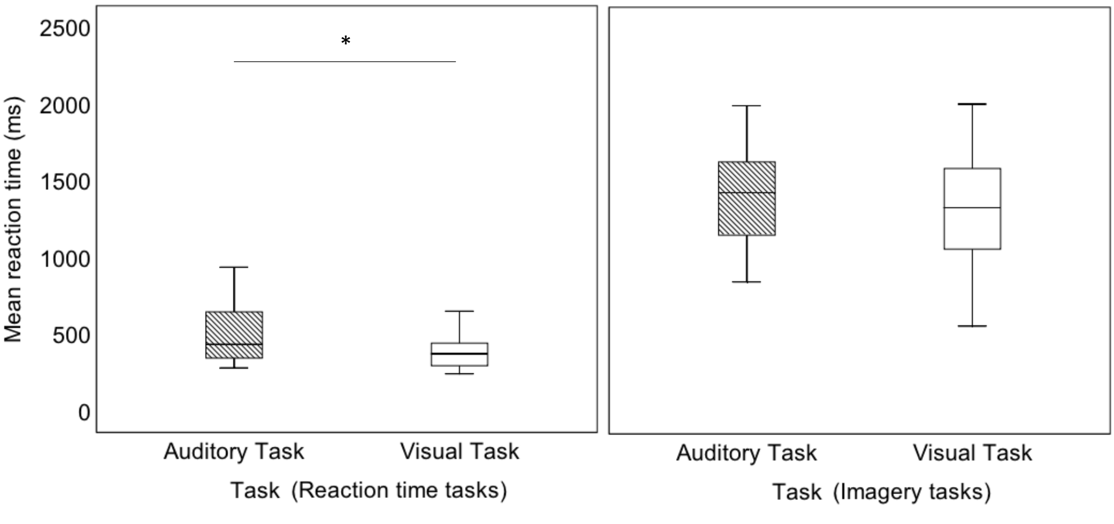


Figure 3.4: Main effect of Task for reaction time in experiment 3. Pure reaction time tasks results (left) show a significant main effect, which is lacking for the imagery tasks (right). * = significant at $p = .05$.

Data were aggregated by Task (auditory vs visual task) and Difficulty (hard vs easy) and a repeated-measures ANOVA was performed to ascertain whether the tasks and difficulty levels were comparable based on accuracy data.

There was a main effect of Task (F(1,16) = 15.55, p = 001, η_p^2 = .493, Hedges' g = -1.126) due to the fact that the visual task (M = 90.4%, SD = 10.7%) showed greater accuracy scores than the auditory task (M = 78.5%, SD = 9.9%). There was also a main effect of Difficulty (F(1,16) = 61.15, p < .001, η_p^2 = .793, Hedges' g = -2.042), since overall easy stimuli (M = 94.6%, SD = 7.4%) showed greater accuracy than hard stimuli (M = 74.3%, SD = 11.5%, see Figure 3.5). The accuracy data also showed a very strong interaction (F(1,16) = 75.85, p < .001, η_p^2 = .826). A similar interaction was found in experiments 1 and 2, so follow-up analyses were performed to investigate if the interaction showed similar trends.

Follow-up t-tests showed that there was a large difference between the easy and difficult stimuli in the auditory task (MD = 33.9%, SD = 13.5%, t(16) = 10.37, p < .001, Hedges' g = -2.796), while there was a smaller difference between the easy and difficult stimuli in the visual task (MD = 6.7%, SD = 11.4%, t(16)= 2.42, p = .028, Hedges' g = -0.535). However, it was clear that the relevant difference could be found with the hard stimuli between tasks: there was a large difference between the higher accuracy for difficult visual stimuli and the lower accuracy for difficult auditory stimuli (MD = 25.5%, SD = 17.4%, t(16) = 6.05, p < .001, Hedges' g = -1.704) compared to the difference across the easy stimuli (MD = -1.7%, SD = 9.4%, t(16) = -0.77 p = .451, Hedges' g = 0.195, see Figure 3.6). This effect conforms to the interactions found in experiments 1 and 2, and shows that although

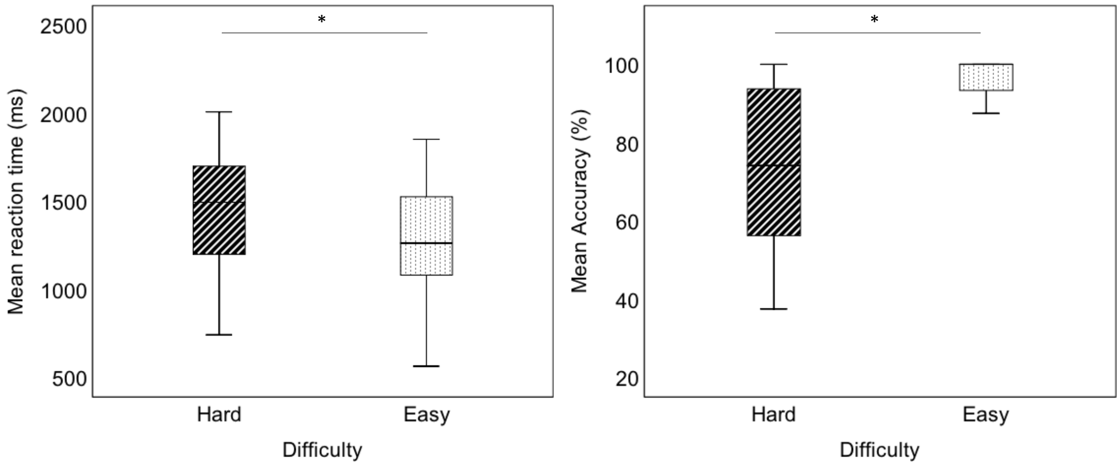


Figure 3.5: Main effects of Difficulty in reaction time and accuracy results for experiment 3 (imagery tasks). * = significant at p = .05.

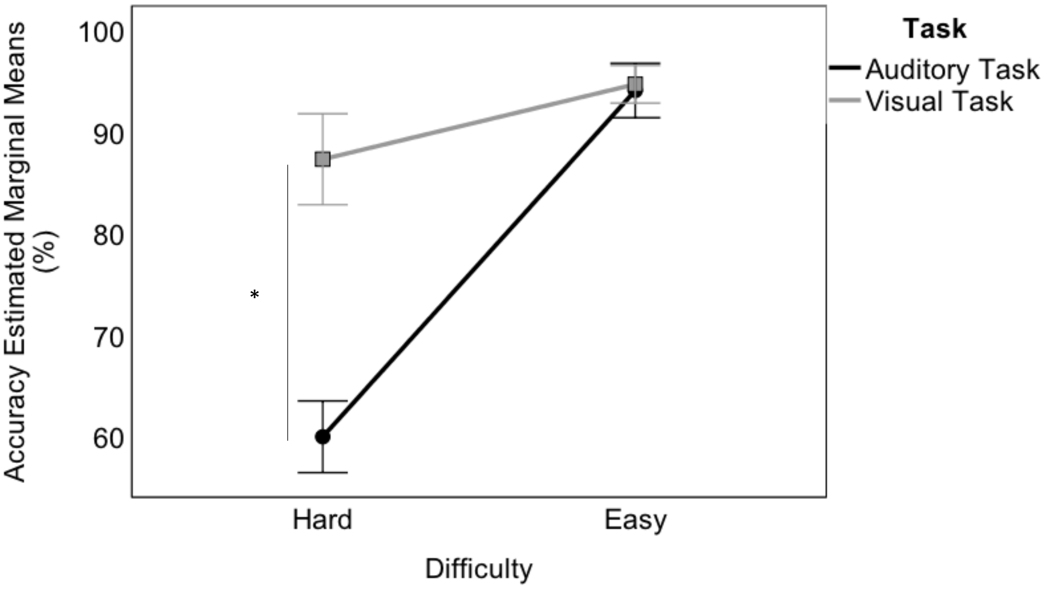


Figure 3.6: Interaction effect of Task and Difficulty across tasks in experiment 3. Error bars indicate +-1 SEM. * = significant at $p = .05$

Difficulty on its own is a primary factor in reaction time and accuracy scores, when accuracy in particular is concerned Task has a large impact on the difference in scores between easy and hard stimuli: the visual task showed small differences depending on task, whereas the auditory task showed large differences depending on task.

3.2.2.3 Visual task

3.2.2.3.1 Visual task: Reaction Time

To investigate the differences in reaction time in the visual task, a repeated-measures ANOVA with the factors Difficulty (hard vs easy) and Hand (left hand vs right hand) was conducted. The results showed that there was a significant main effect of Difficulty ($F(1,16) = 17.15$, $p = .001$, $\eta_p^2 = .517$, Hedges' $g = 0.424$), showing that responses to easy (medial) stimuli ($M = 1238$ms, $SD = 366$ms) were significantly faster than responses to hard (lateral) stimuli ($M = 1376$ms, $SD = 372$ms). There was no main effect of Hand ($F(1,16) = 0.13$, $p = .726$, $\eta_p^2 = .008$). Similarly, there was no interaction Difficulty x Hand ($F(1,16) = 1.09$, $p = .312$, $\eta_p^2 = .064$). These results suggest that the rotation of the stimulus (lateral vs medial) was the main factor resulting in reaction time differences, irrespective of which hand showed the rotation (see Figure 3.7).

3.2.2.3.2 Visual task: Accuracy

A repeated-measures ANOVA was performed to investigate accuracy differences using the factors Difficulty (hard vs easy) and Hand (left vs right hand). The results showed a main effect of Difficulty (*F*(1,16) = 5.78, *p* = .029, η_p^2 = .265, Hedges' *g* = -0.586) showing that responses to easy (medial) stimuli (*M* = 93.7%, *SD* = 8.7%) were significantly more accurate than responses to hard (lateral) stimuli (*M* = 87.0%, *SD* = 14.8%, see Figure 3.7). There was no effect of Hand (*F*(1,16) = 0.78, *p* = .389, η_p^2 = .047) as responses to left hand stimuli (*M* = 89.3% , *SD* = 13.2%) were no more or less accurate than responses to right hand stimuli (*M* = 91.4%, *SD* = 10.3%). There was no interaction Difficulty x Hand (*F*(1,16) < 0.01, *p* = .943, η_p^2 < .001), showing that once again, difficulty was the defining factor for accuracy scores, regardless of the hand which was used to show the rotation. These findings were in line with those found in experiments 1 and 2.

3.2.2.4 Auditory task

As before, the auditory task featured counterbalancing across participants, so a one-way ANOVA was performed to check whether there was a difference between groups that had the rounded and unrounded choice assigned to the left and right buttons on the buttons box, and vice versa. The ANOVA showed that there was no difference of reaction time between groups (*F*(1,15) = 1.06, *p* = .320, η_p^2 = .066), with those who used the former configuration responding no faster (*M* =1511ms, *SD* = 324ms) than those who used the latter configuration (*M* = 1345ms, *SD* = 211ms).

3.2.2.4.1 Auditory task: Reaction time

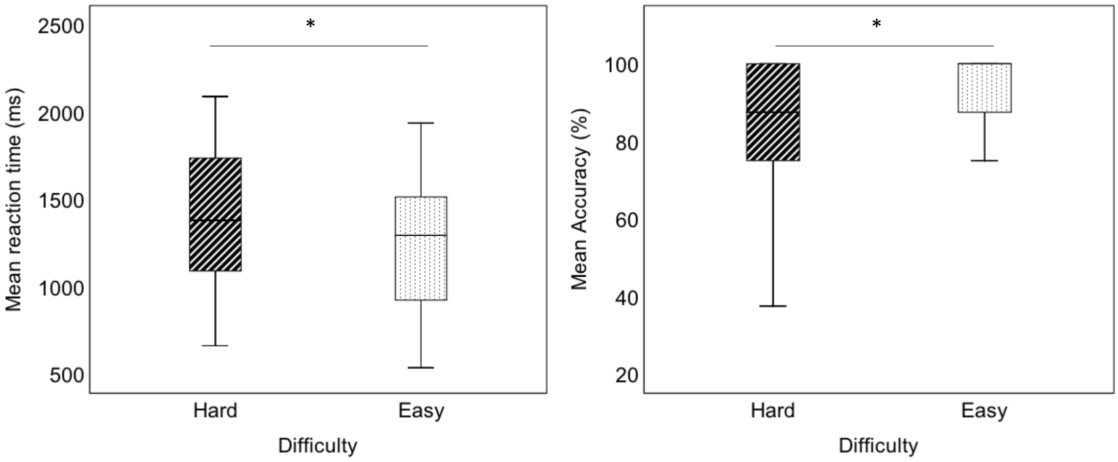


Figure 3.7: Main effects of Difficulty for experiment 3 (visual imagery tasks). * = significant at. *p* = 0.05

To investigate differences in reaction time in the auditory task, a repeated-measures ANOVA with the factors Difficulty (hard vs easy) and Shape (rounded vs unrounded vowel) was performed. The results showed that there was a main effect of Difficulty ($F(1,16) = 14.38$, $p = .002$, $\eta_p^2 = .473$, Hedges' $g = 0.780$), as easy stimuli (M = 1296ms, SD = 232ms) showed shorter reaction times than hard stimuli (M =1578ms, SD = 378ms, see Figure 3.8). However, there was no effect of Shape ($F(1,16) = 0.33$, $p = .573$, $\eta_p^2 = .020$) as there was almost no difference between rounded (M = 1452ms, SD = 329ms) and unrounded stimuli (M = 1422ms, SD = 247ms). Similarly, there was no interaction (F(1,16) = 1.955, $p =.181$, $\eta_p^2 = .109$). These results conformed to those found in experiment 2 but differed from the results in experiment 1.

3.2.2.4.2 Auditory tasks: Accuracy

Differences in accuracy were investigated using a repeated-measures ANOVA with factors Difficulty (hard vs easy) and Shape (rounded vs unrounded). There was a main effect of Difficulty ($F(1,16) = 102.4$, $p < .001$, $\eta_p^2 = .865$, Hedges' $g = -2.779$) due to easy stimuli (M = 95.4%, SD = 9.1%) showing much greater accuracy than difficult stimuli (M = 61.1%, SD = 14.4%). Figure 3.8 shows a clear ceiling effect for easy vowels. Unlike experiment 2, there was no main effect of Shape ($F(1,16) = 1.64$, $p = .218$, $\eta_p^2 = .093$) nor was there a interaction effect ($F(1,16) = 1.58$, $p = .227$, $\eta_p^2 = .090$).

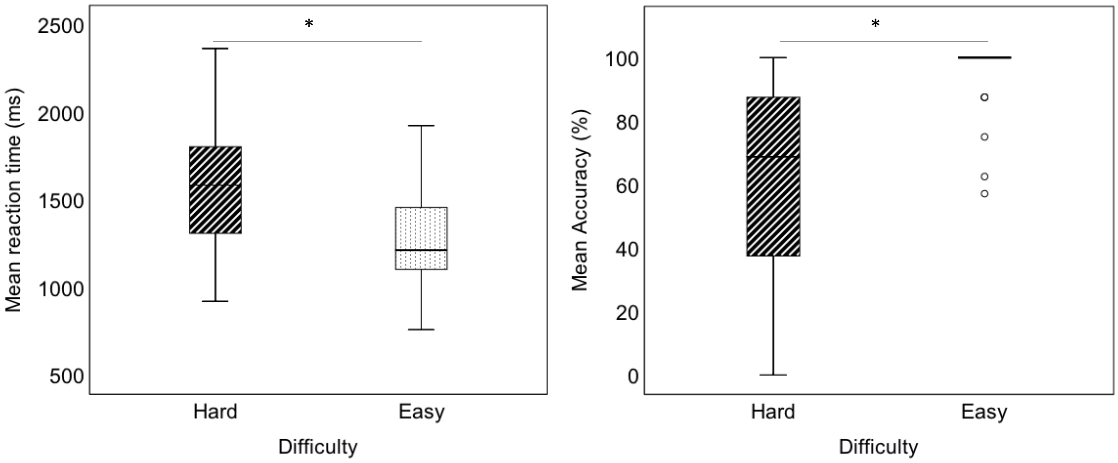


Figure 3.8: Main effects of Difficulty for reaction time and accuracy in the auditory imagery task in experiment 3. * = significant at $p = 0.05$

3.2.3 fMRI Results

3.2.3.1 Task effects: Task (visual, auditory) x Response Type (imagery, reaction time)

A random effects analysis using contrasts visual reaction time > rest, visual imagery > rest, auditory reaction time > rest and auditory imagery > rest was carried out to test for main effects and interactions using a 2 x 2 ANOVA with levels Task (visual, auditory) and Response Type (imagery, reaction time) There was a main effect of Task (visual, auditory) as well as a main effect of Response Type (imagery, reaction time) but there was no interaction effect. The main effects are discussed below.

3.2.3.1.1 Main effect of Task

This experiment involved two different tasks with respect to modality: in one task participants responded to visual stimuli (hands on a screen) while in the other, they responded to auditory stimuli (presented in-ear). The factorial analysis showed that a network of auditory and visual processing areas was active during these tasks. Post-hoc analysis revealed that the two tasks differentially activated these areas (see Table 3.1 and Figure 3.9, all table values and figures computed using an inclusive mask comprised of contrasts of each condition (visual reaction time, visual imagery, auditory reaction time, auditory imagery) > rest. For Auditory > Visual, mask additionally included auditory reaction time > visual reaction time and auditory imagery > visual imagery. For Visual >

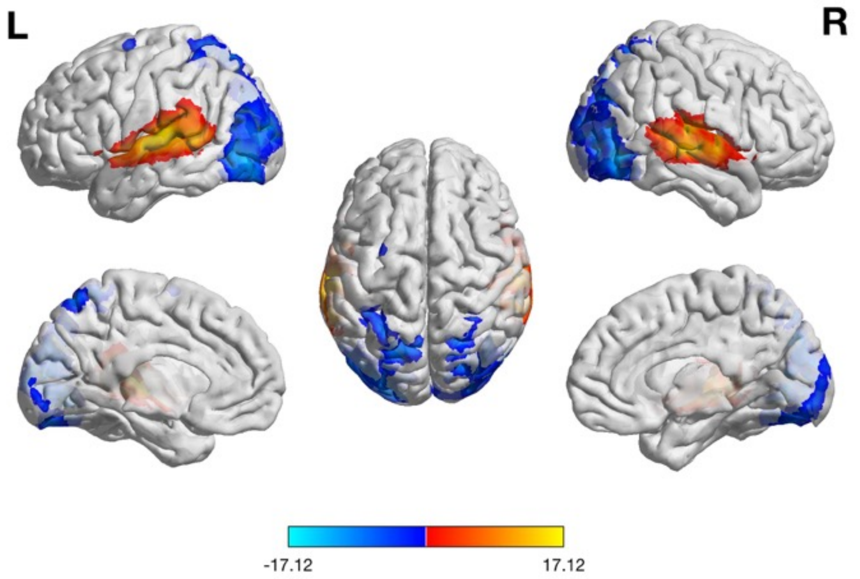


Figure 3.9: Contrasts showing activation in the auditory task (red-gold) relative to the visual task (blue-turquoise). All activation $p < 0.05$ FWE, height threshold $Z = 4.51$, extent threshold $k = 6$ voxels.

Table 3.1: Regions showing main effects of Task. ** = significant at the cluster level ($p < 0.05$ FWE), * = significant at the peak level ($p < 0.05$ FWE). Height threshold $Z = 4.51$, extent threshold $k = 0$ voxels, min. cluster size (FWEc = 6).

Brain region (BA)	MNI coordinates			*Z*-score	K_E	
	x	y	z			
Auditory > Visual						
R Superior Temporal Gyrus (A1)	68	-28	8	Inf.	3856	**
	62	-22	2	Inf.		*
	58	-2	-4	Inf.		*
L Superior Temporal Gyrus (A1)	-62	-24	4	Inf.	3353	**
	-56	-14	2	Inf.		*
L Middle Temporal Gyrus	-56	-46	10	7.21		*
Visual > Auditory						
L Fusiform Gyrus	-46	-76	-8	7.33		**
R Fusiform Gyrus	46	-66	-12	7.03		*
L Mid Occipital Gyrus	-22	-94	10	6.84		*
	-32	-84	4	5.91		*
R Mid Occipital Gyrus	24	-94	14	7.05		*
R Inferior Occipital Gyrus	50	-72	-2	7.12		**

Auditory, mask additionally included visual reaction time > auditory reaction time and visual imagery > auditory imagery). As expected, the auditory tasks activated canonical auditory regions such as STG and A1, whereas the visual tasks activated higher-level visual processing areas in the occipital lobe.

3.2.3.1.2 Main effect of Response Type

The factorial analysis showed a main effect of Response Type, with imagery tasks and reaction time tasks activating substantially different regions (see Table 3.2 and Figure 3.10, all table values and figures computed using an inclusive mask comprised of contrasts of each condition (visual reaction time, visual imagery, auditory reaction time, auditory imagery) > rest. For Imagery > Reaction time, mask additionally included auditory imagery > auditory reaction time and visual imagery > visual reaction time. For Reaction time > Imagery, mask additionally included auditory reaction time > auditory imagery and visual reaction time > visual imagery). The imagery network is far larger and involves a number of frontal areas related to motor processing. Several clusters were found showing activation in the imagery task above that shown during the reaction time

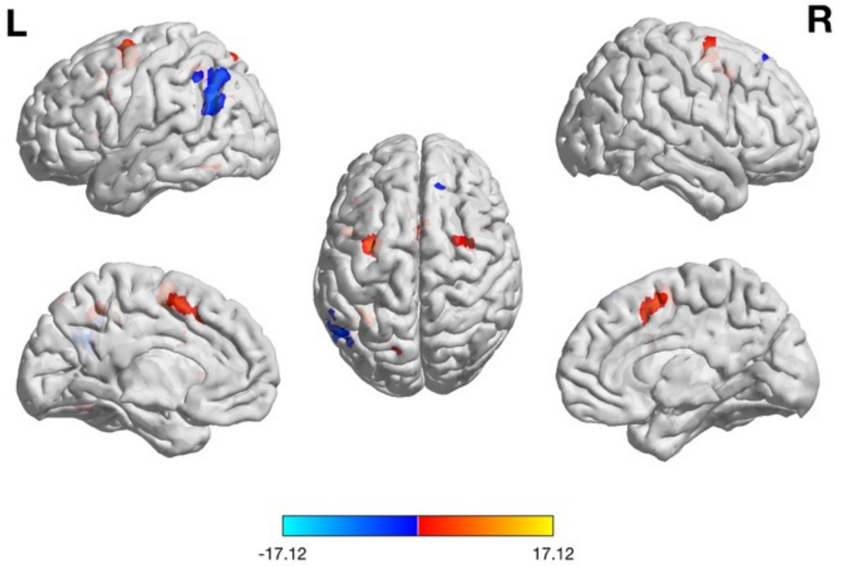


Figure 3.10: Contrasts showing activation in the imagery task (red-gold) relative to the reaction time task (blue-turquoise). All activation thresholded at $Z > 4.51$, representing activation with $p < 0.05$ (FWE) with extent threshold k=6 voxels.

tasks including superior parietal lobule (SPL), likely due to its involvement in visuo-motor coordination and proprioception. Additionally, there was increased activation in pre-supplementary motor area and premotor cortices bilaterally (see Figure 3.11). There was also consistent activation in a small premotor area in IFG and frontal eye fields (FEF) bilaterally, and further activation was noted in insula bilaterally (see Figure 3.12). Notably, cerebellar activation did not differ significantly between the imagery and reaction time tasks.

Table 3.2: Regions showing main effects of Response Type ** = significant at the cluster level ($p < 0.05$ FWE), * = significant at the peak level ($p < 0.05$ FWE). Height threshold $Z = 4.51$, extent threshold $k = 0$ voxels, min. cluster size (FWEc = 6).

Brain region (BA)	MNI coordinates			Z-score	K_E	
	x	y	z			
Imagery > RT (masked by inclusive visual/sound imagery > rest)						
L Medial Frontal Gyrus (pre-SMA /SMA)	-6	10	50	6.83	382	**
R Medial Frontal Gyrus (pre-SMA /SMA)	10	14	48	5.58		*
L Mid Frontal Gyrus (premotor /FEF)	-30	-4	62	5.07		*
R Superior Frontal Gyrus (premotor /FEF)	28	2	68	4.71		*
L Precentral Gyrus (FEF)	-28	-4	54	5.4	296	**
R Precentral Gyrus (FEF)	26	-4	48	5.27		*
L Inferior Frontal Gyrus (premotor)	-44	6	30	5.48	80	**
R Inferior Frontal Gyrus (premotor)	46	8	28	5	36	**
L Superior Parietal Lobule	-30	-48	46	6.03	404	**
L Inferior Parietal Lobule	-32	-38	42	5.44		*
L Insula	-28	20	2	5.15	50	**
R Insula	30	22	4	4.85	32	**
RT > Imagery (masked by inclusive visual/sound RT > rest)						
R Superior Frontal Gyrus	14	36	54	4.81	8	**
L Angular Gyrus	-56	-60	32	5.40	190	**
	-42	-62	26	4.78		*

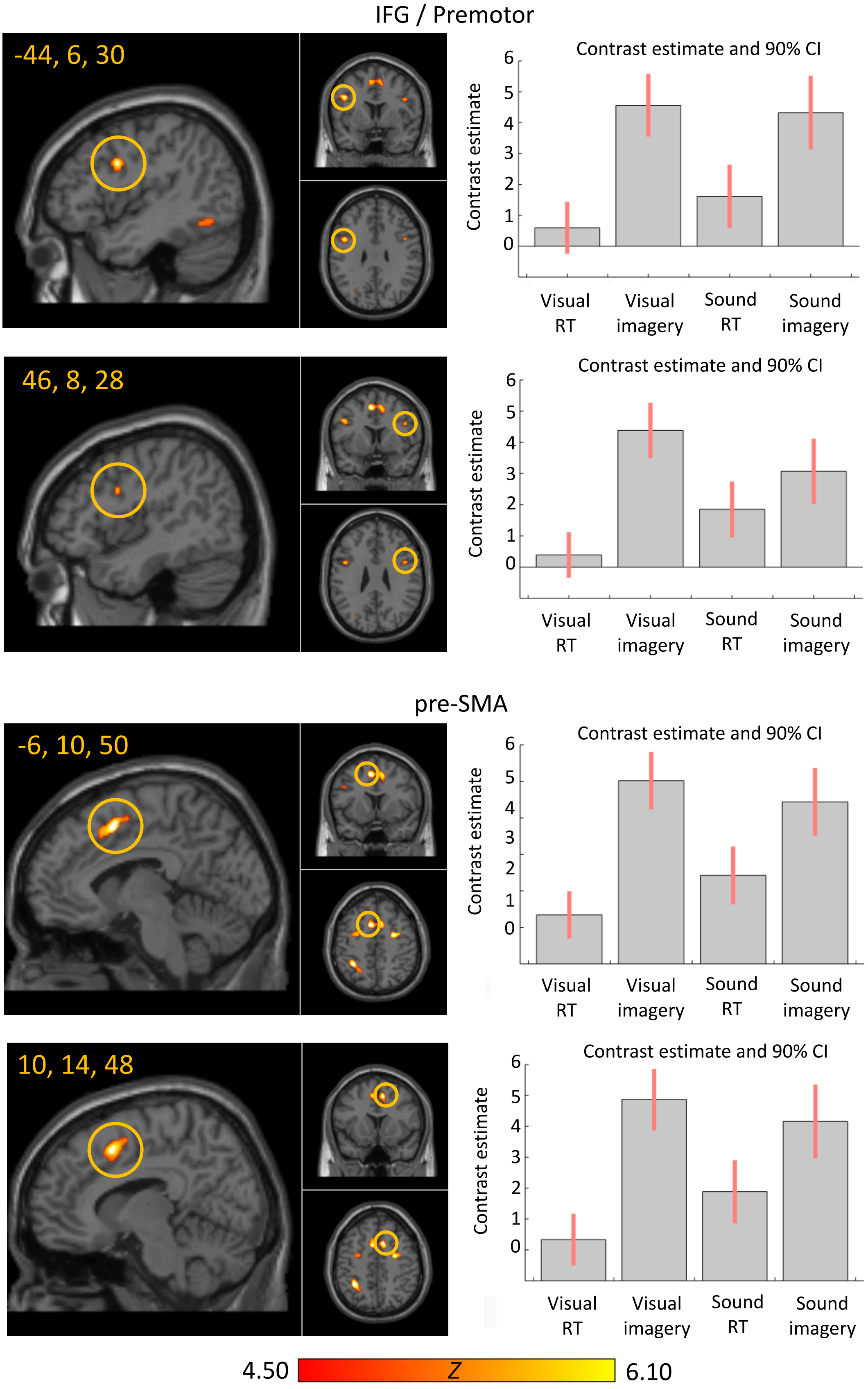
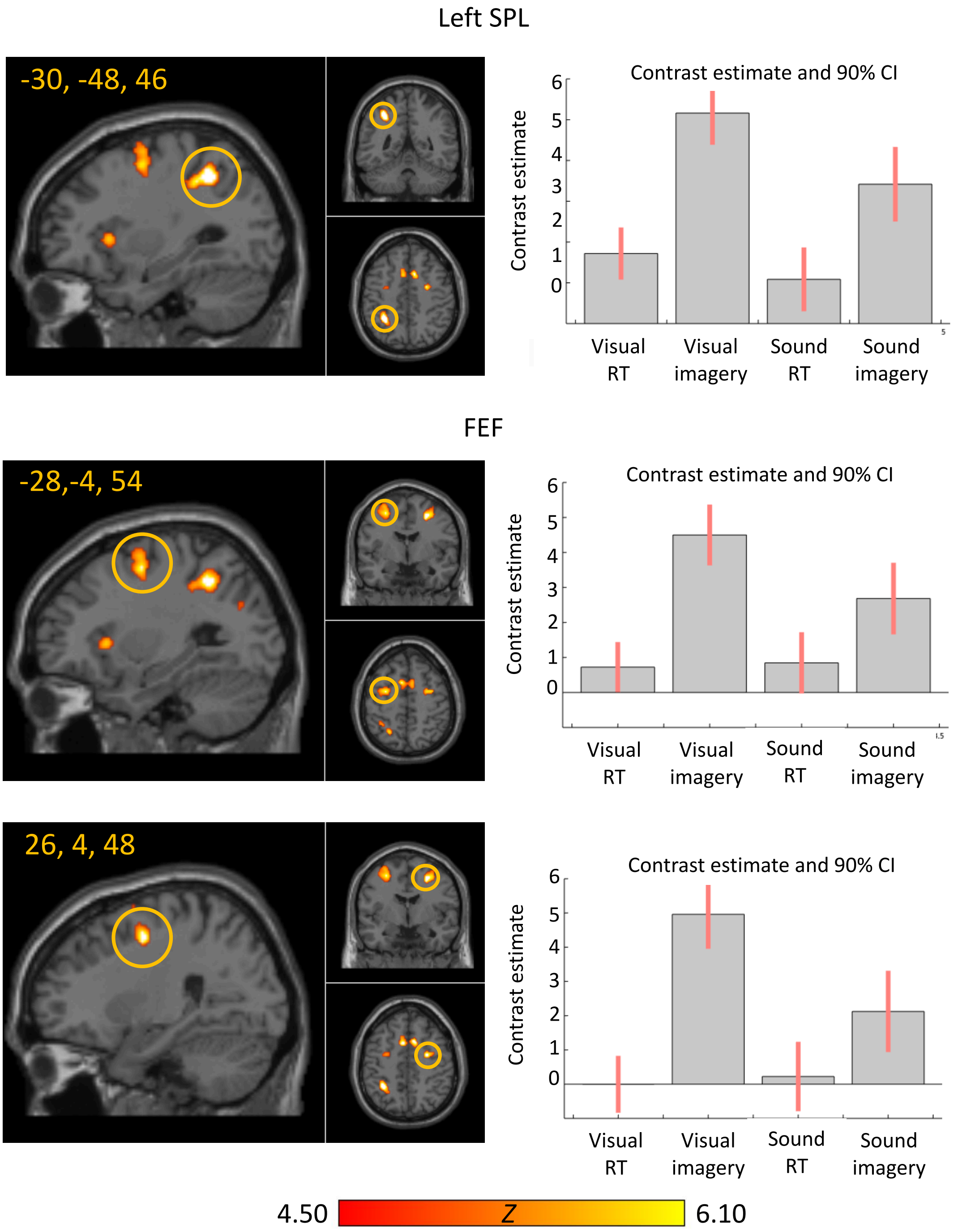


Figure 3.11: Activation in IFG/premotor regions and pre-SMA in experiment 3. Activation shown is thresholded (Z = 4.51, $p < 0.05$ FWE corrected) statistical parametric map of imagery > reaction time masked inclusively (mask uncorrected $p < 0.05$) by visual imagery > visual reaction and auditory imagery > auditory reaction time. Error bars indicate 90% confidence interval.


Left SPL

FEF

Figure 3.12: Activation in Left SPL and FEF in experiment 3. Activation shown is thresholded ($Z = 4.51$, $p < 0.05$ FWE corrected) statistical parametric map of imagery > reaction time masked inclusively (mask uncorrected $p < 0.05$) by visual imagery > visual reaction time and auditory imagery > auditory reaction time. Error bars indicate 90% confidence interval.

3.2.3.2 Imagery-only Task effects (Task x Difficulty)

The fMRI data from the imagery condition was subjected to a 2 x 2 ANOVA with factors Task (visual, auditory) and Difficulty (easy, hard). As before, there was a main effect of Task, with the visual task showing activation in visual areas, and the auditory task showing activation in auditory regions, see Table 3.3 and Figure 3.13. However, there was no effect of Difficulty, nor was there an interaction Task x Difficulty.

Table 3.3: Regions showing main effects of Task. ** = significant at the cluster level ($p < 0.05$ FWE), * = significant at the peak level ($p < 0.05$ FWE). Height threshold $Z = 4.51$, extent threshold $k = 0$ voxels, min. cluster size (FWEc = 6).

Brain region (BA)	MNI coordinates			*Z*-score	K_E	
	x	y	z			
Auditory > Visual						
R Superior Temporal Gyrus (A1)	64	-24	2	Inf.	3091	**
	62	-16	0	Inf.		*
R Superior Temporal Gyrus (A1)	54	-14	2	7.61		*
L Superior Temporal Gyrus (A1)	-64	-30	10	Inf.	2761	**
	-54	-14	2	Inf.		*
	-60	-22	2	7.68		*
Visual >Auditory						
L Middle Occipital Gyrus	-26	-92	10	7.33	1744	**
	-28	-76	28	4.99	42	**
L Fusiform Gyrus	-46	-74	-8	7.32		*
L Cuneus	-16	-98	8	6.53		*
R Fusiform Gyrus	46	-66	-14	6.39	1722	*
	52	-66	-8	6.11		*
R Middle Occipital Gyrus	24	-94	14	6.31		*
R Precuneus	32	-74	32	5.09	101	**
L Precuneus	-20	-60	54	4.97	13	**

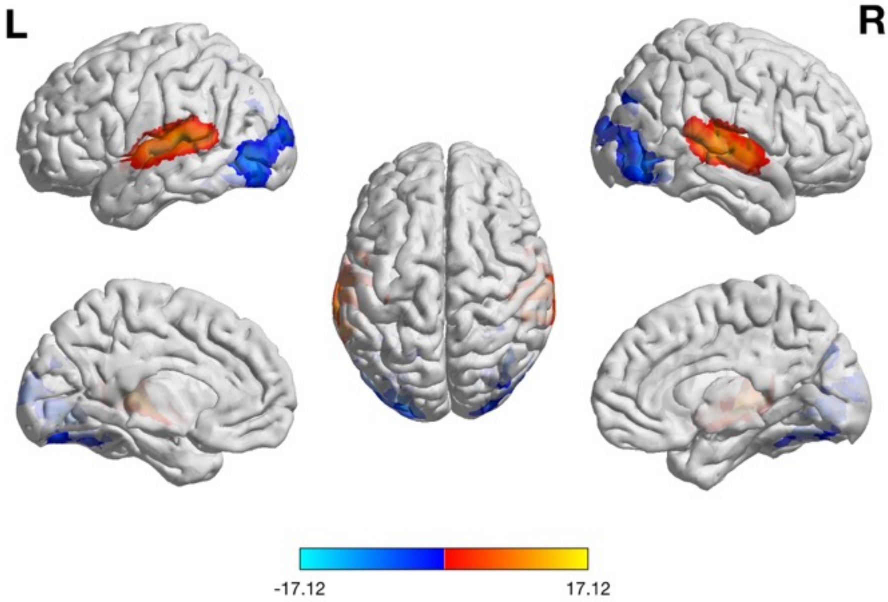


Figure 3.13: Contrasts showing activation in the auditory task (red-gold) relative to the visual task (blue-turquoise) in the imagery condition only. All activation $p < 0.05$ FWE, height threshold $Z = 4.51$, extent threshold $k = 6$ voxels.

Follow-up tests were performed for the visual task with factors Difficulty (easy, hard) and Laterality (left, right), but no effects were found. Similarly, follow-up analyses for the auditory task with factors Difficulty (easy, hard) and Shape (rounded, unrounded) yielded no effects.

3.2.4 Discussion

The primary aim of experiment 3, supplementary to those previously evaluated in experiments 1 and 2, was to measure areas of increased cortical activation during visual and auditory imagery tasks, relative to rest as well as a baseline reaction time condition. While there was an interest in the whole imagery network, there was a specific interest in motor cortex activation.

The results show that there are several areas with increased activation in both imagery tasks relative to the baseline tasks. Both imagery tasks resulted in shared activation in a number of areas that have previously been associated with motor imagery, namely activation in premotor cortices pre-SMA/SMA, left SPL, and superior and middle frontal gyri (Kosslyn et al., 1998; Parsons et al., 1995; Vingerhoets et al., 2002; Zapparoli et al., 2014). Tian et al., (2016) refer to these areas, including insula as activated here as well, as the neural network mediating motor simulation. SMA, premotor cortices and insula, as well as IFG, are involved in motor planning, and as Tian et al., reason, it is sensible that these same areas are also active where motor simulation is concerned since efference copies must still be generated. However, the next step in their simulation-estimation stream suggests expected activation in parietal operculum and somatosensory regions such as supramarginal gyrus, areas not found to be activated over and above the simple reaction time experiment above.

No activation was found in DLPFC, but it is possible that DLPFC activation is dependent on task: this area is known to be engaged in response inhibition (see Blasi et al., 2006 and Rogasch et al., 2015 for examples) and previous activations during motor imagery research designs may have been due to task instructions suggesting that participants inhibit their overt speech. There was consistent activation in a small IFG premotor area which has previously been associated with action observation. More specifically, Hardwick et al., 2018 find that this area is found to be more active during action observation rather than motor imagery. Although this area is near a language area (*pars opercularis*) and phonetic stimuli were used in the auditory task, this location is more dorsal of where auditory IFG activation would be expected. A similar observation is the increased activation of FEF in motor imagery tasks, which may simply be due to the sustained visual attention to the visual stimuli. One area that was found to be more active during the reaction time task than the motor imagery task was posterior angular gyrus. This finding is likely related to the angular gyrus' established role in the default mode network, with activation likely induced due to the restful visual fixation during the reaction time task (Raichle, 2015).

It should be noted that the chosen HRF function for the analysis was canonical HRF + temporal derivative, which allows for +-1s deviation from the trial onset. As both the reaction time and imagery tasks show, the auditory task responses consistently lagged behind the visual responses by around 100 – 150ms. Although it is unlikely, it is possible that the HRF function as modelled did not capture the auditory imagery processes as well as it did the visual imagery processes. One solution would be to instead model the results using the reaction times per trial rather than the trial onset. Alternatively, it might be useful to use parametric modulation, in which the stick function applied to the HRF as modelled is altered depending on the length of the trial, so that the BOLD response variability is partially accounted for by the reaction time. In addition, a recent analysis technique called searchlight representation similarity analysis (RSA) had been proposed as a useful tool to investigate phoneme-level neural representations and may shed further light on speech-related neural activation (Evans & Davis, 2015; Zhang et al., 2020).

Finally, while the tasks were designed to be analysed using a 2 x 2 repeated-measures ANOVA and so some implicit assumptions were made regarding difficulty and attentional demands, it is clear that there are differences in the tasks inherently. For instance, Figure 3.12 shows that the FEF was more active during both imagery tasks, which is likely due to sustained attention and a consequence of participants observing the stimuli for longer (especially in the visual tasks). As such, it may be advisable to change the baseline tasks to tasks which would elicit a similar timespan of observation and attention in order to have a more appropriate basis for comparison with the imagery tasks.

3.3 Discussion

3.3.1 General discussion

Three experiments were conducted to investigate the efficacy of two tasks, one visual imagery task and one auditory imagery task. The visual task was successfully replicated in all three experiments, with results showing higher reaction times and lower accuracy for laterally-rotated hand stimuli than medially-rotated hand stimuli, even when using few stimuli (8 in total) and regardless of whether the experiment was supervised. In addition, several cortical areas related to motor imagery showed increased activation in line with previous neuroimaging results. The novel auditory task successfully showed that native vowel stimuli had higher accuracy and faster reaction times than the non-native stimuli, even with fewer stimuli than used initially. However, the results were more complicated as there was some effect of rounded vs unrounded shape in experiments 1 and 2. In addition, no cortical activation was observed in M1 over and above that found for simple hearing. The results are discussed in detail below.

The first major aim to be addressed (see 2.1.2 for a list of aims and hypotheses) was replicating the visual task successfully. The visual task has been used in several previous studies (Hyde et al., 2017; Kosslyn, Thompson, et al., 2001; Parsons et al., 1995), and the version closest to previous iterations was used in experiment 1 (in terms of general design and number of stimuli). Experiment 1 showed a clear replication of results from previous studies: easy (medial) stimuli showed faster reaction times and increased accuracy relative to difficult (lateral stimuli), irrespective of which hand (left or right) was used to show the rotation. The versions of the task used in experiments 2 and 3 featured a sharply reduced number of stimuli (from 16 to 8), yet the same results were found throughout, with only moderate changes in effect sizes. These results show that the task itself is robust with regard to number of stimuli, as well as how often each stimulus is shown. These results confirm the hypothesis that lateral rotations would results in higher reaction times than medial rotations, the hypothesis that there would be no differences between left and right hand stimuli reaction times per se, and that medial rotations would be judged with significantly lower accuracy than lateral rotations (see 2.1.2).

The second major aim was to create an auditory imagery task using the design of the visual imagery task, that was equally balanced and equally difficult. In all three experiments, it appears that the visual task overall was easier than the auditory task, as suggested by overall effects of reaction time and accuracy (though no reaction time difference in experiment 3). While it was expected that reaction time would differ as a result of different stimulus onset times (auditory stimuli consistently have lower onset time since a sibilant appears before the vowel, while the visual stimulus appears in its entirety immediately), accuracy was not predicted to be different. Since accuracy was consistently lower for the auditory task, the results are not as expected.

In each experiment there was also an effect of the difficulty of the stimuli, showing that the stimuli that were predicted to be difficult (following previous literature), were in fact more difficult than those that were predicted to be easier. In this sense, the auditory task is successful in that, similar to the visual task, there is a clear difference between the stimuli that are easy to imagine and those that are harder to imagine. In this case, the task shows that native vowels are easier to imagine than those that are not native to the participant. This is also underpinned by the consistent strong to medium effect sizes found for this effect (Hedges' g = 0.945 in experiment 1, 0.926 in experiment 2, 0.424 in experiment 3 respectively).

However, in each experiment there was an important interaction in at least the accuracy data, and in the case of experiment 1 also the reaction time data, between the task and difficulty. This is an important point as it shows that the two tasks are differentially affected by the differences between easy and hard stimuli. As the effects across experiments show, it is clear that while the easy stimuli are comparable in terms of reaction time and accuracy across tasks, the difficult stimuli are not: difficult auditory stimuli consistently show lower accuracy and longer reaction times (in experiment 1) compared to their visual counterpart. It appears that the accuracy gap in difficulty in the visual task is narrower than this same gap in the auditory task, and it is likely that this underpins the difference in accuracy performance seen between tasks. While this makes comparisons across tasks somewhat more difficult, it does not invalidate the use of either task.

The final major question was whether the novel auditory task would show similar differences between difficult and easy stimuli. Throughout all three experiments, there was a clear distinction between the easy (native) and difficult (non-native) stimuli, showing faster reaction times and higher accuracy for easy stimuli in all experiments. This confirmed the first hypothesis for this task: non-native vowels were expected to result in longer reaction times than native vowels (see 2.1.2). It was clear that this effect was robust, given the strong effect sizes regardless of number of stimuli, and whether the task was performed under supervision or online. However, experiments 1 and 2 also showed that there was a separate distinction dependent on the shape of the vowel, with rounded stimuli showing faster reaction time (experiment 1 only) or greater accuracy (experiment 2 only) than unrounded vowels. This result refutes the hypothesis that there would be no differences between rounded and unrounded vowels. The effect of shape complicates the results, as theoretically there should be no difference dependent on mouth shape, in the same way that there should be no difference in the visual task dependent on which hand shows the rotation. In experiment 1 this result is found outright, but in experiment 2 it is attenuated by an interaction, which shows that the difference between unrounded and rounded vowels is present only in the case of difficult (non-native) vowels, and not for easy (native) vowels. Additionally, experiment 3 does not show any differences related to mouth shape. It appears that the unrounded, non-native stimuli are particularly difficult for participants, a finding not found in the visual task (an appropriate comparison might be that lateral rotations are more difficult generally than medial rotations, but that left-hand lateral rotations are especially difficult), though again this finding is not present in experiment 3. Overall, the novel auditory task performs adequately, but the expectation that there would be no interplay related to shape was not met in experiments 1 and 2. This precluded the use of the task in further experiments without further modifications and testing.

Several other elements of the study design were investigated over the course of the three experiments. As previously explained (see 2.2.1.3) counterbalancing of response buttons/switches was effected in the auditory task, and not the visual task. No experiment showed any differences on this basis.

In each experiment, there was also an element of learning/habituation, indexed primarily as a reduction in reaction time, and secondarily as an improvement in accuracy. None of the experiments showed any change in accuracy throughout the study, but all experiments showed a similar pattern of reduced reaction time from the first block to the last. Interestingly, the experiments do show a different pattern when taking task into account as well: experiment 1 shows that learning takes place in both tasks, but more rapidly in the visual conditions, experiment 2 shows learning taking place only in the auditory task, with reaction time stagnant for the visual task across blocks, and experiment 3 shows consistent learning rates across tasks. These results are somewhat contradictory since they do not follow the same path, but there does appear to be learning in at least the auditory condition. It is possible that the auditory condition shows more learning because of the fact that the stimulus is shown more briefly (with respect to the visual condition, where the stimulus remains onscreen throughout the trial), which may lead to a longer time-course of learning.

Additionally, I investigated whether the visual and auditory tasks would be suitable for on-line and unsupervised testing, specifically with a reduced number of stimuli (compared to the original experiment 1). The analysis showed that the general effects were replicated. The results showed no evidence to suggest that the tasks were performed differently online and in the fMRI scanner, save for a small delay in reaction time, which confirms the hypotheses that reaction times and accuracy differences between tasks would be maintained with fewer stimuli (see 2.1.2).

3.3.2 Limitations

As with most motor imagery research, the main limitation is a lack of direct, non-self-reported feedback from the participant on what strategy was used to perform the imagery tasks. Previous research, such as Zapparoli et al., (2014) show that motor imagery is more likely to take place during the visual task when using palm-only views, as the back views of hands led to activation in occipital rather than motor cortices. However, none of the experiments can be used to state with certainty that motor imagery processes were used. One anecdotal piece of evidence that suggests motor imagery was used intuitively was that the majority of participants used their hands and lips in an attempt to make a decision during the training task and were asked not to do so during the actual experiment. This suggests that a physical movement trajectory was often attempted, but this evidence is anecdotal and does not preclude the possible changing of strategies some way through the tasks. However, while an increase in blood flow may suggest a certain strategy, it does not allow one to definitively rule out other motion-related strategies. For instance, Zapparoli et al., (2014) suggest that increases in blood flow in occipital regions suggest an image-rotation strategy (i.e., a difficult hand stimulus is rotated to an easier rotation, such as 0°, allowing for facilitated decision-making) while increases in blood flow in motor regions suggest a motor imagery strategy, but the authors cannot rule out the strategy where participants imagine physically moving the stimulus itself, as is the case in Kosslyn, Thompson, et al., (2001). Additionally, many of the areas that were active during imagery are also thought to be active during inhibition of motor execution, and it is difficult to disambiguate which process are exclusively involved in which task. One potential solution is to move towards a more explicit motor imagery task (see Chapter Four and Chapter Five).

Another limitation of this experiment is the lack of electromyography (EMG) to monitor movement in the task-relevant muscles. For instance, McGuigan and Dollins (1989) perform an experiment in which participants were asked to silently read letters ('P', 'T', control figure) while EMG readings were taken from lip and tongue muscles. The results show that when silently reading 'P', EMG readings increased relative to the baseline and relative to when participants read 'T'. Similarly, EMG readings of the tongue muscle while participants read 'T' were significantly higher than baseline and compared to when participants read 'P'. In both cases, there was no difference between the control baseline and the non-relevant letter, showing that the increase in activity was articulator-specific rather than general speech-related. This result in turn suggests that consonant-specific motor imagery of speech took place when specific letters were read silently. A similar set-up was intended for experiment 1, but was discarded primarily because there is no difference expected in lip EMG activation between unrounded and rounded lips, since both positions require similar muscular increases. This point is specifically related to the use of vowels, since only a small number of vowels could realistically be compared in terms of lip involvement (a specific example might be /ə/ vs /u/). Consonants are perhaps the more useful category, but as mentioned in the introduction (2.1.1), there is the potential issue of categorical perception that need to be taken into account when using consonants for such a task.

Related to the previous point, one limitation of these experiments is the lack of a control condition in which no motor imagery action was required. Experiments 1 and 2 do not contain control stimuli but experiment 3 does contain simple reaction time tasks (although not interleaved with the imagery tasks). While not necessarily an issue in the experiments presented here, it is not forward-thinking with respect to the objective of the use of this task in a TMS-MEP paradigm (see 2.1.2). The lack of a built-in control task or control stimuli means that the task would necessitate further changes before it could be used in a TMS design. Similarly, in retrospect, the use of these vowels causes the issue regarding differentiation between lip and tongue muscles previously alluded to, which is likely to be an important element in the TMS design (MEPs measured from lip and tongue muscles separately, and the ability to discriminate on the basis of presented stimulus).

Two additional limitations concern fMRI experiment 3 specifically. Firstly the number of stimuli per participant per difficulty category amounted to 8 in each modality, which is quite low. This is reflected in the fact that no significant clusters were found related to hand laterality even though such a difference was found in Zapparoli et al., (2014) and no differences were found between easy and hard stimuli in either task. As the learning data show, over the time course of the tasks participant reaction time decreases indicating improved performance, limiting the number of times a stimulus can be shown and active imagery expected rather than habituation to individual stimuli. Even so, it is possible that the low stimulus number affected the activation seen, in which case the auditory task in particular might benefit from an increased number of trials.

Secondly, the overt response used in both tasks may have led to activation of motor or sensorimotor cortices being less discernible, especially if M1 activation in motor imagery is quite low. While using an analysis modelled using the trial start time and temporal derivative likely helps in this regard (as opposed to a stick function that is more closely related to reaction time and therefore time of overt action) it is possible that small but significantly different activations between tasks were dwarfed by the preparation and execution of the finger motor response, which was present in each task.

3.3.3 Implications

The results from experiments 1, 2 and 3 have several important implications, both generally for motor imagery and specifically related to the programme of study.

The first implication concerns the robustness of the visual task: it is quite clear that the visual task is extremely robust regardless of number of stimuli shown, and the conditions under which the task takes place. The auditory task, in comparison, is not quite as robust: there are several important differences between the modified tasks, as well as differences dependent on the conditions under which the task is performed. The general implication is that the task needs refinement, further piloting (perhaps with different vowel stimuli) and further replications by others before the task can be considered robust. As a result, with regard to the programme of study, the differences between the experiments are somewhat too great to make use of the task in further experiments.

As previously discussed, while there is a plethora of evidence for motor imagery during the visual task, the auditory task does not benefit from such support at this time. These first results are promising, and the general effects concerning reaction time and accuracy relative to stimulus difficulty correspond well with expectations based on visual task experience. The neuroimaging data also support the notion that a similar network, whether it be for motor planning or active simulation, is used in the auditory task. The implication is that if there is good evidence to assume that motor imagery of hand movement occurs in the visual task, there is similarly good evidence that motor imagery of lip movement occurs in the auditory task, even considering the novelty of the task. Both the behavioural data and fMRI data support this notion.

3.4 Conclusion

In this first chapter, three experiments using a visual and auditory imagery task were presented. While the visual imagery task resulted in the replication of previous behavioural and neuroimaging findings, the auditory task results were more complex. Specifically, the differences observed between native and non-native vowels were at times confounded by differences found in the lip shape of the vowel stimulus.

The results from experiments 1-3 show that reaction time and accuracy measures between the visual and auditory task show similar trends, and that neuroimaging results suggest that the same cortical apparatus is used in both tasks. While the imagery task fMRI results showed a network involved in imagery, the imaging results do not support the notion that M1 is involved in motor imagery, either visual or auditory. However, as previously discussed (see 1.4), neuroimaging may not be the optimal research technique for observing transient, small increases in activation such as those that would be expected in imagery of movement. Instead, the TMS-MEP method offers improved precision temporally and spatially, but is limited to M1. As discussed in section 3.3.2 (Limitations), the auditory task using vowel stimuli as in experiments 1-3 presents significant challenges with regard to being used in a neurostimulatory design such as a TMS-MEP design.

Chapter 4 presents an experiment that uses a consonantal auditory imagery task, embedded in a TMS-MEP design, to establish whether M1 is used in auditory imagery of simple articulatory gestures.

4 Chapter Four

Sections of this chapter were previously submitted for publication to the journal Frontiers in Human Neuroscience with the title "Motor Imagery of Speech: The Involvement of Primary Motor Cortex in Manual and Articulatory Motor Imagery" (published 11th of June 2019, see Maegherman et al., 2019).

4.1 Introduction

In section 1.3, motor imagery was defined as 'motor imagery processes that involve volitional, agentive, and expanded articulatory motor processing'. The concept of motor imagery is inherently linked to that of motor execution, and the former is readily conceptualised as a covert form of – or subset of processes relating to – the latter (Jeannerod, 2001). Fadiga et al., (1998) proposed that motor imagery includes aspects of motor execution experience, on the basis that numerous biological parameters such as blood pressure and heart rate, which increase during action execution, also correlate positively with action effort during motor imagery. Likewise, the time course of motor imagery and motor execution of actions has been found to be similar (Parsons, 1994). Functional imaging has also shown some overlapping activation for both types of processing, including sensorimotor and premotor cortices, putamen and cerebellum (Hardwick et al., 2018). Earlier studies (Porro et al., 1996; Roth et al., 1996) also claimed activation in M1. Since such measures suggest a significant level of shared processing between the two behaviours, the extent to which motor plans are used in motor imagery compared to motor execution has been the subject of debate. Notably, Vogt et al., (2013) proposed a continuum from action observation to motor imagery, along which the extent of the simulation of motor plans differs, though neither effect motor unit activation. Similarly, Jeannerod (2001) proposed that covert and overt action stages also represent a continuum where execution suggests the existence of a covert stage, but covert actions do not evolve into the overt stage. In this sense, it can be said that a continuum of action processing, and so use of motor plans for potential simulation of action, exists from action observation through covert simulation of motor plans to overt execution of motor plans. Some models include primary motor cortex (M1) as an important element in simulation processing (Lœvenbruck et al., 2018) while others suggest it is involved only in execution processes (Tian, Zarate, & Poeppel, 2016). The central question remains to what degree motor imagery engages motor processing in the absence of overt action execution, and which regions are involved in these processes.

Early studies of motor imagery employed mental rotation paradigms and investigated how complexity can affect reaction time. Two well-known examples of this are the Shepard-Metzler visual rotation task (Shepard & Metzler, 1971) and the hand laterality task (Parsons, 1994), which were discussed in chapter 2. Such tasks have been adapted for use with imaging methods such as positron emission tomography (PET) (Deiber et al., 1998; Kosslyn, Digirolamo, Thompson, & Alpert, 1998) and functional magnetic resonance imaging (fMRI) (Porro et al., 1996; Richter et al., 2000, see Munzert et al., 2009 for review). Many of these imaging studies found activation in supplementary motor area (SMA) and premotor areas (Kosslyn, Thompson, Wraga, & Alpert, 2001; Richter et al., 2000), while some studies also find activation in M1 (Solodkin, Hlustik, Chen, & Small, 2004; Stippich, Ochmann, & Sartor, 2002; see Munzert et al, 2009, p. 308 for an extensive overview), and especially early papers (see Porro et al., 1996) disagree on whether motor imagery increases activation in M1.

Several studies examined the involvement of M1 in motor imagery, primarily in the hand area of M1, using transcranial magnetic stimulation (TMS). For instance, Izumi et al., (1995) asked participants to REST, THINK (motor imagery), or CONTRACT (motor execution) with regard to index finger abduction, with electrodes measuring electrical activity at the *first dorsal interosseous* (FDI – index finger abductor). A TMS pulse was administered 3-5 seconds after the instruction was given. Fifteen motor-evoked potentials (MEPs) were collected per participant per condition. Significant differences were found between the positively modulated THINK and non-modulated REST conditions, as well as between CONTRACT and THINK conditions, and CONTRACT and REST conditions, indicating that M1 was active not only during motor execution, but also during motor imagery. Further studies investigating motor imagery effects showed differences between motor imagery and rest conditions in forearm (Yahagi et al., 1996), as well as biceps and thumb muscles (Fadiga et al., 1998). With respect to other muscles, F. Tremblay et al., (2001) similarly examined leg muscle responses (quadriceps and thigh muscles) using MEPs and found similar facilitation effects in M1 for imagery respective to non-facilitation during a rest baseline. In summary, a number of TMS studies have shown that motor imagery increases the excitability of primary motor cortex as measured by larger MEPs.

While the above studies focus on MEP increases, there is little discussion regarding the chronometry of motor imagery processes, as few studies included several timepoints across which MEPs were compared. However, one study by Hashimoto and Rothwell (1999) used TMS to investigate a simple imagery task with flexion and extension of the wrist, measuring MEP amplitude at the *first dorsal interroseous (*thumb abductor, here used as the control muscle), *flexor carpi radialis* and *extensor carpi radialis* (two muscles engaged in radial abduction of the wrist). Nine participants were told to imagine performing repetitive wrist movements at a rate of 1Hz, with TMS applied over the hand area of M1 at 0ms, 250ms, 500ms and 750ms after an auditory cue. Using averages of 8 MEPs, the authors found increases in MEP amplitude over the time course 0ms - 250ms, with a plateauing and decrease of MEP amplitude at 750ms in *flexor* and *extensor carpi radialis*. The authors did not compare activation in the imagery condition with an execution condition. For experiment 4, one aim was to replicate these chronometric results with the *first dorsal interosseous* as muscle of interest.

Hyde et al., (2017) used the hand laterality task to measure the involvement of M1 in motor imagery of simple and extensive manual rotation using TMS. Hyde et al. measured MEPs from the hand (FDI) region in M1 to measure its involvement in motor imagery. Single-pulse TMS was administered at latencies of 50ms, 400ms and 650ms post-stimulus presentation. The latter two time-points (400 and 650ms) were included as previous studies (Ganis, Keenan, Kosslyn, & Pascual-Leone, 2000; Tomasino, Borroni, Isaja, & Rumiati , 2005) showed increased reaction times during the task when TMS was administered at these time points. The early (50ms) timepoint was added as an early stimulation reference point. Twenty-two participants were tested and subdivided into groups of 'likely motor imagery users' and 'likely non-motor imagery users' on the basis of whether performance was better in case of simple biomechanical actions (e.g. a 90° internal rotation) than of complex biomechanical action (e.g., a 90° external rotation). Those likely to have used motor imagery were thought to use a strategy whereby they mentally rotated their hand to fit the hand image displayed. Conversely, those likely not to have used motor imagery were thought to have used a strategy whereby they rotated the on-screen picture to fit their hand position. When response times were not modulated on the basis of physical difficulty, the authors assumed that participants used a different strategy, e.g., rotating the visual stimulus rather than imagining movement. In this case, subjects were placed in the likely non-motor imagery group. In the likely motor imagery users group, there was an increase in MEP at all three timepoints (vs. baseline), while this was not the case for likely non-motor imagery users. This increase in MEPs during motor imagery presented evidence for increased cortical excitability of hand M1 during imagery of manual movement.

The aim of experiment 4 was to replicate and extend findings of previous work (Hashimoto & Rothwell, 1999; Hyde et al., 2017) on motor imagery in hand muscles by studying the effects of motor imagery for facial muscles, specifically the articulatory muscles, i.e., *orbicularis oris* (OO) in a sub-phonemic context. The sub-phonemic quality of a simple lip task is on par with the non-linguistic nature of a simple hand task – a task containing linguistic elements is eschewed in favour of establishing the viability of basic imagery data collection from hand and lip effectors. While motor imagery has been investigated using several hand muscles as well as leg muscles, there is a lack of studies in other muscles –there has been no investigation into articulatory muscles, such as lips or tongue, using TMS during speech imagery even though speech imagery is a well-known phenomenon (Alderson-Day & Fernyhough, 2015; McGuire et al., 1996; Oppenheim & Dell, 2008; Sokolov, 1972). Moreover, the effect of observation of speech on cortical excitability of lip area is similar to observation of other motor sequences, e.g. manual sequences (Fadiga et al., 2002; Watkins et al., 2003). A substantial number of researchers have examined motor imagery of speech actions using fMRI with motor tasks (Hardwick et al., 2018; Huang et al., 2001; Shuster & Lemieux, 2005; Szenkovits et al., 2012; Tian et al., 2016). For instance, Tian and Poeppel (2010; 2012) used fMRI and motor imagery tasks to investigate the topography and time course of imagined speech actions and found that such imagery actions are likely part of a simulation process. This simulation process is likely a sub-process of an internal forward model used in prediction and online correction of movement errors (see section 1.3.2). Manual motor imagery has been investigated with neuroimaging and by measuring MEPs, and motor imagery of speech has been investigated using neuroimaging techniques only. No study so far has investigated cortical excitability as measured using MEPs related to motor imagery for speech actions and no study has directly compared cortical excitability of hand and lip actions in a motor imagery context.

Experiment 4 examined motor imagery in hand and lip muscles while participants performed a simple motor imagery task. While a number of previous studies used a multi-finger finger tapping task to investigate hand motor imagery, such a task is difficult to extend to the lip muscles due to the smaller number of joints and muscles involved. Instead a simple imagery task was used, such that it could be performed equally with both effectors using simple compression, similar to joint flexion (Hashimoto & Rothwell, 1999). Participants were asked to either imagine pressing together a set of wooden tweezers between their lips (lip motor imagery task), or between their thumb and index finger (hand motor imagery task), or to press together the wooden tweezers between the lips (lip motor execution task), or fingers (hand motor execution task), or to do nothing (remain at constant contraction - baseline). The squeezing together of the lips in particular can be considered a sub-phonemic articulatory gesture, used in the production of plosive sounds, such as /p/. In line with Hyde et al., (2017) and Hashimoto & Rothwell, (1999), the experiment used a chronometric design and measured MEPs at various time points from trial onset, including 50, 150, 250, 350, 450 and 550ms. A wide range of time points was used as it is unclear to which extent motor imagery and motor execution of speech muscles follow the same time course. This experiment aimed to clarify to which extent M1 is engaged during motor imagery of simple speech and hand tasks, as well as how this involvement develops over time. The primary hypotheses for this experiment are as follows:

(1) Motor execution conditions show significantly larger facilitation than imagery and baseline conditions, since M1 is most active;
(2) Motor imagery conditions show significantly larger facilitation than baseline conditions, since M1 is active above baseline due to simulation processes.

Furthermore, MEP amplitude is expected to increase with later timepoints, but this expectation is exploratory as the time course of imagery and action is uncertain.

4.2 Experiment 4

4.2.1 Methods

TMS is a non-invasive neurostimulation technique that uses strong magnetic pulses to stimulate neuronal populations in the cortex. It is based primarily on Faraday's law of electromagnetic induction, which states that an electric current sent through a wire generates a time-varying magnetic field perpendicular to that wire, which in turn generates a current in a secondary nearby conductor (Ruohonen, 2003). In the case of TMS, the wire is a wound coil, most often in the shape of a figure-of-8, placed on the head such that when a current is sent through the coil, it generates a magnetic field that rapidly (< 1ms) decays, which generates a current in a nearby conductor, here the brain. The magnetic field passes unimpeded through the skull and if the current is of a sufficient intensity, depolarises neurons in a local population in a specified cortical area. It is believed that TMS modulates activity in grey matter more than white matter, and that stimulation occurs along axons (specifically in axonal bends or where discontinuities exists, e.g., at axonal boundaries) rather than the cell body. In the research presented, TMS is used in its single-pulse form to evoke motor-evoked potentials (MEPs) as measured through electromyography (EMG). MEPs are the results of stimulation of the motor cortex (M1) which is organised somatotopically, allowing for the stimulation of cortical areas related to specific muscles, such as those of the hand, lip and tongue. Motor neurons form motor units with skeletal muscle fibres which are innervated by the stimulated neurons. An increase in MEP size relates to the number of activated motor neurons, and are therefore reflective of underlying cortical motor activation: the more activated the motor region, the larger the MEP (Devlin & Watkins, 2007). MEP amplitude changes are therefore said to reveal the relative excitability of the related M1 region and has been used to detect M1 involvement various language processes, such as speech listening and speech observation (Watkins et al., 2003). Increases in MEP amplitude due to other cognitive processes, relative to baseline measurements, are referred to as corticospinal facilitation. For example, Aziz-Zadeh et al., (2004) found that left-lateralised hand MEP amplitude increased when participants were exposed to sounds related to hand actions, such as typing, revealing facilitatory effects of auditory manual sounds on hand M1.

This experiment contains sections in which TMS is used to investigate hand and lip motor areas. Subcortically hand MEPs, and lip (as well as tongue) MEPs arise via different efferent motor pathways. The corticospinal pathway innervates the hand while the corticobulbar pathway innervates the lip and tongue muscles (Adank et al., 2018). As a result, there are differences between MEPs recorded from hand, lip and tongue motor areas. One difference concerns MEP latency, which is longer in the hand due to the length of the corticospinal tract compared to the corticobulbar tract, resulting in an MEP around 20ms post-pulse as opposed to 8-10ms post-pulse in the lip and tongue. A second difference concerns MEP morphology: hand and tongue MEPs often show a single peak (monophasic) waveform, whereas lip MEPs often show multi-peaked (polyphasic) waveforms. As a result, the MEP extraction method, usually area-under-the-curve (AUC) or peak-to-peak (P2P), is critical and should be standardised across tasks. Throughout the thesis preference is given to AUC for in-depth analysis, although P2P values are also provided and analysed to a limited degree as they provide a useful intuitive comparison, especially in line with older literature.

One of the key design choices in a TMS study is whether to use active or resting motor threshold (aMT versus rMT). The use of motor threshold is itself a choice: MEPs can be elicited by using a consistent threshold expressed in percentage of stimulator output (maintained for all participants), or as the percentage of stimulator output required to elicit MEPs of a given amplitude in each participant (Rossini et al., 1994, 2015). One advantage of a pre-determined threshold is that less time is spent setting up the experiment, at the risk of using a stimulation threshold that is too low or too high for a subset of participants. An advantage of setting individual thresholds is that MEPs are more representative of each participant's motor pathway idiosyncrasies (e.g., regarding sensitivity to TMS pulses) and therefore any subtle changes in MEPs, at the cost of time spent finding the threshold. If the choice is made for individual thresholds, active or resting motor threshold can be used. Active motor threshold (aMT) represents the percentage of stimulator output required to reach a predetermined MEP amplitude in an active muscle in 5 out of 10 pulses delivered (as per Groppa et al., 2012) or 10 out of 20 pulses delivered (as per Rossini et al., 2015) during a pre-testing thresholding procedure. Resting motor threshold (rMT) represents much the same, but for a resting muscle. aMT is preferable to use in muscles such as lip or tongue, which are rarely at rest and are difficult to keep in a natural resting position (Watkins et al., 2003). For this reason, individual aMT thresholds were used in experiments 4, 5 and 6.

4.2.1.1 Participants

Twenty participants (11F, 9M, mean age = 22y 7m, *SD* = 3y 8m, age range = 19-34) took part in the experiment. Handedness was established via the Edinburgh Handedness Inventory (Oldfield, 1971), which found all participants to be right-handed. Participants reported no history of neurological or psychiatric disease, and none reported use of any long or short-term medication. No medical conditions, relevant or otherwise, were reported, and neither was any history of specific repetitive motor activity or muscle disorders. All subjects completed all conditions on the same day in one session and had a minimum high school-level education. Experiments were undertaken with the understanding and written consent of each subject, according to the University College London Research Ethics Committee (UREC #0599/001). Data collection and reporting conventions are as suggested in Chipchase et al., (2012).

4.2.1.2 Materials

A tweezer-like tool was used to effect constant muscle contraction as well as to provide a tool for the motor execution condition (Figure 4.1). The choice was made to use active motor threshold in both muscles to make sure the tasks were as comparable as possible (see also 1.5.2 on the choice of aMT for lip MEPs). Prompts were displayed on a 21.5" computer monitor screen while participants sat approximately 70cm away. Prompts were a combination of symbols (font size 24), with '%%' representing hand use and '&&'

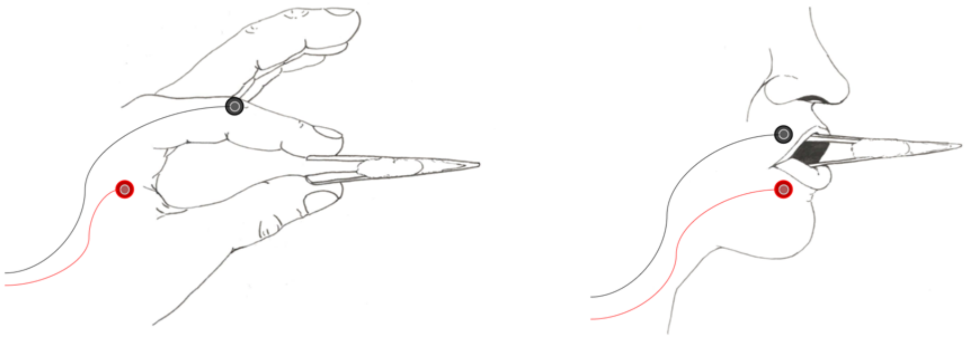


Figure 4.1: Examples of the tweezers as used in the hand (left panel) and lip (right panel) conditions. Participants were instructed to maintain a baseline level of activity as trained on in a pre-thresholding session. New tools were provided for each participant.

representing lip use. Colour was used to indicate whether the action should be imagined (blue) or overt (red), or whether no action should be taken (black). Prompts were displayed on a light-grey background and were preceded by a white fixation cross (see Figure 4.2).

4.2.1.3 Procedure

Each trial began by displaying a white fixation cross for 1000ms. The prompt was shown for 2000ms, with TMS pulses administered at 50ms, 150ms, 250ms, 350ms, 450ms, or 550ms post-stimulus (counterbalanced). The screen was cleared of all prompts and remained clear for at least 2000ms, but longer dependent on when the TMS pulse was given. The inter-pulse-interval was kept constant at 6000ms. The next trial then began with a new fixation cross. Blocks consisted of 25 trials presented in two minutes, with a break of one minute in between each block. Each break was a minimum of one minute, after which the second tester pressed a control key to continue, unless the participant requested more time for a break. Trials were blocked by effector in separate blocks (first effector counterbalanced across participants), with motor execution, motor imagery and baseline prompts as interleaved, mixed trials to avoid potential muscle activation carry-over effects from one stimulus to the next. Each prompt was presented 15 times so that 15 MEPs could be used to derive an average for each trial type at each of the chronometric time points. The experiment consisted of 450 trials in total (150 per task, 30 per time point). The experiment lasted 2 hours (45 mins of TMS).

Participants were recruited using the UCL online participant pool by advertising the study as a session which examined the effects of attention on task processing without explicit reference to MEPs, reaction time, grip strength, or imagery.

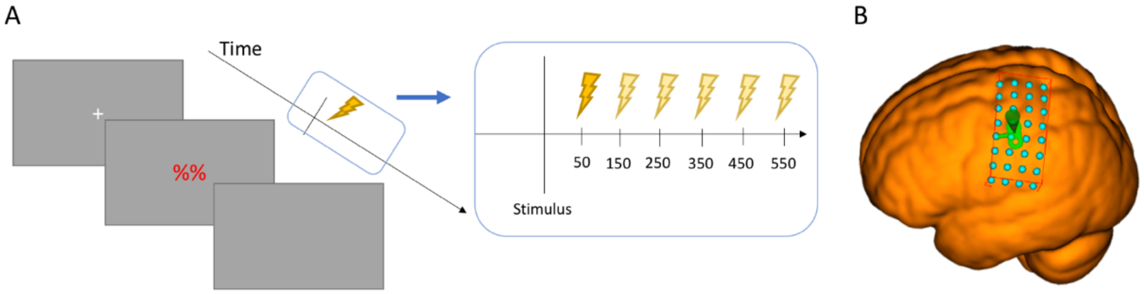


Figure 4.2: (A) Chronometric design of the study showing the six time points in relation to the stimulus; (B) the grid surface for the MNI brain used in thresholding.

Upon arrival, the study was explained to the participant, and they were given information about the TMS procedure. The participant was then shown what they were expected to do for each symbol. A training session showed each trial type four times allowing the experimenter to make comments on the actions performed. Training lasted 2 minutes and all participants were able to successfully perform the actions.

Upon completion of the TMS section of the study, an audiometry test was completed (frequencies of 1000Hz, 2000Hz, 4000Hz, 8000Hz, 250Hz and 500Hz in each ear in accordance with guidelines from the British Society of Audiology, 2012), as was the Montreal Cognitive Assessment (Nasreddine et al., 2005) for baseline cognitive assessment. Finally, participants completed the Varieties of Inner Speech Questionnaire (VISQ, McCarthy-Jones & Fernyhough, 2011). The VISQ was included to test for a potentially positive correlation between any VISQ factors (level of condensed inner speech, level of dialogic inner speech, level of other people represented in inner speech and level of motivational inner speech) and MEP area-under-the-curve (AUC), as it is expressly designed to assess the nature of inner speech and the extent to which it is used on a per-subject level. See Appendix B - E for all tests, and see Appendix A for the correlational analyses.

All MEPs were initially extracted from Spike^2 using custom-made scripts that allowed for the extraction of the electromyography (EMG) signal for a timeframe between 1000ms pre-TMS pulse and 40ms post-TMS pulse. The area-under-the-curve (AUC) of each MEP was then calculated, with hand MEPs spanning the region 13-40ms post-TMS pulse, and lip MEPs spanning the region 8-35ms post-TMS pulse due to their shorter onset time. An equally long section of 27ms was extracted pre-pulse to allow post-hoc checks of equivalent baseline contraction across conditions. The first MEP in each block was then removed as the first MEP is non-representatively larger than those that follow, and this ensures stable neuronavigated coil placement is in place. The mean and standard deviations for the data of each effector was calculated and used to standardise the results and so enable cross-effector comparisons (*z*-scores).

4.2.1.4 Transcranial Magnetic Stimulation

Following the successful conclusion of the training session, the electrodes (Ag/AgCl, ⌀10mm) were attached to the FDI (tendon-belly montage) and the OO (belly-belly montage), and two minutes were used for training to maintain baseline activity at 20% of maximum voluntary contraction. This training was intended to ensure consistent muscle activation maintained throughout the experiment, except when responding in the action condition. The EMG signal was sampled at 5000Hz, amplified by a factor of 1000 and band-pass filtered between 100-2000Hz using a 1902 amplifier, with digital-to-analog conversion using a Micro1401 unit (both Cambridge Electronic Design, Cambridge, UK) connected to a Windows 7 PC. Frameless stereotaxy (Brainsight, Rogue Resolutions, Montreal, Canada) was used to localise the area of stimulation for each effector, making use of the built-in MNI-152 model. A virtual 8 x 4 grid was placed over the motor cortex with the centre corresponding to MNI coordinates -64, -4, 39 (see Figure 4.2). Grid placement was used to search for the hand and lip motor areas in a simple and consistent manner. This system was also used to ensure correct coil location, and to ensure stability was maintained throughout. A figure-of-eight coil (⌀ 70 mm) was placed at a 45° angle relative to the sagittal plane, inducing a posterior-anterior current flow approximately perpendicular to the lateral fissure. Once all grid points had been given a minimum of 3 TMS pulses, the MEPs were extracted from Spike^2, averaged and visualized using a heatmap display command in MATLAB, displaying the best area to stimulate for each effector. Thresholding was performed using a standard thresholding procedure in which 5/10 MEPs must be elicited (Rossini et al., 1994; Rossini et al., 2015; Watkins et al., 2003). A MEP was defined as having a peak-to-peak amplitude of 500μV for FDI and 200μV for OO. Once active motor threshold (aMT) was established, testing intensity was set to 120% of aMT. This resulted in effector-specific motor threshold and testing intensities as necessitated by the inherent differences between the motor thresholds of facial and manual muscles (Groppa et al., 2012).

4.2.1.5 Analysis

Descriptive statistics were performed for the VISQ data to gain insight into participants' individual experience of inner speech. A one-way Analysis of Variance (ANOVA) was performed on baseline data per effector (1 x 3) to check that pre-pulse baselines were roughly equivalent across conditions. This was necessary to ensure that any changes in

MEP were due to changes in response to stimuli, as opposed to due to different baselines. A 2 x 3 x 6 repeated-measures ANOVA was then performed on the combined lip and hand data to establish the effect of the factors Effector (hand vs. lip), Condition (motor execution, motor imagery, baseline) and Timepoint (50, 150, 250, 350, 450, 550ms) on the *z*-scores of MEPs. This overall ANOVA was followed up by two separate ANOVAs for each effector separately to investigate key interactions where applicable: raw data were used for the effector-specific analysis to gain as much understanding as possible. Follow-up *t*-tests were also performed where applicable. Correlation analyses were conducted to investigate possible correlations between VISQ scores, MOCA scores, and MEP size. Due to violations of normality and the ordinal nature of the VISQ, Spearman's *rho* is reported, see Appendix A. All results were Greenhouse-Geisser corrected where assumptions of sphericity were violated. All significance values were corrected for multiple comparisons (Bonferroni) where applicable. MEPs were removed from the analysis if they exceeded 3 SDs from the effector-specific mean, which resulted in the removal of 64 MEPs (~0.7%).

4.2.2 Results

4.2.2.1 VISQ

Participants rated each question on the VISQ on as scale of not applying to them (1 -3) or applying to them (4-6, see chapter 2 for more information on the VISQ). Overall

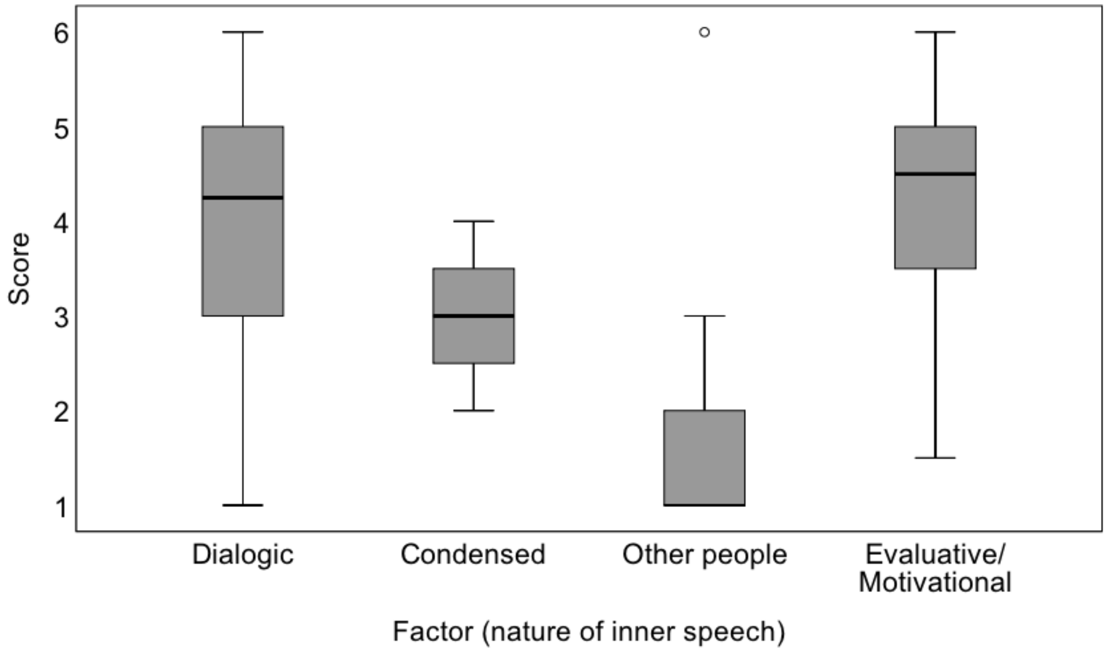


Figure 4.3: VISQ scores per type of inner speech for experiment 4. Scores are amalgamated across five questions that relate to each factor. See Appendix C for the questionnaire. (1 = Certainly does not apply, 6 = Certainly does apply*).*

participants reported that they experienced evaluative/motivational speech most often, followed by dialogic inner speech (see table 3.1). Condensed inner speech was reported to not occur rather than occur (on average), and participants overwhelmingly stated that they did not hear other people's voices in their own inner speech (except in one case, see Figure 4.3)

4.2.2.2 Overall data

No participants were excluded on the basis of the MOCA since all participants scored >26 points. Average MNI coordinates for hand were [-40, -15, 67] and average MNI coordinates for lip were [-59, -8, 46]. Due to issues with the frameless stereotaxy system, it was possible to extract localisation data from 13 out of 20 participants. A one-way ANOVA comparing baseline muscle contraction levels found no significant differences in baseline contraction across conditions in either the hand (motor imagery *M* = 5.882mV·ms, *SD* = 2.628mV·ms; motor execution *M* = 5.956mV·ms, *SD* = 2.601mV·ms; baseline *M* = 5.864mV·ms, *SD* = 2.744mV·ms; *F*(2,57) < .01, *p* = .993, η_p^2 < .001, n.s.) or lip (motor imagery mean = 9.895mV·ms, *SD* = 5.202mV·ms; motor execution mean = 10.163mV·ms, *SD* = 5.044mV·ms; baseline mean = 9.926mV·ms, *SD* = 5.120mV·ms; *F*(2,57) = .02, *p*= .984, η_p^2 = .001, n.s.), showing that any difference in MEPs was not due to different levels of pre-pulse contraction.

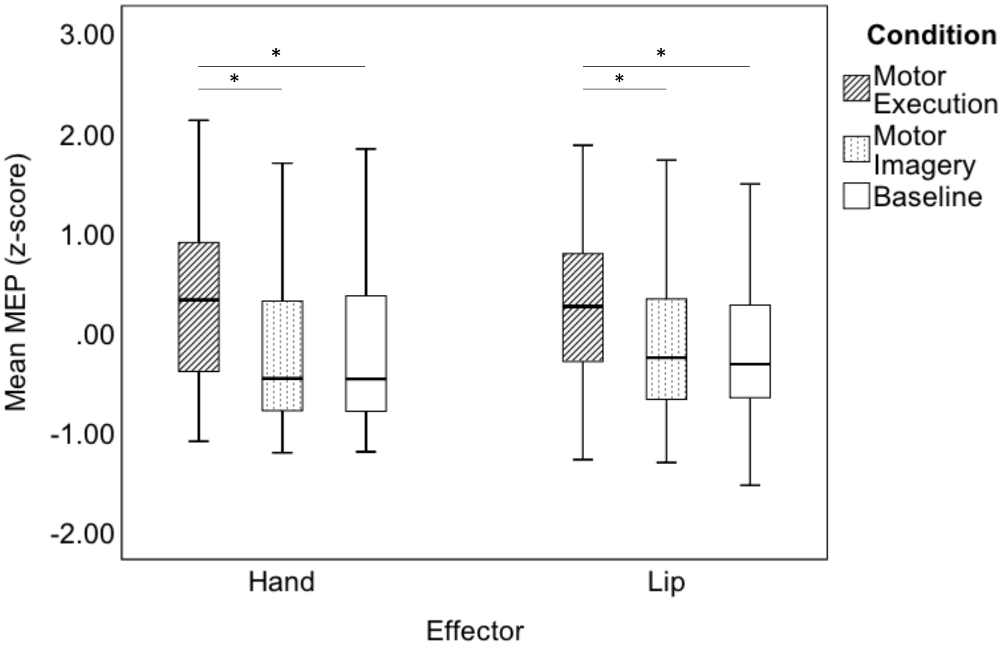


Figure 4.4: Main effect of Condition and post-hoc significant tests for mean MEP z-scores in hand and lip effectors. * = significant at *p* = 0.05.

Z-scores of the averaged data were used for a 2 x 3 x 6 omnibus ANOVA comparing Effector (hand vs. lip), Condition (motor execution, motor imagery, baseline) and Timepoint (50, 150, 250, 350, 450, 550). There was no significant main effect of Effector ($F(1{,}19) = .01$, $p = .906$, $\eta_p^2 = .051$, n.s.), which was expected since z-scores were used. There was a significant main effect of Condition ($F(1.103, 20.957) = 54.20$, $p<.001$, $\eta_p^2 = .740$). Post-hoc comparisons show that motor execution MEPs ($M = 0.297$, $SE = 0.097$) were significantly larger than motor imagery ($M = -0.177$, $SE = 0.122$, $p < .001$) and baseline MEPs ($M = -0.188$, $SE = 0.116$, $p < .001$), but that there was no difference between motor imagery and baseline MEPs ($p = 1.000$), see Figure 4.4). A second main effect was found for Timepoint ($F(2.600, 49.401) = 22.27$, $p < .001$, $\eta_p^2 = .540$). Post-hoc comparisons showed a significant difference between the first timepoint and all subsequent timepoints (minimum $MD = -0.166$, $SE = 0.036$, $p = .003$), except timepoint 150 ($MD = -0.030$, $SE = 0.022$, $p = 1.000$), see Figure 4.5.

There was also a significant Condition × Timepoint interaction ($F(3.760, 71.444) = 24.65$, $p < .001$, $\eta_p^2 = .565$), showing that the trend of change across timepoints was different depending on the condition involved. Figure 4.6 shows the difference across conditions: the data show that starting from timepoint 250 onwards, there is a significant difference between the motor execution condition and the other conditions. No other interactions

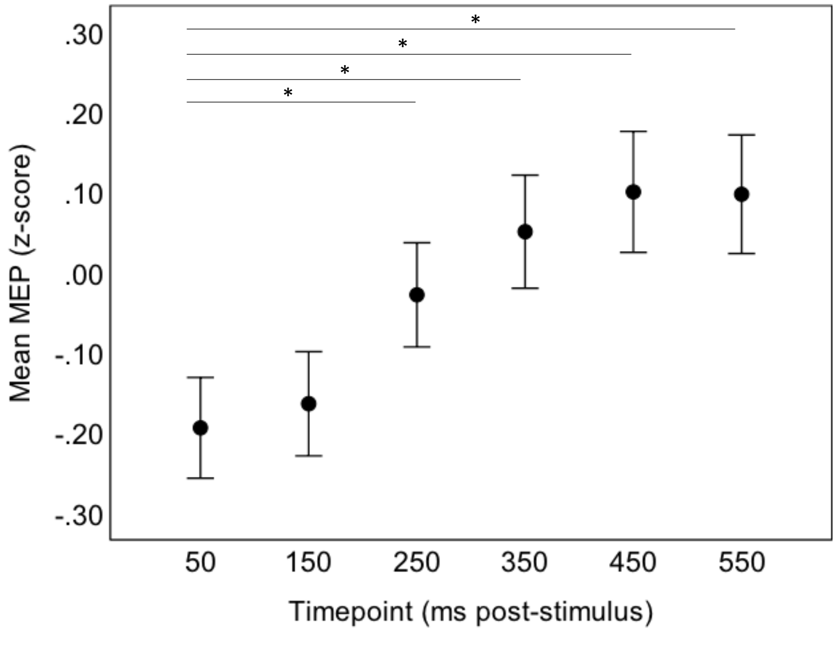


Figure 4.5: Significant increases in MEP size per timepoint across effectors and conditions. Error bars indicate +- 1SEM, * = significant at $p = 0.05$.

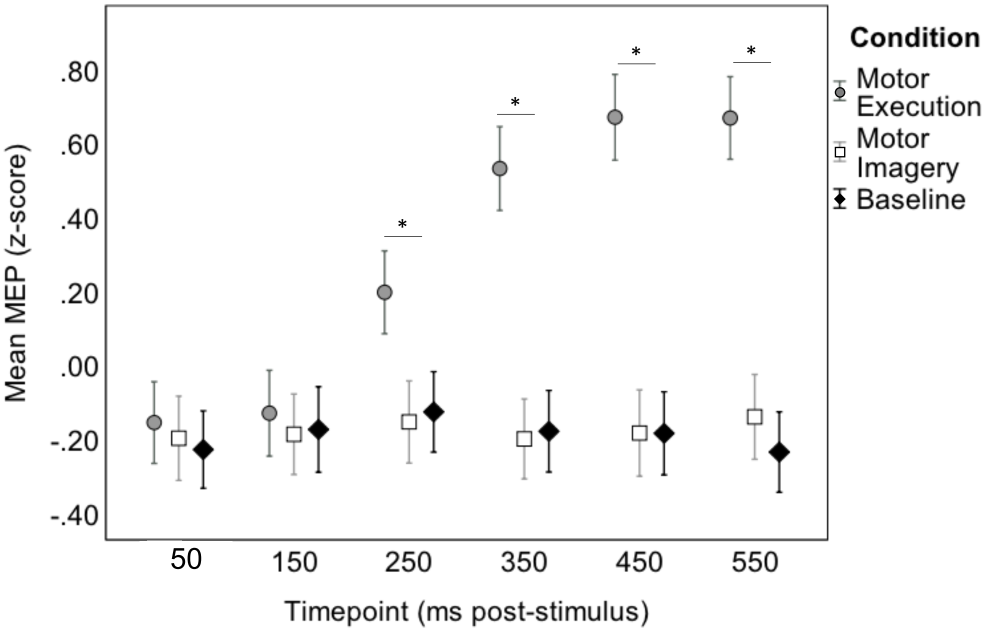


Figure 4.6: Interaction of Condition and Timepoint showing significant differences in MEP *z*-scores across effectors. Error bars indicate +- 1SEM. * = significant at $p = 0.05$.

were significant, suggesting that this trend was likely the same across effectors. Since there was no difference between effectors (either as a main effect or interaction effect) after standardisation around the mean (*z*-scores), the raw data was used for each effector in independent analyses.

4.2.2.3 Lip data

A 3 x 6 repeated-measures ANOVA comparing Condition (motor execution, motor imagery, baseline) and Timepoint (50, 150, 250, 350, 450, 550ms) was run on the lip data. The analysis of the lip MEPs showed a main effect of Condition ($F(1.199, 22.772) = 30.99$, $p < .001$, $\eta_p^2 = .620$), with contrasts revealing motor execution MEPs to be significantly larger than motor imagery MEPs ($MD = 3.512$mV·ms, $SE = 0.582$mV·ms, $p < .001$) and baseline MEPs ($MD = 3.583$mV·ms, $SE = 0.647$mV·ms, $p < .001$), but with no significant difference found between the motor imagery and baseline conditions ($MD = 0.070$mV·ms, $SE = 0.233$mV·ms, $p = 1.000$), as per the z-score analysis. There was also a main effect of Timepoint ($F(2.864, 54.422) = 10.89$, $p < .001$, $\eta_p^2 = .364$), with contrasts revealing MEPs at 50ms to be significantly smaller than at 350ms ($MD = -1.526$mV·ms, $SE = 0.418$mV·ms, $p = .025$), 450ms ($MD = -2.287$mV·ms, $SE = 0.538$mV·ms, $p = .006$) and 550ms ($MD = -2.234$mV·ms, $SE = 0.478$mV·ms, $p = .002$). This main effect was driven by a significant interaction between Condition × Timepoint ($F(3.883, 73.778) = 12.63$, $p < .001$, $\eta_p^2 = .399$). Specifically, the MEP size only increased with time in the motor execution

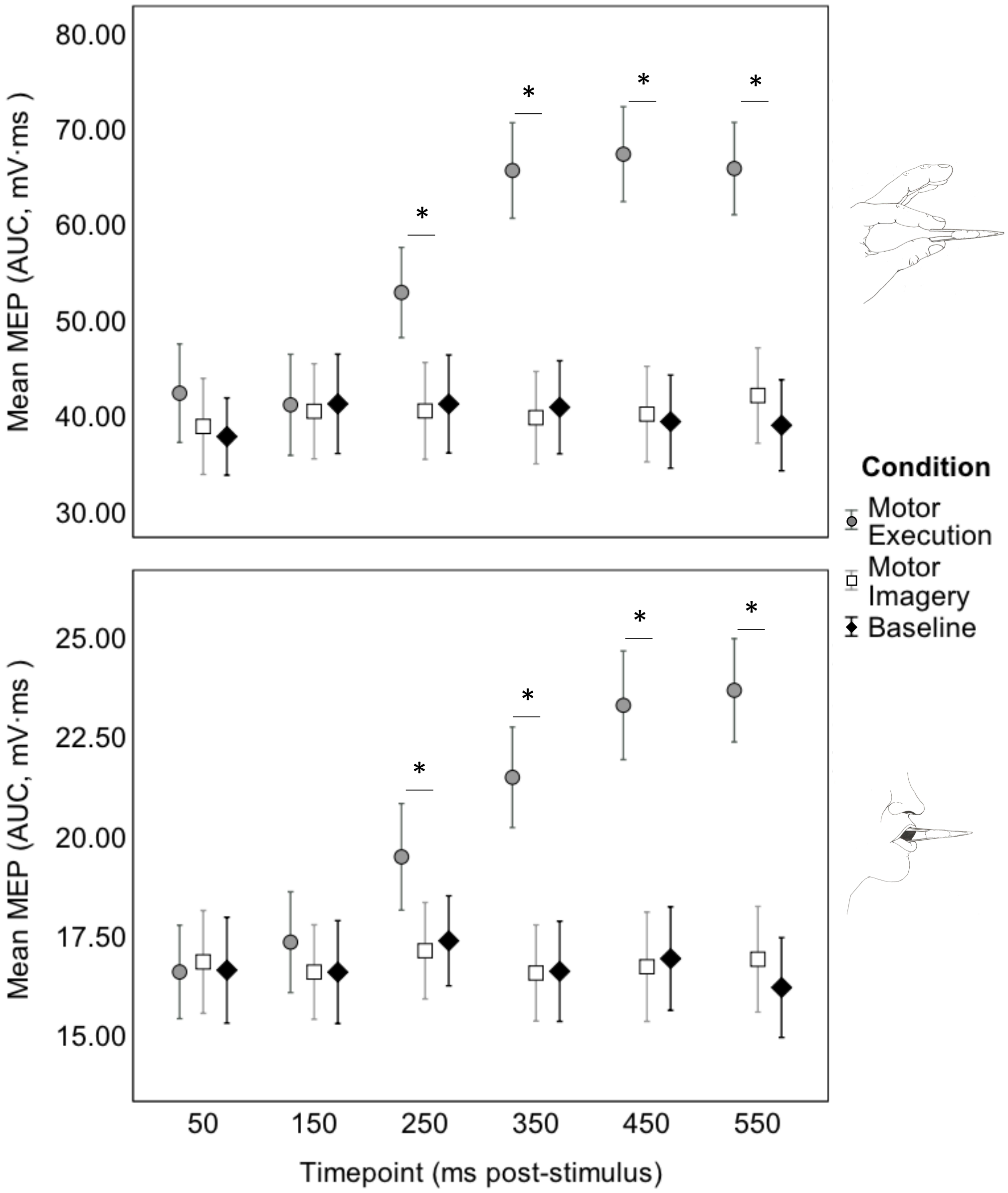


Figure 4.7: Interaction MEP AUC plots for Hand (top) and Lip (bottom) over time point. Note the scale of the y-axis is not equivalent, though the trends are similar. Error bars indicate +-1SEM. * = significant at $p = .05$

condition but not in the motor imagery or baseline conditions (see Figure 4.7). Planned follow-up tests confirmed that while there were no differences across timepoints in the motor imagery and baseline conditions, there were significant differences across timepoints in the motor execution conditions, with MEPs at 50ms significantly smaller at than at 250ms ($t(19) = -3.89$, $p = .001$), 350ms ($t(19) = -4.54$, $p < .001$), 450ms ($t(19) = -5.466$, $p < .001$), and 550ms ($t(19) = -6.59$, $p < .001$). In other words, MEP size only changed when participants actually squeezed the tweezers with their lips, and not when they only imagined doing so.

4.2.2.4 Hand data

A 3 x 6 repeated-measures ANOVA comparing Condition (motor execution, motor imagery, baseline) and Timepoint (50, 150, 250, 350, 450, 550ms) was run on the hand data. There was a main effect of Condition *(F(*1.120, 21.272) = 46.08, *p* <.001, η_p^2 = .708), indicating that motor execution MEPs were significantly larger than motor imagery MEPs *(MD* = 15.549mV·ms, *SE* =2.285mV·ms, *p* < .001) and baseline MEPs *(MD* = 15.944mV·ms, *SE* =2.267mV·ms, *p* < .001), but with no significant difference found between the motor imagery and baseline conditions *(MD* = 0.396mV·ms, *SE* = 0.638mV·ms, *p* = 1.000), as per the z-score analysis. There was also a main effect of Timepoint *(F(*2.424, 46.062) = 14.49, *p* < .001, η_p^2 = .433), with contrasts revealing MEPs at 50ms to be significantly smaller than at 250ms *(MD* = -5.186mV·ms, *SE* = 1.116mV·ms, *p* = .003), 350ms *(MD* = -9.089mV·ms, SE =1.421mV·ms, *p* < .001), 450ms *(MD* = -9.282mV·ms, *SE* =1.702mV·ms, *p* < .001) and 550ms *(MD* = -9.301mV·ms, *SE* = 2.008mV·ms, *p* = .003). As was the case for the lip, this was once again driven entirely by a significant interaction between Condition × Timepoint *(F(*3.504, 66.572) = 15.16, *p* <.001, η_p^2 = .444), see Figure 4.7. Hand MEPs only increased with time in the motor execution condition but not in the motor imagery or baseline conditions. Follow-up tests confirmed that while there were no differences across timepoints in the motor imagery and baseline conditions, there were significant differences across timepoints in the motor execution conditions, with MEPs at timepoint 50 significantly smaller than at 250ms *(t(*19) = -3.84, *p* = .001), 350ms *(t(*19) = -5.00, *p* < .001), 450ms *(t(*19) = -5.86, *p* < .001), and 550ms *(t(*19) = -5.42, *p* < .001). In other words, MEP size only changed when participants actually squeezed the tweezers by hand, and not when they imagined doing so, as per the lip condition.

4.3 Discussion

4.3.1 General discussion

This study aimed to examine whether primary motor cortex was facilitated for speech motor imagery by examining a simple motor imagery task in lip and hand muscles. A secondary aim was to describe the time course of both motor imagery and motor execution in speech and hand muscles. In experiment 4, participants performed a speech or manual action, imagined performing the action, or kept pressure at a baseline level. The hypotheses for this experiment were as follows (see 4.1):

(1) Motor execution conditions was predicted to show significantly larger facilitation than imagery and baseline conditions, since M1 is most active;
(2) Motor imagery conditions was predicted to show significantly larger facilitation than baseline conditions, since M1 is active above baseline due to simulation processes;

The results showed that M1 showed facilitation during action execution for both effectors, but there was no evidence supporting the prediction that mental imagery involves M1. The pattern of increasing MEPs in the results for lip and hand during action execution followed a comparable time course, but differences in area-under-the-curve of the MEPs were more pronounced for hand muscles.

The effect of motor execution in both lip and hand muscles showed that MEPs successfully captured M1 facilitation during execution of actions, with the time course showing the expected increase between 150-350ms, plateauing thereafter. Compared to the lip, it might appear that the hand showed changes in reaction times more quickly given that it was different to baseline at an earlier timepoint, though this is likely to be a result of lip MEP changes being overall smaller rather than a result of actual reaction time differences, as the hand action likely engages a larger pool of motor units relative those engaged in lip.

The results showed no evidence of cortical facilitation for the imagery condition compared to the baseline condition for either effector. This result was unexpected as previous studies focusing on hand actions (Hashimoto & Rothwell, 1999; Hyde et al., 2017) reported cortical facilitation during imagery. Both of those studies used tasks which are broadly comparable to the task used in the present study, namely a mental rotation task and a mental flexion task. However, it should be noted that Hyde and colleagues. only reported an increase in MEP amplitudes for imagery compared to baseline for a subset of their participants (namely the ones that were classified as using motor imagery to a higher extent).

The lack of an increase in MEPs related to motor imagery in these results might be attributable to task effects. Specifically, it could be the case that the experiment 4 task did not sufficiently encourage participants to engage in motor imagery. The task was deliberately simplified to allow for a direct comparison across hand and lip muscles. Studies that use a more complex task tend to report larger increases in MEPs, for instance a study by Roosink and Zijdewind (2010) did find such a modulation of MEPs. In this study, the authors measured the difference in corticospinal excitability between action observation and motor imagery processes to infer the usefulness of these processes in motor rehabilitation. The authors found significant differences between active action observation and motor imagery conditions, with active action observation showing greater MEPs than simple or complex motor imagery conditions (which are not between them significantly different). As a result of their statistical analysis in which the rest condition mean and *SD* was used to obtain *z*-scores for the experimental conditions, the rest condition was not included in the multi-level analysis, and the authors did not claim that the motor imagery conditions were significantly different to the rest condition. However the authors did report significant differences between simple and complex actions. Based on the results for a complex task reported in Roosink and Zijdewind and the lack of a clear result for simple tasks reported in Hyde et al., (2017) it is possible that the task in experiment 4 was too easy to encourage participants to engage in motor imagery.

With respect to the time course results, it was expected that earlier timepoints would be associated with smaller MEPs than any later timepoints. Data from both effectors showed the expected pattern of increasing MEP amplitudes for later time points, which was likely due to motor preparation and motor response. This increase over timepoints addressed the secondary aim and clarified the time points post-stimulus at which motor excitability was notably modulated during motor processes. Follow-up analyses showed that this increase over timepoints was a result of motor execution only, and not motor imagery which showed no difference to the baseline at any timepoint.

There was only limited evidence of a general positive correlation between the VISQ scores and the MEP data across participants for either effector (or condition): specifically, only one factor of the VISQ correlated with MEP size in the lips, and the strength of the relation and its meaning is dubious (see Appendix A). Although this finding may at first glance reveal something about how levels of experienced dialogic inner speech are related to lip motor excitability, there are several caveats to be considered: the most important issue is that the range of condensed inner speech involved does not span the whole scale, but in fact has a range between scores of 2 (‘Possibly does not apply to me’) and 4 (‘If anything, applies to me slightly’). Scores in this range are therefore not strongly determinative of inner speech use – if responses spanned the whole scale from 1-6 but clustered around points 4-6 (i.e., ‘possibly applies’ to ‘certainly applies’) the correlational findings would be more indicative of a relationship. Additionally, these scores are self-reported rather than measured, which could have led to self-report biases.

In sum, TMS can be used to measure changes in cortical excitability associated with overt movement in simple tasks, but TMS may not be suitable for measuring motor imagery of movement in simple tasks. There are several possible reasons for the lack of an increase in MEPs associated with motor imagery compared to the baseline condition. The first could be that MEP data can index larger changes in cortical activity for active movement, but not for subtler changes, such as imagery of such movement. While certain studies are able to show increases in blood flow to motor regions during imagery (see Munzert et al., 2009, p. 308), including articulation imagery, direct measurement via TMS is more difficult to achieve and is perhaps dependent on effort.

It could also be the case that participants did not engage in motor imagery as requested, which would yield results similar to those above. A number of participants pointed out that while the task was intuitive, imagery was not always attained without difficulty. Another reason could be that while active muscle movement activates corresponding motor regions, imagery of muscle movement does not. In this case, the results would contradict several imaging studies that have found action imagery activates motor regions, but would fit with a number of TMS studies that have found no increase in EMG in motor imagery using simple tasks, (e.g., Fadiga et al., 1998; Tremblay et al. ,2001; and Yahagi et al., 1996). One final possibility is that the premotor cortex and SMA actively inhibit primary motor cortex in motor imagery, but not motor execution, precisely to ensure that no overt movement occurs. This may in turn cancel out the activation of primary motor cortex as it is engaged in motor simulation. While this process would affect MEP size negatively, the fMRI Blood-Oxygen-Level Dependent (BOLD) signal would not dissociate between deactivation and activation, potentially explaining studies that show premotor area activation (e.g., Deiber et al., 1998; Kasess et al., 2008). Additionally, there may be inhibition further downstream (e.g., the brainstem or spinal column) which would not be observable in the MEP.

4.3.2 Limitations

The design of the study presented certain limitations. As it is not feasible to consistently obtain lip MEPs without muscle contraction (see Cattaneo & Pavesi, 2014), participants were required to perform active isometric contraction throughout the task. As a result, participants effectively performed two simultaneous tasks, the motor execution and imagery task, and the isometric contraction task. As a result, the EMG signal could have masked underlying subthreshold voluntary activity. This issue could be addressed in future studies, e.g., by attempting a replication of the hand task without voluntary contraction, so that the consistency of task results between active and relaxed muscles can be further examined. Another possible solution is the use of a specialised population with regard to lip muscles, e.g., woodwind and brass instrumentalists, to explore whether a lower threshold or indeed the use of resting motor threshold is possible. In addition, this specialist population may already routinely engage in motor imagery of lip movements through mental rehearsal, which could be exploited in an imagery task.

One other limitation is the use of a very simple task over the use of a more complex task. While the task here was designed specifically to encourage similar imagery in hand and lip muscles, the (imagined) action is not sufficiently speech-like to make statements about the use of articulatory motor imagery in motor imagery of speech. In this sense, the task used was intended to be a steppingstone towards a more complex task.

The correlation analyses showed that the lip data showed an expected correlation between pre-pulse window AUC and MEP AUC, but surprisingly this effect was not found for the hand data. It is possible that this is related to the high threshold that was set for the hand, at 0.5mV. A follow-up study might benefit from using a lower threshold, e.g., 0.2mV.

The VISQ represented another issue since for the factor showing a relationship with MEP size, participant responses were clustered around the responses that indicated uncertainty. Since the VISQ does not include an option equivalent to 'not applicable' or 'neither applies nor does not apply' participants may have clustered their responses around the midpoint (3-4) instead. Although it requires changing an established questionnaire, it might prove fruitful to include a 'neither applies nor does not apply'.

Finally, this experiment is primarily useful in indexing M1 excitability via motor units and does not clarify the processes that take place prior to M1 involvement. For instance, any role for DLPFC or pre-SMA in inhibition of M1 remains unclear, and the experiment 4 design does not allow for disambiguation between a lack of M1 involvement and inhibited activity in M1.

4.3.3 Implications

This experiment explored the use of MEPs in measuring the excitability changes in M1 during imagery of simple actions in hand and lip muscles. The use of MEPs to address these questions specifically in lip muscles, using both execution and imagery conditions and recording the time course of excitability changes, represents a combination of research techniques to deliver a novel approach towards investigating motor imagery. Given the limited usability of MEPs (i.e., they can be used to index M1 activation and nerve innervation along the motor pathway, but not other cortical regions), this experiment does not impact on those models of motor simulation that do not necessarily implicate M1. Specifically, while the lack of increase in MEP size in the imagery tasks suggests that M1 is not involved in low-level imagery, this result does not provide direct evidence for the hypothesis that M1 is not involved in imagery at all. The results also do not serve to suggest that higher-level imagery tasks, such as those suggested in expanded inner speech, do not involve M1 – for this, further research into higher-level phonemic tasks is required.

One important implication is the time course of action as indexed via MEPs. The results are useful in narrowing the focus for action and imagery of action research via TMS, allowing future designs to use only a subset of time points, thus facilitating research design.

4.4 Conclusion

This experiment aimed to establish whether motor imagery of simple tasks differed from a baseline and action execution condition at the cortical level for hand and lip muscles, and to assess the time course of cortical changes associated with each effector. The results showed that while action execution of simple tasks could be indexed using TMS and MEPs, the same was not found for motor imagery of these tasks. Similarly, while it was possible to record how excitability increased over the six time points for the action execution condition, a similar increase was not recorded for the motor imagery condition. The involvement of motor cortex regions during motor imagery may be dependent on task complexity and may not be implicated in the simulation of simple tasks.

In chapter 5, several of the limitations of experiment 4 are addressed. Most importantly, the articulatory task is more complex and physically requires more time to complete. Additionally, the motor execution condition is removed in favour of an action observation condition, in line with previous literature. To investigate the speech musculature further, the tongue is the focus of chapter 5, as it may be more sensitive to articulatory imagery than the lips.

5 Chapter Five

Sections of this chapter were previously submitted for publication to the journal NeuroImage with the title "The involvement of articulatory motor cortex in motor imagery of speech" and have been made available as part of a preprint at https://doi.org/10.31219/osf.io/5k9wa .

5.1 Introduction

As noted in the introduction (section 1.2.1), speech motor imagery is defined as motor imagery processes that involve articulatory motor processing – a subset of processes involved in inner speech. The suggestion that speech motor imagery involves concrete and modality-specific processes, such as low-level articulatory specification and consequent motor planning, has become more prevalent through studies looking at inner speech errors at the phonemic and feature level (Nooteboom, 2005; Oppenheim & Dell, 2008). Additionally there is support from neuroimaging studies using inner speech conditions that note activation of voice-selective phonological brain regions, such as those involved in prosody (temporal voice area, Belin et al., 2018; Perrone-Bertolotti et al., 2014) and voicing (Kell et al., 2017), at least in such cases where expanded (non-condensed, see 1.3.1) inner speech is expected (Grandchamp et al., 2019). A better understanding of the cognitive mechanisms supporting speech motor imagery and their neural loci is an important step to elucidating pathologies associated with inner speech processing, such as auditory verbal hallucinations (Fernyhough, 2004; Jones, 2010; Seal et al., 2004). Another condition that benefits from speech imagery research is ASD (Alderson-Day & Fernyhough, 2015; Whitehouse et al., 2006, see 1.6), especially in terms of understanding the developmental impact of compromised inner speech processes on executive function (Russell-Smith et al., 2014; Williams & Jarrold, 2010).

Speech imagery is thought to be the result of one of two mechanisms, both of which may provide the resultant perceptual experience (i.e., the experience of speech perception without the presence of an overt stimulus) in different ways. The first mechanism is a *direct simulation* mechanism in which speech perception is simulated through memory retrieval of speech percepts – the mental concepts developed through prior perception – in long-term memory, which are used to re-activate relevant sensory cortices (Kosslyn, 2005). The second mechanism is an *indirect simulation* mechanism in which speech motor plans are simulated and somatosensory consequences are estimated on the basis of this simulation (Jeannerod, 1994, 2001; Tian & Poeppel, 2010). The difference between these two mechanisms lies primarily in the transformation from motor to somatosensation in the indirect simulation mechanism, which is absent in the direct simulation mechanism: direct reactivation of sensory cortices without motor involvement should not involve motor areas. Note that this difference does not pertain directly to the overall thesis aim of establishing whether M1 is involved in motor imagery: direct simulation could, but does not necessarily recruit M1 – in fact, Tian and Poeppel argue against this very notion (Tian & Poeppel, 2012). One prediction that can be made is that direct simulation would likely recruit fewer areas than the indirect simulation, since direct simulation requires only memory retrieval and activates sensory cortices, and that indirect simulation would recruit additional integration areas, such as parietal cortex.

Tian and Poeppel (2012) investigated whether such a simulation-estimation mechanism is sequential in nature and whether the consequences generated extend beyond the somatosensory into the auditory domain, specifically for speech. The authors revisited the internal forward model framework for speech as proposed by among others Guenther et al. (2006) and Rauschecker and Scott (2009) (see also section 1.2.3), and expanded it so that the simulation-estimation mechanism supports speech in two ways. In overt speech, the simulation-estimation mechanism provides internal auditory and sensory predictions (the predicted auditory and sensory consequences of simulated action), which can be compared with actual auditory and sensory feedback to monitor and correct motor errors in an online and rapid manner.

In inner speech where no motor output is present, the simulation-estimation mechanism still outputs somatosensory and auditory predictions and perceptual consequences, which results in the inner speech experience. As a result, an internal forward model is a necessary component of both overt and inner speech. Tian, Zarate, and Poeppel (2016) further proposed that in the case of motor imagery of speech, articulatory planning areas produce motor plans that would ordinarily be forwarded to primary motor cortex (M1) are instead sent to parietal regions for simulation and sequential estimation processes, resulting in the inner speech experience. Crucially, this simulation-estimation processing pathway excludes M1. In contrast, Lœvenbruck et al., (2018) postulated that even in speech motor imagery, M1 integrates motor programmes computed by auditory and somatosensory regions, but that any action is suppressed via inhibitory commands from anterior prefrontal cortex. The proposed involvement of M1 in speech motor imagery therefore remains unclear. This chapter investigates the involvement of tongue M1 during speech motor imagery, comparing M1 excitability during speech motor imagery, speech hearing and a baseline condition.

A recent neuro-imaging meta-analysis by Hardwick, Caspers, Eickhoff, and Swinnen (2018) on motor imagery of body movements – excluding facial muscles due to a low number of relevant studies – highlighted the shared neural substrates between action observation, motor imagery and motor execution. This meta-analysis showed overlap between cortical regions involved in action observation, motor imagery, and motor execution, including pre-supplementary motor area (pre-SMA), SMA-proper, dorsal and ventral premotor cortex (PMC), and sensorimotor cortex. It also emphasised regions where there are differences in activation between these conditions. For example, motor imagery was consistently associated with greater activation in premotor regions than motor execution or action observation. In contrast, no evidence of M1 involvement was found in motor imagery or action observation, echoing earlier findings by Hétu et al., (2013).

Both studies do note that whole-brain functional Magnetic Resonance Imaging (fMRI) may be suboptimal for investigating motor imagery, due to the possibly rapid inhibition of M1 from other brain areas, such as pre-SMA. A meta-analysis by Munzert, Lorey, and Zentgraf (2009) including neurostimulation studies showed a more divided picture of evidence for M1 involvement in general motor imagery during fMRI (Alkadhi et al., 2005; Kuhtz-Buschbeck et al., 2003; Stippich et al., 2002), but also showed stronger support for M1 involvement by including the transcranial magnetic stimulation (TMS) literature (Kumru et al., 2008; F. Tremblay et al., 2001; Yahagi et al., 1996). However, none of these imagery studies investigate speech specifically.

While Hardwick et al., (2018) show a lack of M1 involvement for general movement, there is a paucity of speech motor imagery studies focusing on M1 involvement during speech imagery in both neuroimaging and neurostimulation research. This remains the case despite the fact that speech imagery represents an omnipresent phenomenon in the general population, and that an inability to perform speech imagery may present a fundamental hurdle for executive function (Baddeley et al., 2001; Hill, 2004).

While the number of imagery studies using neurostimulation research methods is low, several studies have shown that action observation increases excitability in the associated area of M1. For instance, Fadiga, Fogassi, Pavesi, and Rizzolatti (1995) collected motor-evoked potentials (MEPs) from hand muscles during four tasks, two of which involved observation of hand movement. MEPs are muscle responses induced through TMS of the primary motor cortex (M1) and that index the excitability of the corticospinal or corticobulbar path(s) (Adank et al., 2017).

An increase in MEP waveform amplitude is thought to reflect an increase in local cortical processing: if an MEP is enhanced during a task, it provides correlational evidence that the task involves the targeted region of M1 (Devlin & Watkins, 2007). For both hand movement conditions, MEPs were found to be significantly larger than the non-hand movement conditions. Other studies have extended this method to measure the excitability of lip motor cortex (Möttönen & Watkins, 2012; Murakami et al., 2013; Nuttall et al., 2016; Watkins et al., 2003) or focused on the engagement of tongue motor cortex while observing speech stimuli (D'Ausilio et al., 2009; Fadiga et al., 2002; Roy et al., 2008).

For instance, Watkins et al., (2003) used visual and auditory speech stimuli to investigate the excitability of lip motor cortex. Participants were presented with lip-related speech stimuli in auditory and visual modalities, auditory white noise, or visual stimuli of eyes, which served as control stimuli. MEPs were collected from bilateral M1 lip areas and left-hemisphere M1 hand area only. The results showed that hand MEPs did not differ between conditions, but that left M1 lip MEPs were significantly larger for the auditory and visual speech stimuli relative to the non-speech stimuli.

Moreover, Fadiga et al., (2002) measured MEPs from tongue muscles while Italian participants listened to Italian words and pseudo-words that incorporated either the phoneme /r/ (requires considerable tongue involvement) or /f/ (requires little to no tongue involvement). The authors reported increased activation for both types of stimuli when the stimulus word or pseudo-word contained /r/ compared to when it contained /f/. Tongue M1 was facilitated for /r/ and /f/ in the word condition and for /r/ in the pseudo-word condition, relative to a non-speech baseline consisting of bitonal sounds. The difference in results between phonemes showed a somatotopic distinction in speech perception. Additionally, the contrast between meaningful and non-meaningful words suggested that meaning further increases the excitability of speech motor areas. However, while these studies demonstrate that M1 is facilitated somatotopically during observation of both hand and speech-related actions, they do not clarify the use of M1 in motor imagery as the neural substrates between the two processes are only partially shared (Hardwick et al., 2018).

While there are few TMS-MEP studies investigating speech motor imagery, there are a number of TMS-MEP studies that have measured hand M1 excitability during hand action imagery (Guillot et al., 2012; Meers et al., 2020) and hand rotation imagery (Eisenegger et al., 2007; Ganis et al., 2000; Hyde et al., 2017; Tomasino et al., 2005). Eisenegger et al., (2007) used a paradigm where MEPs were elicited during a mental hand rotation task, during a reading aloud task, and during a silent reading task. Although this study included reading tasks and no articulation, MEPs were collected only from hand muscles throughout – therefore no inferences could be made about lip M1 excitability.

Participants performed each task as part of what they were told was training for a later part of the study. In the rotation task, participants were shown Shepard-Metzler shapes (see 2.1) at two different rotational angles and had to decide whether the two shapes matched or if they were different. In the reading aloud and mouthing silently tasks, participants read short weather reports. The authors found that MEPs during the rotation task were significantly larger than baseline MEPs, as well as larger than MEPs in the reading tasks. Importantly, the reading aloud task also produced significantly larger MEPs than the baseline condition, indicating a generalised excitability increase in M1, although mental rotation MEPs remained significantly larger than reading aloud MEPs. Hyde et al., (2017) showed a similar result using a different kind of rotation task, the hand laterality task by Parsons (1994). In the hand laterality task, participants were shown an image of a hand on-screen and were asked to decide whether they saw a left or a right hand (see 2.1). TMS was delivered to left M1 with MEPs collected from right hand muscles. Once again, hand MEPs were larger than baseline MEPs during the rotation task. These studies support the notion that M1 may be involved in simulation of hand rotation, in contrast to what was proposed by Tian et al., (2016).

The aim of the experiments in this chapter is to establish the extent to which articulatory M1 is engaged in motor imagery or speech actions. Only a single study so far has directly investigated motor imagery of speech using TMS to stimulate articulatory M1 regions (see chapter 3), though MEPs are likely useful markers in tracking changes in M1 excitability during speech imagery processes. TMS and MEPs were used in experiment 4 to investigate whether excitability varied during imagery processes of simple articulatory gestures using lip muscles. Participants were asked to perform or imagine performing a simple task that recruited the lips in the same way they are recruited during the production of labial plosives (/p/, /b/). Although MEPs were found to be larger when the action was executed, there was no difference between a performance imagery and a baseline condition, in which participants did nothing. This result was likely due to the lack of complexity of the action: Roosink & Zijdewind (2010) found increased excitability in hand M1 during imagery of complex hand motor tasks, but experiment 4 did not find increased excitability in simple hand motor imagery tasks or simple lip motor imagery tasks. Experiments 5 and 6 used a complex tongue motor task to investigate whether complexity of the task plays an important role in cortical excitability.

A second aim of experiments 5 and 6 was to further explore the time course of motor imagery, which has been studied in several experiments (Fadiga et al., 1998; Ganis et al., 2000; Hashimoto & Rothwell, 1999; Hyde et al., 2017) although less so specifically for speech motor imagery. Ganis et al., (2000) notes that a good estimation of when motor imagery is likely to occur can be derived from magneto-encephalography (MEG) time course literature using a similar task. Tian and Poeppel (2010) presented an MEG task with a similar paradigm. In their fourth experiment (Tian & Poeppel, 2010, p. 3) three equally-spaced auditory cues (1s apart) were presented with participants performing articulation, articulation imagery, hearing or hearing imagery of the syllable /da/ after the three cues (i.e., where the fourth cue should be if the sequence continued). Median peak amplitudes in the MEG waveform signal were found around 1050ms after the last auditory cue: that is, when directed to follow the preceding tempo in their response, participants showed a peak amplitude – an indication of response processing – 50ms after where the response itself was expected.

I chose to use a similar design in experiment 5 (and 6), with some changes which were necessitated due to the use of TMS (rather than the aforementioned MEG). Piloting showed that participants often waited for the TMS pulse to respond, rather than simply responding as normal - a common occurrence as the pulse is a salient auditory and sensory stimulus in itself. Therefore, in addition to the spacing information provided through a 1s distance between cues, I included a final visual-only cue that would limit variability of responses and would focus response times further. This additional visual-only cue allowed me to instruct participants to use the preceding cues for guidance, but to wait until the final visual-only cue was shown to respond.

Tian & Poeppel (2010) also used an articulation condition, which demonstrated articulatory responses between 250 and 550ms post-cue. As previous motor imagery work has shown that the time course of performed and imagined movement are often very similar (Kumru et al., 2008), with motor imagery sometimes faster (Parsons, 1994) and sometimes slower by milliseconds (Tian & Poeppel, 2010), I decided to include two timepoints for the TMS pulse in experiment 5 (and 6): one at 200ms after the visual-only cue, which was likely to capture early responders, and one at 500ms after the visual-only cue, likely to capture later responders.

A number of studies have looked at the differences in motor imagery between simple and complex manual actions, the use of different effectors and the time course of motor imagery processes, but very few have focused on motor imagery of speech actions and to what degree such imagery involves articulatory motor cortex, and when. The aim of this chapter is to elucidate whether the tongue motor cortex is differentially activated when performing imagery of tongue articulation or hearing a stimulus that involves the tongue *relative to* a baseline condition that does not involve the tongue. An additional aim is to better understand the potential differences in tongue motor cortex excitability during motor imagery of speech and hearing of speech. The relevant hypotheses are as follows:

(i) MEP amplitude will be greater in a motor imagery condition than in a hearing condition or baseline condition, since motor imagery is expected to use M1 in action simulation;

(ii) MEP amplitude will be greater in a hearing condition than in a baseline condition, since action observation may lead to activation of M1 cortices due to covert imitation processes;

(iii) Based on time course results from chapter 4, MEP differences are expected to be larger in a later timepoint (500ms post-pulse) than in an early timepoint (200ms post-pulse) as motor simulation is more likely to occur there.

5.2 Experiment 5

5.2.1 Methods

5.2.1.1 Participants

Twenty participants took part in experiment 5 (19F, 1M, mean age = 22y 6m, *SD* = 2y 4m, age range = 19-27). Handedness was established using the Edinburgh Handedness Inventory (Oldfield, 1971): 18 participants were right-handed and 2 were left-handed. Participants reported no history of neurological or psychiatric disease, and none reported the use of any long- or short-term medication at time of participation. Participants reported no medical conditions, and participants were asked about history of specific repetitive motor activity or muscle disorders, with none reported. All participants completed all conditions on the same day in one session, and all had a minimum of 12 years of education. Participants were screened for contraindications for TMS from guidelines by Rossi et al., (2009). Experiments were undertaken with informed verbal and written consent of each participant, following guidelines set out by the University College London Research Ethics Committee (UREC #0599/001). All research was carried out in accordance with the Declaration of Helsinki.

5.2.1.2 Materials

To collect electromyographical data, two electrodes were attached to a tongue depressor to form a mouthpiece against which participants pressed their tongue. The ground electrode was placed on the right temple. Prompts were displayed on a 21.5" computer monitor while participants sat ~70cm away from the screen. Prompts consisted of representations of an ear in a grey speech bubble, a mouth in a grey thought bubble, and a grey circle, all presented on a black background (Figure 5.1). Fixation crosses and on-screen text were white, presented on a black background. Auditory stimuli were presented at a volume of 70dB SPL, as established for comfort during an earlier piloting phase of the experiment.

5.2.1.3 Procedure

Participants first watched a video showing a trained phonetician providing spoken instructions on how to articulate the target speech sound – a voiceless alveolar plosive followed by an elongated voiced apical trill (/tr/) – by stating the following: "Place your tongue behind your teeth, as if to say /t/. Then, relax the tip of the tongue and let the air flow past it so that the tip of the tongue vibrates against the back of your teeth." The video then showed the phonetician pronouncing the cluster four times, after which the video ended. Participants could practice and watch the video up to five times. One of the exemplars given by the phonetician was also used as the auditory stimulus in the hearing condition (length = 1150ms). The experimenter gave advice on how to pronounce the sound when requested. After a maximum of a 15-minute period, the experimenter stopped the practice session and continued to the testing session.

Next, the tongue electrodes (Ag/AgCl, diameter 10mm) and mouthpiece were fitted. Participants were instructed to position the mouthpiece so that the positive electrode was placed ~1cm behind the tip of the tongue, and the negative electrode was placed 1cm behind the positive electrode (this inter-electrode distance was set when the electrodes were attached to the mouthpiece by the experimenter and was not changeable by the participant). Participants trained for several minutes to press up with their tongue blade and dorsum, thereby exerting pressure on the electrodes. This setup was used to establish that the participants could comfortably hold tension in the tongue muscle that amounted to 20% of their maximum voluntary contraction.

Participants were told that they should be as consistent as possible when pressing against the electrodes, and that the experimenter would notify them if the pressure fluctuated excessively (e.g., an amplitude increase or decrease of ±100 μV above or below the pre-established threshold of 20% of maximum voluntary contraction). Participants were not able to view their MEPs, but the experimenter was able to observe the EMG trace. Participants then performed a practice version of the experiment, which did not include delivery of TMS pulses but was otherwise identical to the main TMS experiment.

As shown in Figure 5.1, each trial began by displaying a white fixation cross for 2000ms. The condition prompt was then shown for 1000ms. Next, participants were shown a countdown displaying the numbers 3,2,1 for 1000ms consecutively, accompanied by a tone (audio-visual cues). After another 1000ms, a white asterisk was shown, which was the visual-only cue to perform the action associated with the condition while

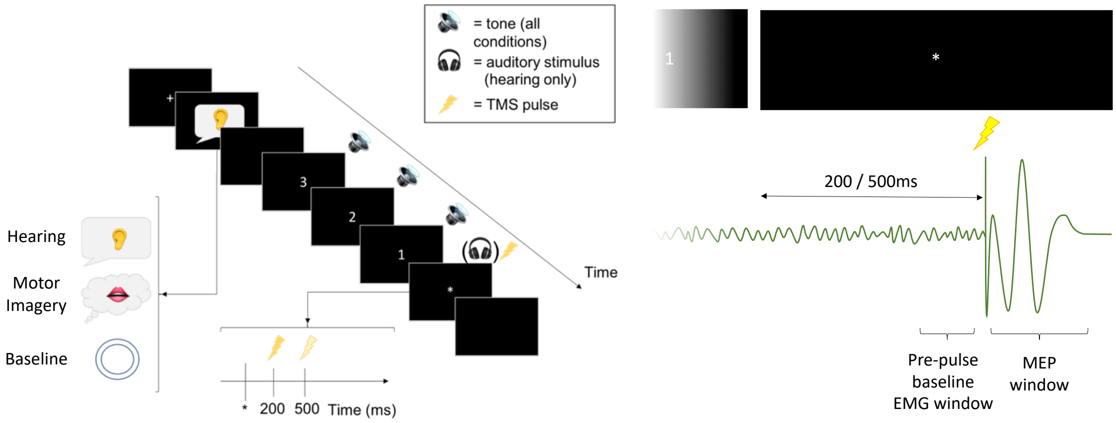


Figure 5.1: Single trial in experiment 5 showing baseline and MEP window relative to EMG trace and prompts.

maintaining 20% of maximum voluntary contraction. There were 3 conditions: (a) imagine saying /tr/, in which participants were asked to imagine saying /tr/ while maintaining contraction; (b) hearing /tr/, in which participants listened to the auditory stimulus /tr/ while maintaining contraction; (c) do nothing, in which participants performed neither action and maintained contraction. A TMS pulse was delivered either 200ms or 500ms after the visual-only cue. The screen then turned black and a new trial began. The practice version of the experiment was made up of 12 trials, showing each condition four times, interleaved.

The main experiment involved eight blocks of 30 trials (8 x ~3min), for a total of 240 trials/MEPs. Each block was followed by a 1min resting period, except the fourth block, which was followed by a rest period with a duration determined by the participants. Note that this flexible rest period had a minimum duration of one minute as this timepoint marked the mid-way point of the experiment. After the experiment had finished, the participant was run through a battery of tests including the Montreal Cognitive Assessment (Nasreddine et al., 2005) to check for mild cognitive impairment, the Varieties of Inner Speech Questionnaire (McCarthy-Jones & Fernyhough, 2011) to have a qualitative and quantitative record of how participants experience inner speech, and an audiometry test (frequencies of 1000Hz, 2000Hz, 4000Hz, 8000Hz, 250Hz and 500Hz in each ear in accordance along guidelines of the British Society of Audiology, 2012) to ensure participants had hearing within the normal range. MEPs were collected over a 40min period and the total testing session ran for about two hours.

5.2.1.4 Transcranial magnetic stimulation

As in experiment 4, individual motor thresholds were used for each participant, meaning that a minimum amplitude for tongue MEPs was set at which 5 out of 10 pulses elicited MEPs (see 4.2.1.4 and Groppa et al., 2012). However, the lack of studies investigating speech motor imagery to tongue M1 poses the problem that certain variables of TMS-MEP setup are not well-established when investigating the tongue muscles. First, several approaches to MEP collection have been described in the literature. Some have measured MEPs with a TMS intensity based on the measurement of active motor thresholds (aMT) (Ghezzi & Baldini, 1998; Muellbacher et al., 1994; Neef et al., 2011; Rödel et al., 2003; Whillier et al., 2018), while other have set this intensity based on measurement of resting motor thresholds (rMT) (Baad-Hansen et al., 2009; Fadiga et al., 2002; Meyer et al., 1997; Panouillères & Möttönen, 2018; Peter Svensson et al., 2003). Second, since the early guidelines (Rossini et al., 1994), there has been a shift in the minimum amplitude criterion in MEPs, with authors stating during active muscle contraction a minimum amplitude criterion of 100μV (Rossini et al., 2015) or 200μV (Groppa et al., 2012) should be used when determining MT, although certain recent studies increase this even further, using a criterion as high as 1000μV for hand muscles (Möttönen et al., 2010)(rMT) or up to 250-450μV for tongue muscles (Barwood et al., 2013). There is much variation in the literature at the level of using active or resting motor threshold, peak-to-peak amplitude criteria, coil size and shape, and mean MEP size of collected MEPs. An in-depth literature search was performed to find the correct parameters for experiment 5. In preparation for these experiments, SCOPUS was used to find studies collecting tongue MEPs. A SCOPUS search was performed using the string 'MEP AND tongue OR lingual' as well as "motor evoked potential" AND "Transcranial magnetic stimulation" AND tongue OR lingual'. Search 1 yielded 41 results of which 23 were relevant, while search 2 yielded 58 results of which 34 were relevant. All 23 results from search 1 were also present in search 2, so search 2 yielded an additional 11 relevant results. An extensive review of the references in these studies yielded an additional 2 studies. Studies were included if they included healthy subjects, or a healthy control population – patient-only studies were excluded. All data in the table describes baseline data: if a study involved manipulations, the data included in the table is representative of the baseline data pre-manipulation. To be included, a study must have reported MEPs – rTMS-only papers were not included. Studies focusing on intra-operative stimulation of tongue motor cortex were not included. The table of studies can be found in Appendix F.

On the basis of the literature search and the design parameters of previous similar paradigms, guidelines set out in Rossini et al. (2015) were followed and piloting was performed on a select group of participants to ascertain the MEP amplitude that constituted a tongue MEP. On the basis of the overall size and shape of the MEP compared to baseline EMG activity at 20% of maximum voluntary contraction with the mouthpiece, and following discussion with colleagues (Helen Nuttall, Patti Adank, Dan

Kennedy-Higgins, personal communication, February 2019) the decision was made to use an amplitude criterion of 400μV for tongue MEPs.

The literature search also showed a variety of testing intensities used: most studies, whether using rMT or aMT, used testing intensities between 80% and 140% of motor threshold. As is common in TMS studies targeting lip and hand muscles, one testing intensity of 120% aMT was fairly common, but so was a lower intensity of 90-110% of aMT. On this basis it was decided that the experiment would be run twice, with two different testing intensities: experiment 5 proceeded to use a testing intensity of 100% aMT, while experiment 6 proceeded to use a testing intensity of 120%.

Participants were familiarised with TMS after the training session, but prior to the practice version of the experiment. The tongue area was localised using the hot-spot technique: starting from previously-established coordinates (-60, -10, 25, D'Ausilio et al., 2009) and moving the coil radially by 5-10mm to ascertain the precise location at which the most robust MEPs were elicited. The EMG signal was sampled at 5000Hz and amplified by a factor of 1000. The signal was band-pass filtered between 100-2000Hz using a 1902 amplified with analogue-to-digital conversion using a Micro1401 unit (Cambridge Electronic Design, Cambridge, UK) connected to a Windows 10 PC. Frameless stereotaxy (Brainsight, Rogue Resolutions, Montreal, QC, Canada) was used to localise the area of stimulation for the tongue motor cortex, making use of the built-in MNI-152 model (scaled to the participant's head dimensions). This system also collected TMS accuracy throughout the session in the form of target error, angular error and twist error. A figure-of-eight coil (⌀ 70mm) was placed at a 45° angle relative to the sagittal plane, inducing a posterior-to-anterior current flow approximately perpendicular to the lateral fissure. Thresholding was performed using a standard descending-ascending thresholding procedure in which 5 out of 10 MEPs must be elicited (Rothwell et al., 1999). An MEP was defined as having a minimum peak-to-peak amplitude of 400μV. Once active motor threshold was established, testing intensity was set to 100% of aMT.

5.2.1.5 Analysis

A 1040ms sweep was extracted from Spike^2 (Cambridge Electronic Design, Cambridge) for each MEP, with 1000ms preceding the pulse and 40ms following the pulse. From this

1040ms MEP sweep, two time windows were extracted. First, a pre-MEP time window was extracted, lasting from 10ms immediately preceding the TMS pulse to the pulse itself, to ascertain muscle activation before the TMS pulse. Second, a MEP time window containing the MEP was extracted, i.e., 8-35ms after the TMS pulse (see fig 5.1). For analysis area-under-the-curve (AUC) values were used as calculated by rectifying the signal in the MEP window and applying a trapezoidal function to the curve to ascertain the area covered, used as an index of MEP size. This procedure is preferable when MEP waveforms are polyphasic rather than monophasic (Adank et al., 2017), though peak-to-peak amplitudes were also calculated for additional inspection. Mean AUC and peak-to-peak values are reported, though AUC only is used for in-depth analysis.

A linear mixed-effects (LME) analysis was performed on the pre-MEP time window data to check that baseline EMG activity was not different between conditions. If pre-MEP baseline EMG was significantly different across participants and condition, a per-participant one-way ANOVA with factor Condition (motor imagery, hearing, baseline) was run to establish whether the participant could be included in the final analysis. This was done to ensure that differences in increases in MEP amplitude in each condition were the result of cortical processes of that condition, rather than increases in muscle contraction prior to and during the pulse which is known to increase MEP amplitude (Darling et al., 2006). The pre-MEP LME analysis was to be followed by an LME analysis of the MEP data with Participant as random effects, and fixed-effects factors Condition (motor imagery, hearing, baseline) and Time point (200, 500ms). See Figure 5.2 for a flow chart that details the data analysis process. Follow-up *t*-tests were performed where appropriate. All significance values were Bonferroni-corrected for multiple

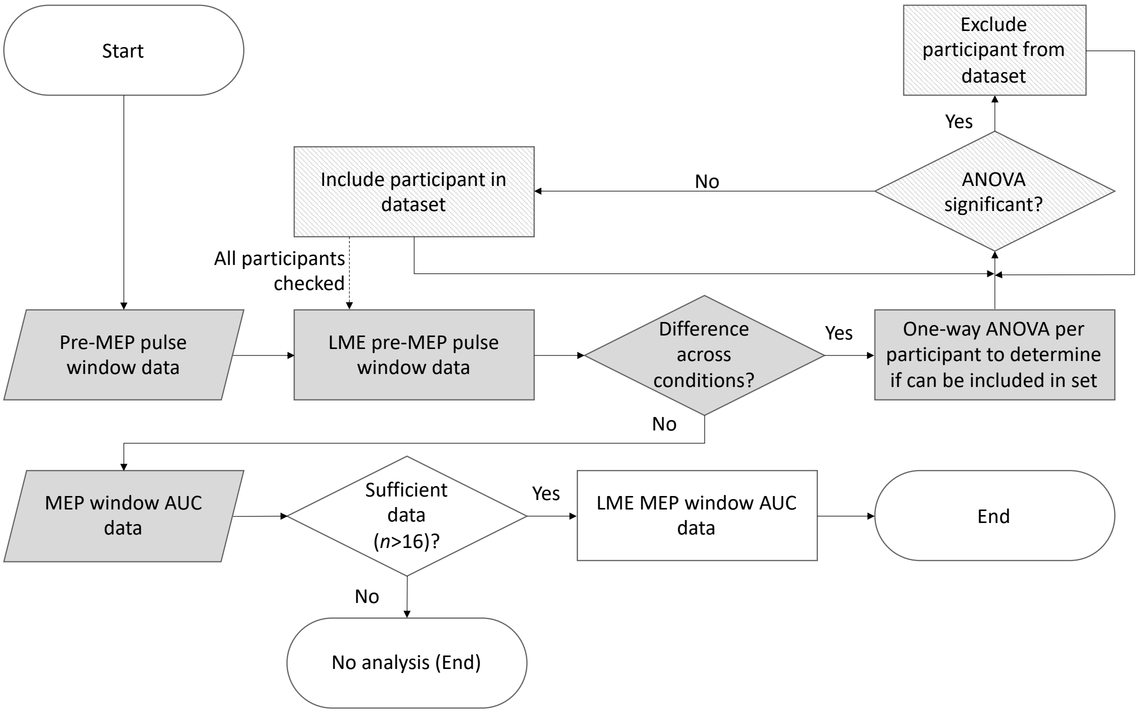


Figure 5.2: Flow chart of data analysis for experiments 5 and 6. LME = Linear mixed effects model, MEP = motor-evoked potential, AUC = area-under-the-curve.

comparisons, where applicable. MEPs were removed from the analysis if they exceeded 3SDs from the overall AUC mean (52 trials). MEPs were further removed if they represented the 1st trial in a block (160 trials), if they exceeded pre-established location error thresholds (16 trials), or if they were manually found to contain noise during the 1040ms sweep (419 trials). In total, 647 trials out of 4800 trials (~13.5%) were judged to be unusable. Participant data was removed from the dataset if fewer than 30 usable MEPs per condition per timepoint were available, resulting in the removal of two participants (P1: range 9-15 MEPs, P2: range 16-19 MEPs), leaving a total of 3997 trials (83.3%) for analysis.

Data were analysed in R (R Core Team, 2020) using a series of generalized linear mixed-effects models (LME) using the `lme4` package (Bates et al., 2015), allowing for the inclusion of random intercepts for individual participants. This LME approach differs from a more traditional MEP analysis, in this case a 3 x 2 x 2 repeated-measures ANOVA, in that it does not standardise the data prior to analysis but accounts for interindividual differences using random effects instead. This provides a better understanding of individual MEP trends and reduces type I error rates for this type of data (Boisgontier & Cheval, 2016). Models were built up iteratively, adding fixed effects and interactions sequentially, and performing likelihood ratio tests following each such addition (Barr et al., 2013). Fixed effects were retained in the final model if they resulted in significantly improved model fit, either per se or in an interaction.

5.2.2 Results

Frameless stereotaxy tracking allowed us to extract in-session localisation data for 18 out of 20 participants (two sets of tracking data were lost due to tracking hardware issues). Average aMT was 53.6% (range 41-73) of stimulator output, which was also the testing output as testing intensity was 100% of aMT.

5.2.2.1 VISQ

Participants rated each question on the VISQ on as scale of not applying to them (1 -3) or applying to them (4-6, see Chapter 2 for more information on the VISQ). Overall participants reported experiencing evaluative/motivational speech most often, followed by dialogic inner speech, as was also the case in previous chapters. Condensed inner speech was less prominent, and participants again stated that they did not hear other people’s voices in their own inner speech (except in two cases, see Figure 5.3).

5.2.2.2 Pre-MEP window

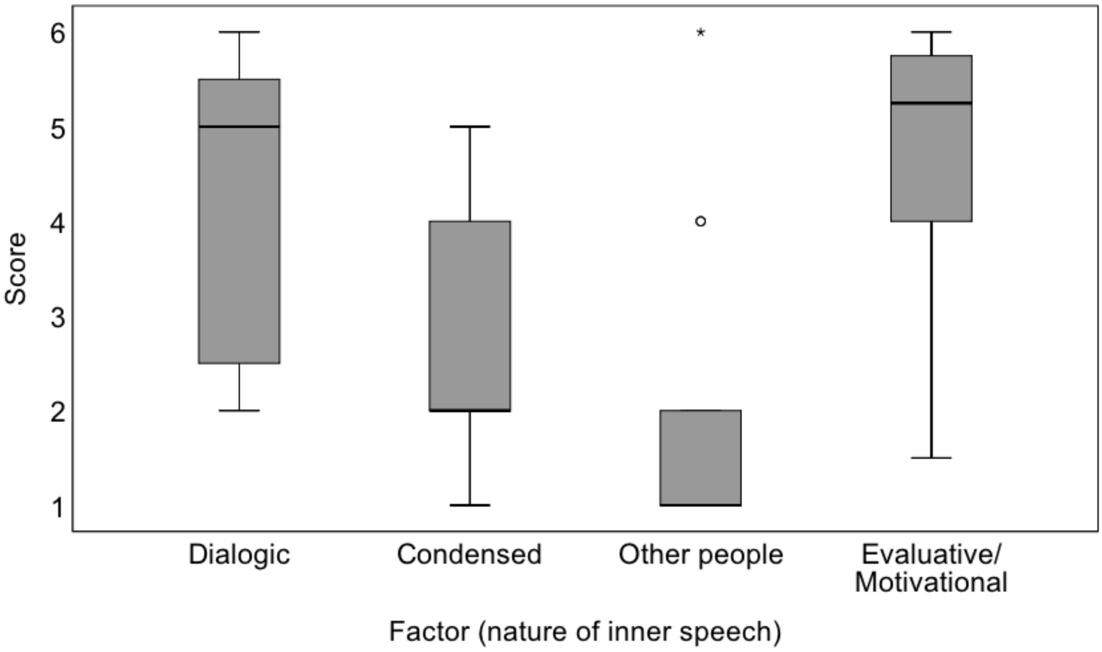


Figure 5.3: VISQ scores per type of inner speech for experiment 5. Scores are amalgamated across five questions that relate to each factor. See Appendix C for the questionnaire. (1 = Certainly does not apply, 6 = Certainly does apply).

A first analysis involved running a linear mixed-effects (LME) analysis on the pre-MEP data to ensure that pre-pulse EMG activity was the same across conditions (see Figure 5.2). This check analysis was necessary to ensure that the MEP analysis reflects cortical excitability rather than simply increased activity through stronger muscle contraction. The check analysis involved the factor Condition (levels motor imagery, hearing, and baseline). Random intercepts were included per participant, with a fixed factor of Condition added. The addition of the fixed factor condition to a baseline model with random intercepts for participants yielded an improved model fit for the pre-MEP EMG data ($\chi^2(2) = 44.398$, $p < .001$), showing that a significant difference in EMG data existed between conditions prior to the TMS pulse having been administered. Follow-up analysis showed a significant difference between the motor imagery and hearing conditions ($t(17.718) = 5.863$, $p < .001$) and between the motor imagery and baseline conditions ($t(17.444) = 7.551$, $p < .001$), but not between the hearing and baseline conditions ($t(17.723) = 0.237$, $p = .813$). Participants provided more contraction prior to the motor imagery condition relative to the other conditions, but there was no difference in contraction force prior to the hearing and baseline conditions.

To account for the differences in muscle contraction across condition, a one-way ANOVA was run on the pre-MEP window data for each participant individually to see how many and which participants showed a pattern of pre-MEP EMG differences across conditions. Eight out of 18 participants showed such a pattern. Table 5.1 shows a one-way ANOVA for these 8 participants combined, as well as the 10 participants left in the dataset, showing clearly the differences of Condition in pre-MEP window data between the two groups. The resultant sample size of 10 participants was deemed to be too low to continue our full-fledged analysis (see Figure 5.2). However, the data was used in a subsequent analysis, see section 5.4.

Table 5.1: Statistics and one-way ANOVA results for experiment 5 (pre-MEP window, 10ms, mV · ms). * = significant at p = .05

One-way ANOVA for n = 8 (significantly different pre-MEP window baselines)

Condition	*M*	*SD*	*n*
Motor Imagery	6.029	6.148	604
Hearing	4.287	3.709	600
Baseline	4.151	4.306	593
Total	4.828	4.912	1797

Source	*df*	Mean Square	*F*	*p*	η_p^2
Corrected Model	2	659.011	28.142	*<.001	0.030
Intercept	1	41784.563	1784.369	*<.001	0.499
Condition	2	659.011	28.142	*<.001	0.030
Error	1794	23.417			

One-way ANOVA for n = 10 (no significantly different pre-MEP window baselines)

Condition	*M*	*SD*	*n*
Motor Imagery	4.388	4.500	730
Hearing	4.167	4.355	729
Baseline	4.273	4.457	741
Total	4.276	4.437	2200

Source	*df*	Mean Square	*F*	*p*	η_p^2
Corrected Model	2	8.900	0.452	0.636	0.000
Intercept	1	40222.295	2042.367	*<.001	0.482
Condition	2	8.900	0.452	0.636	0.000
Error	2197	19.694			

5.2.3 Discussion

Due to unexpected differences between pre-MEP baseline levels of muscle contraction, and the resulting low number of remaining participants, data analysis had to be suspended before investigating the main hypotheses. While a majority of participants (60%) did not exhibit pre-MEP muscle contraction differences, it was sufficient to halt in-depth analysis. However, as experiment 6 was run concurrently, I decided to attempt a unification of data sets once both experiments had completed data collection, while still accounting for differences in testing intensities (100% or 120% of aMT). Whether such a unification would be legitimate depended on whether there would be a difference between participant groups in overall MEP AUC in pre- and post-pulse windows.

5.3 Experiment 6

5.3.1 Methods

5.3.1.1 Participants

Nineteen participants took part in experiment 6 (10F 9M, mean age = 24y 1m, *SD* = 3y 10m, age range = 19-33). Handedness was established using the Edinburgh Handedness Inventory (Oldfield, 1971): 17 participants were right-handed and 2 were left-handed. Participants reported no history of neurological or psychiatric disease, and none reported the use of any long- or short-term medication at time of participation. Participants reported no medical conditions, and participants were asked about history of specific repetitive motor activity or muscle disorders, with none reported. All participants completed all conditions on the same day in one session, and all had a minimum of 12 years of education. Participants were screened for contraindications for TMS from guidelines by Rossi et al., (2009). The experiment was undertaken with informed verbal and written consent of each participant, following guidelines set out by the University College London Research Ethics Committee (UREC #0599/001). All research was carried out in accordance with the Declaration of Helsinki.

5.3.1.2 Materials

The materials for experiment 6 were the same as those used in experiment 5.

5.3.1.3 Procedure

The procedure for experiment 6 was the same as that used in experiment 5.

5.3.1.4 Transcranial magnetic stimulation

The TMS setup for experiment 6 was the same as that used in experiment 5, with the following unique difference: while experiment 5 used a testing intensity of 100% of aMT, experiment 6 used a testing intensity of 120% of aMT. Once active motor threshold was established, testing intensity was set to 120% of aMT.

5.3.1.5 Analysis

The analysis for experiment 6 was identical to the analysis for experiment 5 but is repeated here for convenience and to highlight differences (see Figure 5.2). As above, an LME analysis was performed on the pre-MEP time window data to check that baseline EMG activity was not different between conditions. The pre-MEP LME analysis was to be followed by an LME analysis of the MEP data with participant as random effects, and fixed-effects factors Condition (motor imagery, hearing, baseline) and Time point (200, 500ms). Follow-up *t*-tests were performed where appropriate. All significance values were Bonferroni-corrected for multiple comparisons, where applicable. MEPs were removed from the analysis if they exceeded 3SDs from the overall AUC mean (30 trials). MEPs were further removed if they represented the 1st trial in a block (152 trials), if they exceeded pre-established location error thresholds (55 trials), or if they were manually found to contain noise during the 1040ms sweep (137 trials). In total, 374 trials out of 4560 trials (~8.2%) were judged to be unusable. Participant data was removed from the dataset if fewer than 30 usable MEPs per condition per timepoint were available, resulting in the removal of one participant (P1: range 27-35 MEPs), leaving a total of 3993 trials (87.6%) and 18 participants for analysis.

5.3.2 Results

Frameless stereotaxy tracking allowed us to extract in-session localisation data for 18 out of 19 participants (1 set of tracking data was lost due to tracking hardware issues). Average aMT was 49.4% (range 39-62) of stimulator output, with average testing output 59.2% (range 47-74%) of stimulator output.

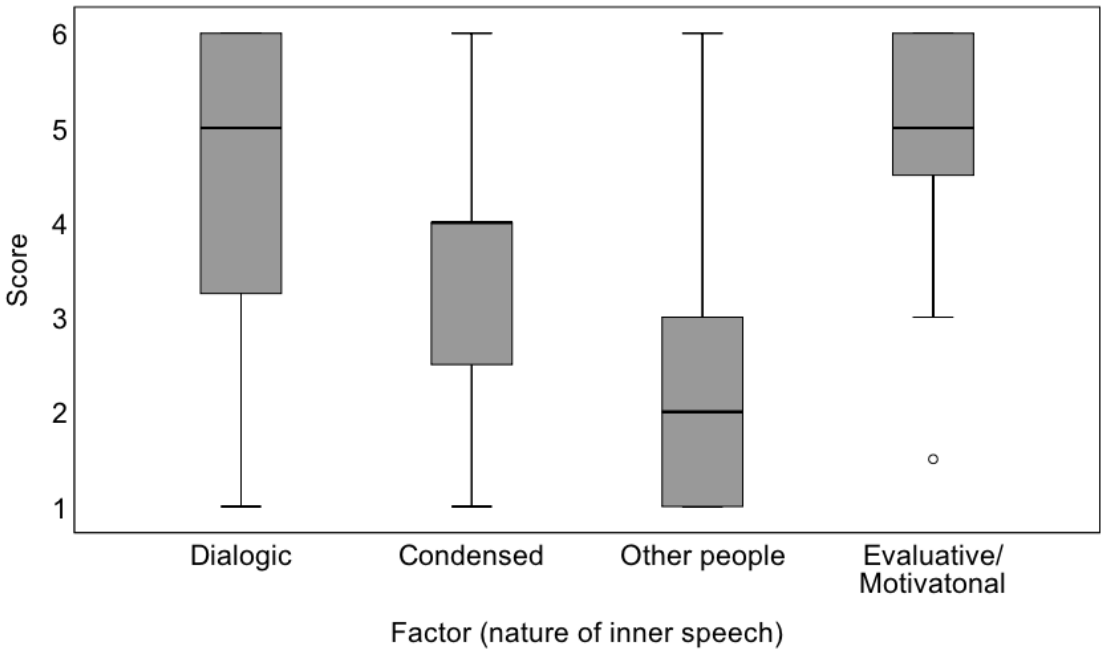


Figure 5.4: VISQ scores per type of inner speech for experiment 6. Scores are amalgamated across five questions that relate to each factor. See Appendix C for the questionnaire. (1 = Certainly does not apply, 6 = Certainly does apply).

5.3.2.1 VISQ

Participants rated each question on the VISQ on as scale of not applying to them (1 -3) or applying to them (4-6, see Chapter 2 for more information on the VISQ). Overall participants reported experiencing evaluative/motivational speech most often, followed by dialogic inner speech, as previous experiments also show. Participants show no specific preference for condensed or expanded inner speech, and overall other people's voices in their own inner speech was the factor with the lowest score (see Figure 5.4).

5.3.2.2 Pre-MEP window

As in experiment 5, a first LME analysis was performed on the pre-MEP data to ensure that pre-pulse EMG activity was the same across conditions. The check analysis involved the factor condition (levels motor imagery, hearing and baseline). Random intercepts were included per participant, with a fixed factor of condition added. The addition of the fixed factor condition to a baseline model with random intercepts for participants yielded an improved model fit for the pre-MEP EMG data ($\chi^2(2) = 14.884$, $p < .001$), showing that a significant difference in EMG data existed between conditions prior to the TMS pulse having been administered. Follow-up analysis showed a marginally significant difference between the motor imagery and hearing conditions ($t(17.894) = 1.944$, $p = .052$) and between the motor imagery and baseline conditions ($t(17.894) = 3.861$, $p < .001$), as well as a marginally significant difference between the hearing and baseline conditions ($t(17.879) = 1.191$, $p = .055$). Participants provided more contraction prior to the motor imagery condition relative to the baseline condition, with additional trends in differences in contraction force between hearing and baseline conditions.

As in experiment 5, a one-way ANOVA was performed on the pre-MEP window data for each participant individually to see how many and which individuals showed a pattern of pre-MEP EMG differences. Seven out of 18 participants showed such a pattern (see Table 5.2). The resultant sample size of 11 participants was deemed to be too low to continue our full-fledged analysis, similar to experiment 5.

Table 5.2: Statistics and one-way ANOVA results for experiment 6 (pre-MEP window, 10ms, mV · ms). * = significant at *p* = .05

One-way ANOVA for *n* = 7 (significantly different pre-MEP window baselines)

Condition	*M*	*SD*	*n*
Motor Imagery	3.428	1.973	517
Hearing	3.166	2.325	528
Baseline	2.761	1.685	531
Total	3.115	2.029	1576

Source	*df*	Mean Square	*F*	*p*	η_p^2
Corrected Model	2	59.288	14.657	*<.001	0.018
Intercept	1	15322.403	3788.008	*<.001	0.707
Condition	2	59.288	14.657	*<.001	0.018
Error	1573	4.045			

One-way ANOVA for *n* = 11 (no significantly different pre-MEP window baselines)

Condition	*M*	*SD*	*n*
Motor Imagery	4.002	4.341	802
Hearing	3.870	4.050	805
Baseline	3.740	4.074	810
Total	3.870	4.156	2417

Source	*df*	Mean Square	*F*	*p*	η_p^2
Corrected Model	2	13.854	0.802	0.449	0.001
Intercept	1	36211.816	2095.863	*<.001	0.465
Condition	2	13.854	0.802	0.449	0.001
Error	2414	17.278			

5.3.3 Discussion

Similar to the data from experiment 5, data in experiment 6 showed pre-MEP baseline differences in levels of muscle contraction, and as the number of participants remaining was insufficient further analysis was suspended. However, as previously discussed, if it is the case that there is no difference between the experiment 5 and 6 participants groups in overall MEP AUC data, it would be justified to combine the data from both experiments and analyse the data as such.

5.4 Experiment 5 & 6 - unified data analysis

To investigate whether we could collapse the two studies' datasets to improve our sample size, we performed independent samples *t*-tests on the MEP AUC data between groups A (participants from experiment 5, 100% aMT) and B (participants from experiment 5, 120% aMT), excluding those participants with insufficient data (3 out of 39, leaving 18 per group). An independent samples *t*-test was run to investigate whether there was a difference between the two groups when all data were collapsed over condition as well as timepoint. Levene's test for Equality of Variance was significant (*F*(2,34) = 6.02, *p* = .019) so equal variance was not assumed and *t*-tests were carried out using the Welch-Satterthwaite method. Using Welch's *t*-test maintains the assumption of normal distribution but accounts for unequal sample distribution variance. The *t*-test showed that there was no difference in MEP AUC between group A (*M* =16698 μV·ms, *SD* = 8049 μV·ms) and group B (*M* =22650 μV·ms, *SD* = 16788 μV·ms); *t*(24.669) = -.1376, *p* = .181). To ensure there were no effects even within the conditions, I also ran an independent samples *t*-test using per-condition data (see Table 5.3). These data show that while group A had a lower mean amplitude than group B, there was no significant difference between any of the conditions, across groups. Participants from both groups were therefore combined from this point forward and analyses conducted on this pooled dataset (21 participants, 7990 trials).

Having assessed and aggregated the data as appropriate, an LME analysis of the MEP data was performed with participant as random effects, and fixed-effects factors condition (motor imagery, hearing, baseline), time point (200, 500ms) and in the first instance, group (A, B). Follow-up *t*-tests were performed where appropriate. All

Table 5.3 Independent *t*-tests between groups in experiment 5 (100% aMT) and experiment 6 (120% aMT).

Groups	Levene's Test		*t*-test for Equality of Means				
	F	*Sig.*	*t*	*df*	*Sig.*	*MD*	*SED*
100% - 120%	6.037	0.019	-1.376	24.669	0.181	-5.95228	4.32465
AI (100% - 120%)	3.051	0.09	-1.293	27.99	0.207	-5.77713	4.4674
H (100% - 120%)	7.993	0.008	-1.411	23.327	0.171	-6.3284	4.48381
B (100% - 120%)	7.528	0.01	-1.358	23.828	0.187	-5.74873	4.23463

significance values were Bonferroni-corrected for multiple comparisons, where applicable.

5.4.1.1 Linear mixed-effects analysis

Mean peak-to-peak amplitude across factors was 1305μV (*SD* = 1154μV). Mean peak-to-peak amplitude for motor imagery was 1036μV (*SD* = 1185μV), for hearing mean amplitude was 1041μV (*SD* = 1198 μV) and for baseline mean amplitude was 1014μV (*SD* = 1144μV). Mean AUC across conditions and groups was 19943μV·ms (*SD* = 19348μV·ms). For motor imagery, mean AUC was 20023μV·ms (*SD* = 18787μV·ms), for hearing mean AUC was 19993μV·ms (*SD* = 19851μV·ms) and for baseline mean AUC was 19814μV·ms (*SD* = 19402μV·ms).

Random intercepts were included per participant, while fixed effects including Condition (levels motor imagery, hearing, baseline), Time point (levels 200, 500ms) and Group (levels group A, B) were added iteratively. R notation has been added for clarity and is typeset in `Courier`. The baseline model (`MEP_AUC ~ (1|Participant)`) was first compared to a model also including Group. The addition of Group to the baseline model (`MEP_AUC ~ Group + (1|Participant)`) did not improve model fit ($\chi^2(1)$ =

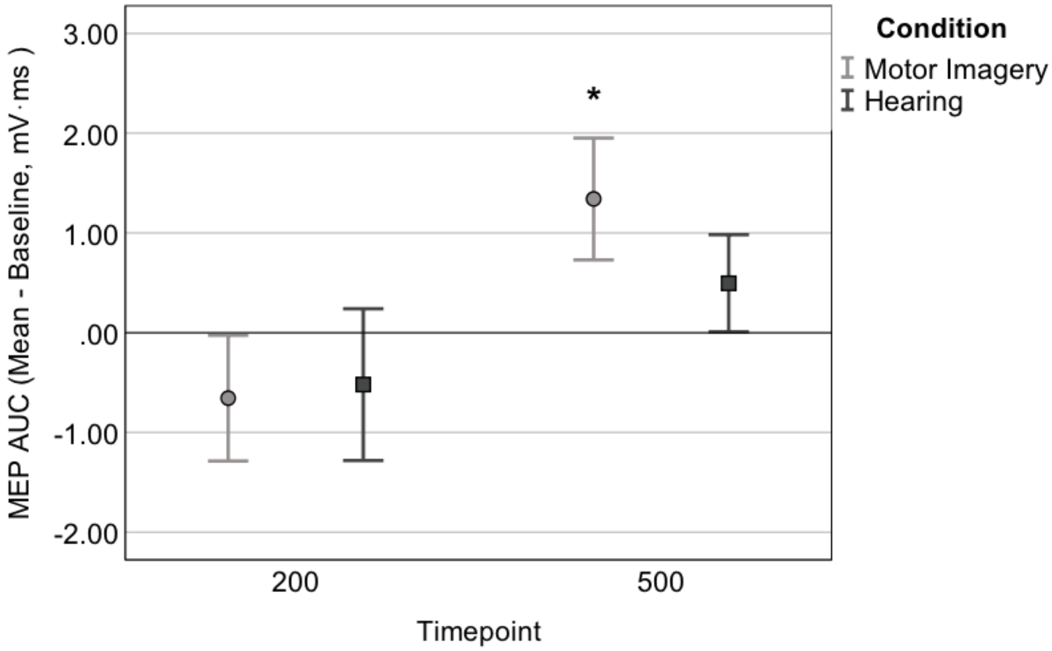


Figure 5.5: MEP AUC values relative to average baseline values for included participants (*n.= 21)*. Scores were calculated by subtracting mean AUC of baseline per condition per time point from mean AUC of motor imagery/hearing per condition per time point. * shows significance (*p* = 0.05) from baseline (0.00, black horizontal line). Error bars indicate +- 1 SEM.

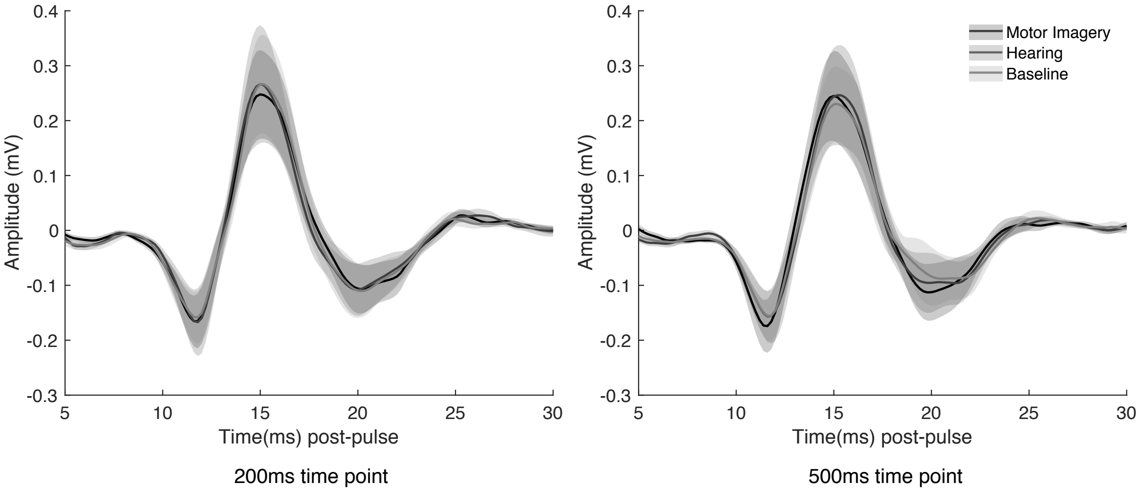


Figure 5.6: Raw MEP data waveforms for each timepoint, including motor imagery, hearing and baseline conditions. Note that waveform polymorphy and latency differences results in average peak-to-peak amplitudes and AUC appearing as smaller here than when calculated on per-participant basis (see Appendix G for individual plots). Shaded areas indicate +- 1SEM.

1.227, *p* = .268) indicating again that performance did not differ between the two groups. Since it is once again clear that there were no differences between groups, the factor Group was abolished henceforth and not included in further models. The addition of Condition to the baseline model (`MEP_AUC ~ Condition + (1|Participant)`) did not improve model fit ($\chi^2(2) = 1.061$, *p* = .588). The addition of Time point to the model baseline model (`MEP_AUC ~ Time_point + (1|Participant)`) did not improve model fit ($\chi^2(1) = 2.873$, *p* = .090). The addition of Condition and Time point to the baseline model (`MEP_AUC ~ Condition + Time_point + (1|Participant)`) did not significantly improve model fit ($\chi^2(3) = 3.925$, *p* = .269). However, comparing this model to the two-way interaction between Condition and Time point (`MEP_AUC ~ Condition * Time_point + (1|Participant)`) did significantly improve model fit ($\chi^2(2) = 7.001$, *p* = .030). Fixed effects follow-ups showed that there was a significant difference between motor imagery and baseline at 500ms (*t*(20) = 2.645, *p* =.008) but not between motor imagery and hearing (*t*(20) = 1.305, *p* = .192) or hearing and baseline (*t*(20) = 1.337, *p* = .181, see Figure 5.5).

For added clarity and completeness, addition of Group to this interaction model as either an additional interaction (`MEP_AUC ~ Group * Condition * Time_point + (1|Participant)`) or as simple main effect (`MEP_AUC ~ Group + Condition * Timepoint + (1|Participant)`) did not significantly improve model fit as judged by the comparative Akaike Information Criterion (AIC) or Bayes Information Criterion

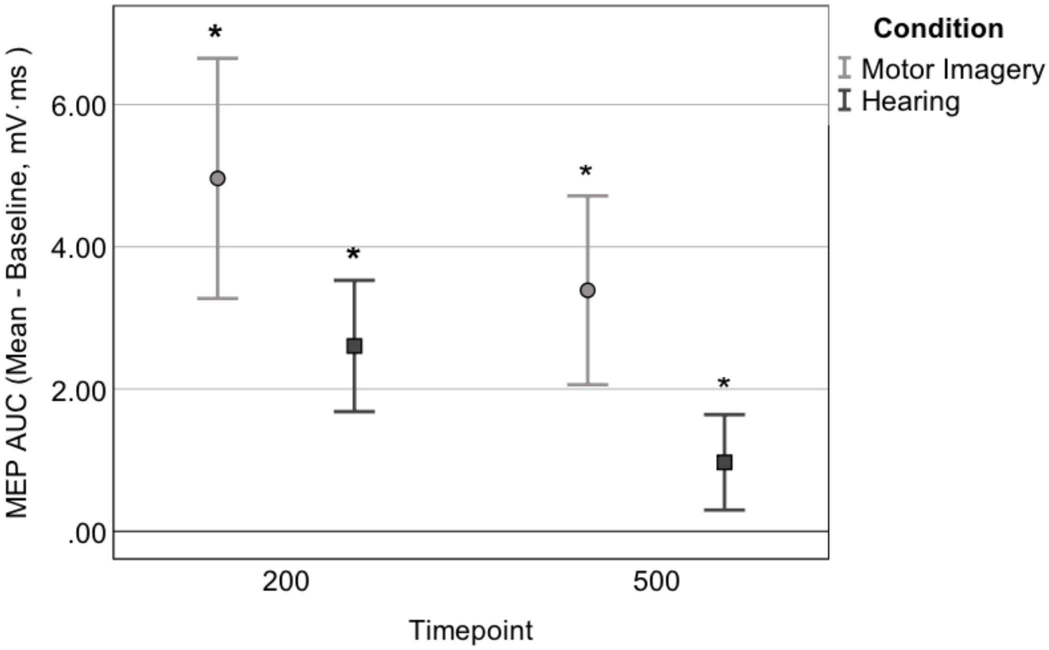


Figure 5.7: MEP AUC values relative to average baseline values for excluded participants. Scores were calculated by subtracting mean AUC of baseline per condition per time point from mean AUC of motor imagery/hearing per condition per time point. * shows significance (p = 0.05) from baseline (0.00, black horizontal line). Error bars indicate +- 1 SEM.

(BIC), which consistently identified the two-way interaction between Condition and Time point as the best fitting model. See Figure 5.6 for overall raw MEP plots, and Appendix G for individual averaged MEP plots.

The LME with the same model procedure was also carried out on the pre-MEP EMG data, which showed no effects as expected. This analysis was also carried out on the excluded group of participants and their pre-MEP EMG data, showing the expected differences between a baseline model and a model with the factor Condition ($\chi^2(2)$ = 166.34, $p < .001$) or Time point ($\chi^2(1) = 9.6815$, $p = .002$) which are known to be due to pre-existing differences in the pre-MEP EMG data as shown through the earlier ANOVAs (see previous paragraphs for `R lmer notation`, and previous Table 5.1 and Table 5.2 as well as Figure 5.7). This result in itself shows the importance of excluding those participants with differences in conditions in the pre-MEP window from the main dataset, as allowing them to remain would yield a stronger but misleading effect (see Figure 5.7).

5.4.2 Discussion

The experiments in this chapter aimed to establish cortical excitability changes in the tongue M1 during motor imagery of speech, hearing of speech, and a baseline condition. Cortical excitability, as measured through MEPs, was predicted to be greater in the motor imagery condition than in a baseline condition if M1 was involved in speech motor imagery. An additional prediction was that cortical excitability in a hearing condition would be greater than that in a baseline condition if M1 was involves in action observation processes. Neither prediction was met outright. The results show increased engagement of tongue M1 for the motor imagery condition relative to baseline, but only 500ms after the onset of motor imagery. The earlier timepoint (200ms) showed no increase in M1 activity. In addition, in contrast to previous studies reporting increased motor cortex engagement during listening to speech (Fadiga et al., 2002; Watkins et al., 2003), there is no increase in MEPs for the hearing condition relative to the baseline condition at either time point.

These results show an increase in cortical excitability at the later timepoint during imagined production of a complex consonant cluster (see Figure 5.5). Models of speech production such as Indefrey and Levelt (2000), based on picture-naming and word/pseudo word reading tasks performed with magneto-encephalography (MEG) and subdural grids, suggest that silent word reading shows activation in the occipital lobe as early as 100-200ms, with picture-naming tasks showing further activation in parietal areas around 200-400ms, or in the case of sensory cortex, as late as 600ms. Motor areas appear to be active 500-600ms after picture presentation, though silent word reading shows activation in motor areas between 200-400ms post-presentation. The key differences between their experiments and experiment 5 and 6 are the wait time (i.e., the condition was known well before a response was required) and the non-semantic nature of this task. In this task, it is likely that participants were prepared to perform the required action, as opposed to having to access lexical stores and mental syllabaries to (silently) produce a word. Salmelin, Service, Kiesilä, Uutela, and Salonen (1996) reported occipital activation at 100-200ms, with motor activity found around 200-400ms post-presentation when participants performed a silent word-reading task in a MEG study. Since in this case the condition and task were already known to the participants, one can assume a time period roughly between 100-300ms, not accounting for any additional reaction time, in which motor activation might have occurred. These results, however, do not suggest a change in cortical excitability changes in motor cortex during this time period. As stated in the introduction, the main difference between this paradigm and that in experiment 4 in Tian and Poeppel (2010) was the additional visual cue necessary to prevent participants from waiting until the TMS pulse was administered. If participants did not wait for this additional cue and used only the preceding inter-cue distance of 1s to prepare and perform imagery of the stimulus, then it is possible that even the earliest TMS pulse did not capture the imagery in this time window. The results from Tian and Poeppel (2010) suggest a mean reaction time of 1047ms from the final auditory cue (their table 2, p. 11), or around 150ms prior to the first TMS pulse, though this is based on MEG component identification. Although the length of the presented /tr/ stimulus was 1150ms and the /tr/ stimulus was presented as an auditory/production model both during learning and throughout (as part of the hearing condition), there is no way of knowing with absolute certainty whether participants performed the imagined execution for the same duration when performing motor imagery of speech. These results suggest, however, that the presence of an additional final visual cue likely slowed down the reaction time process, and given the earliest reaction time for motor performance of 200-400ms as noted in Indefrey and Levelt (2000), MEP differences should be observable during at least one of these time points if motor cortex is involved in motor imagery.

The finding that the motor imagery condition showed larger M1 excitability at the 500ms time point suggests that participants employed a strategy of waiting for the final cue and subsequently performing imagery, rather than using the alternate strategy of using the even spacing in preceding cues to anticipate and predict when to perform the imagery itself. It is also possible that inter-individual variability (see Appendix G) coupled with the length and robustness of the imagery stimulus gave rise to a result different from baseline only at the later timepoint as insufficient change occurred across participants in the earlier time point. That is, the nature of the stimulus may entail that while participants' reaction times are normal, the difficult nature of the stimulus required additional time compared to those used in other studies, such as /da/ in Tian and Poeppel (2010). The /da/ stimulus is a comparatively easy stimulus for a native British

English speaker, while /tr/ often requires practice to master. This difference may be reflected in the large difference between baseline and imagery MEPs at the later timepoint, whereas it would logically be absent at the earlier timepoint if it requires significant time to construct a motor plan.

Contrary to expectations, there was no increase in the hearing condition at either timepoint. This null effect might be due to the fact that the stimulus was a non-meaningful string of consonants. Fadiga et al., (2002) reported that listening to words elicits greater excitability than listening to pseudo-words, so it is possible that the semantic status of the stimulus in their study could have engaged M1 to a larger degree. A follow-up study by Roy et al., (2008) further showed that hearing pseudo-words involving tongue movement results in significantly larger tongue MEPs than hearing pseudo-words that require minimal tongue movement. Similarly, Watkins et al., (2003) showed increases in MEPs while listening to continuous prose, results that were confirmed by Murakami, Restle, and Ziemann (2011). However, work by Nuttall et al., (2016) showed increases in lip motor cortex for listening to meaningless distorted syllables such as /apa/ and /aba/. Work by Sato, Buccino, Gentilucci, and Cattaneo (2010) showed that auditory presentation of /ga/ (involving tongue) resulted in greater MEPs than presentation of /ba/ (involving lips), replicating results from Fadiga et al., (2002) for syllables rather than words and pseudo-words. One difference between earlier studies and experiment 5 and 6 is the time at which the TMS pulse is delivered: Watkins et al., (2003) and Murakami et al., (2011) both used continuous prose, so the consideration is irrelevant. However, Nuttall et al., (2016) delivered their pulse 100ms post-onset of the intervocalic consonant, while Sato et al., (2010) delivered their pulse at both 100ms and 200ms. Sato et al., (2010) do not find a difference in MEP amplitudes between the two timepoints.. Roy et al., (2008) investigated 4 timepoints (0, 100ms, 200ms, 300ms after intervocalic consonant onset) and found a difference between pseudo-words involving tongue movement and those not involving tongue movement at 100ms and 200ms. It is reasonable then to assume that MEPs during hearing the stimulus should be larger than baseline MEPs at the 200ms time-point, however this was not the case here (nor was it the case at 500ms).

One possibility is that the motor imagery condition difference from baseline is so strong that any effect of hearing versus baseline is lost due to low power, however this seems unlikely given the large percentage change that previous papers have reported, e.g., Watkins et al., (2003) showed a percentage increase of ~30% for speech stimuli. An alternative explanation might be that due to the novelty of the stimulus and the relative difficulty of its production, participants did not automatically engage pre-learned motor plans in the same way as in the aforementioned studies, in which the presented stimuli were relatively easy to produce. The question of stimulus novelty relates to the opposing views of Pickering and Garrod (2013), whose theory suggests that observation of learned/known actions leads to greater M1 excitability, and Wilson and Knoblich (2005), whose theory suggests that observation of novel actions lead to greater M1 excitability (Nuttall et al., 2016). The results from experiment 5 and 6 appear more in line with Pickering and Garrod's account that suggests observation of novel stimuli results in less M1 activation than observation of known sounds, since the MEP amplitude for the hearing conditions does not differ significantly from baseline. However, since there is no comparison between novel and known sounds in experiments 5 and 6 the results are not suited to make decisive statements of support for either theory.

Another possibility is that attentional demands set these experiments apart from previous studies. As two conditions were passive and one was active (in the sense that one required an active imagery response, where the others required only passive listening or nothing at all), it may be that attentional demand drives the increase in excitability. Fadiga et al., (2002) included a task where, at certain intervals, participants were asked to provide a response, and Sato et al., (2010) also included a response task. However, neither paper included a task where participants know in advance whether they need increased attention. It was assumed that they always paid equal attention to all stimuli. Nuttall et al., (2016) and Watkins et al., (2003) did not include an attention task, though Nuttall et al., (2016) noted that lip MEPs significantly increased with the addition of noise, suggesting that some level of cognitive demand may relate to increases in MEP sizes. In this sense too, these findings might showcase increased MEPs in the one condition which requires an active, and therefore more cognitively demanding task. The fact remains however, that if this were the case one would expect such a difference between conditions to be found in both early and late timepoints, given that the attention and task demand is known well before the response is required.

Finally, compared to studies looking at hand action observation and the resultant increase in M1 excitability, an important difference is the perception of movement: while hand action, including our own, is often observed visually, self-produced speech sounds can only be observed in terms of their auditory and proprioceptive consequences as the articulators are outside our field of vision. The relative lack of visual information in self-generated speech might make them harder to imagine and participants may employ a different strategy. That being said, the motor imagery condition may have led some participants to perform not articulation imagery but hearing imagery instead. Whereas articulation imagery implies egocentric behaviour, hearing imagery implies allocentric behaviour – that is, one relies on a motor strategy (as if speaking) while the other relies on a memory reactivation strategy (as if being spoken to). The instructions attempted to circumvent this issue by specifically eliciting a motor strategy ("Say the sound in your head"), but again there is no certainty as to how this instruction was interpreted by the participants. Very few studies have investigated similar issues for speech actions specifically, but the mental rotation neuroimaging literature distinguishes between more egocentric tasks, such as the Parsons (1994) hand rotation task (Ganis et al., 2000; Parsons et al., 1995) and more allocentric tasks, such as the Shepard & Metzler task (Kosslyn et al., 1998; Richter et al., 2000; Shepard & Metzler, 1971; Wraga et al., 2003) and rotation of other non-corporeal items (Vingerhoets et al., 2001, 2002) .These studies showed greater involvement of motor areas in egocentric tasks than in allocentric tasks. While speech studies using both types of task such as Tian et al., (2016) observed motor areas common to both tasks, such as mid premotor cortex and inferior *pars opercularis*, stronger motor activation (bilateral sensorimotor cortex, articulatory M1) is reserved for articulation imagery, and stronger parietal activation appeared in hearing imagery, suggesting a similar picture for motor imagery of speech. The results at the later time point suggest that, while it is difficult to ascertain every participant's strategy, a sufficient number of participants may have used the egocentric articulatory imagery strategy over the allocentric strategy.

5.4.2.1 Limitations

The pre-MEP window EMG differences between conditions raise some important questions. As the pre-MEP data analysis shows, there was a difference between the motor imagery and baseline conditions prior to the TMS pulse. This made it necessary to remove nearly half of participants from the data pool. This difference between conditions might be due to micromovements from the tongue during inner articulation. Sokolov, (1972, p. 167) described that, when faced with a mental arithmetic problem, the "[…] electrical activity of speech musculature […] while listening to a problem in many cases increased to 10-15 μV, and during subsequent mental execution […] it increased still further, not infrequently being accompanied by 'bursts' or 'volleys' of impulses which reached 50μV or more in intensity." It is this difficulty in achieving complete relaxation of the tongue that influenced the decision to use active motor threshold, but it seems that rather than resolve this potential issue, active motor threshold may have instead exacerbated it. This issue was not observed in-session due to the interleaved nature of the stimuli. Future studies may benefit from adopting a resting motor threshold procedure to avoid this problem.

One issue that needs to be addressed is the lack of difference between groups in experiments 5 and 6, even though one group underwent stimulation at 100% of aMT and the other group underwent stimulation at 120%. Expected was that there would be increased MEP amplitude in participants stimulated at 120% relative to participants stimulated at 100% as increased stimulation intensity leads to increased MEP amplitude (Rösler & Magistris, 2012). Since no such difference was observed, it is possible that MEP results show a ceiling effect. The reason for this may be threshold for MEPs, which in this case was 400μV. Although an extensive literature review was performed, future studies would benefit from choosing a lower threshold. Another possibility is that OO MEP are affected differently by increases in stimulus intensity. That is, it may be the case that MEPs collected from OO are less susceptible to increases in stimulus intensity compared to hand muscles, requiring larger increases for MEP responses to be significantly enlarged. To investigate this possibility, an investigation into input-output curves for lip muscles is suggested as a follow-up study.

Another potential issue with the design is that the upcoming condition for each trial was presented to the participant well before they had to perform what the conditions required. There was at least a 4s delay between the showing of the stimulus and the cue, which gave the participant ample time to prepare for motor imagery, hearing, or simply maintaining pressure. This delay was initially implemented to prevent potential issues from Go – No Go strategy processes. In these Go – No Go strategies, immediate decision-making was thought to potentially cause increased muscle movement, thereby skewing the results, or conversely excessive M1 inhibition due to No Go strategies (baseline) used on Go (imagery) trials. However, the pattern in this data could be due to two distinct participant strategies: (a) perceiving articulation and hearing stimuli causes conscious or unconscious increases in the tongue tension relative to the baseline condition, or (b) perceiving the baseline stimulus causes conscious or unconscious decreases in the tongue tension relative to the other stimuli. It may be difficult to tease apart which process is responsible as the pre-MEP data window is relative to the pulse, and not the cue. That is, looking at the longest pre-MEP baseline time window gives information from 300ms pre-cue to 200ms post-cue in the case of the 200ms timepoint condition, whereas it gives information from 0ms pre-cue to 500ms post-cue in the case of the 500ms timepoint conditions. Further study using this design should eliminate this wait period to ensure there are no condition-specific effects in the baseline as a result of pre-existing baseline EMG differences between conditions.

5.4.2.2 Implications

Experiments 5 and 6 provide evidence in favour of cognitive models that include M1 in a speech motor imagery processing stream. For instance, Tian et al. (2016) proposed a simulation-estimation stream that uses articulatory planning in premotor areas but bypasses M1 itself in favour of performing simulation processes in parietal areas. In their fMRI experiment comparing articulation imagery and hearing imagery, the authors noted increased activity in articulatory regions of M1 during an articulatory imagery task relative to a hearing imagery task, and highlighted that the latter task engaged parietal areas more. Lœvenbruck et al. (2018) proposed that even during speech motor imagery, M1 plays the important role of integrating auditory and somatosensory-derived motor programmes.

Working memory tasks and executive control tasks have shown that verbal motor processes are strongly linked to successful performance of action control (Baddeley et al., 2001), and studies that focus on the developmental aspects of inner speech in ASD have shown that ASD populations are far less likely to make use of inner speech in executive processing, likely preferring a visual-only strategy compared to typically-developed (TD) participants (Holland & Low, 2010; Russell-Smith et al., 2014; Whitehouse et al., 2006). Specifically, task designs that disrupt inner speech processes cause a decrease in accuracy and increase in reaction time in TD populations, but not ASD populations, suggesting more limited use of inner speech in ASD populations (Holland & Low, 2010; Russell-Smith et al., 2014). Coupled with evidence that M1 facilitation in action observation has also been observed to be lower in individuals with ASD than in TD participants (Théoret et al., 2005), these results suggest a role for articulatory M1 in inner speech and to some degree, executive processing. These results also support previous findings that articulatory M1 is active during periods of AVH (Rapin et al., 2013), suggesting its use should be integrated in forward models of inner speech (Jones & Fernyhough, 2007; Seal et al., 2004; Tian & Poeppel, 2012) as it already is integrated in forward models of overt speech (Tian et al., 2016).

As the models described above predict the involvement of specific cortical areas, future studies could use repetitive TMS to disrupt these areas and test their causal role in motor imagery. A specific area of interest is the parietal locus of the simulation-estimation stream, where motor simulation is thought to take place instead of motor execution in M1 (Tian et al., 2016). If this area is crucially involved in imagery then behavioural performance might be affected post-stimulation. Such a task would involve a reaction time and accuracy measurement task comparing execution and imagery of complex articulatory sequences after repetitive TMS to parietal areas (see Cona et al., 2017, for an example using a hand rotation task). This area could also be used in an MEP paradigm for speech actions. Tian and Poeppel (2012) predict there to be sequential nature to simulation estimation, whereas Lœvenbruck et al., (2018) predict parallel processing, a change in MEP size after rTMS to parietal areas could show which proposal is more likely to be true. Similarly, one could perform rTMS to preparatory areas, e.g., rostral prefrontal cortex, pre-SMA or SMA, to establish effects on MEPs. If these areas are involved in motor plan preparation, one would expect smaller MEPs after rTMS, however if they are involved in the inhibition of overt responses then one would predict expect larger MEPs as M1 would be disinhibited.

5.5 Conclusion

The experiments in this chapter investigated cortical excitability changes in the tongue motor cortex during motor imagery of speech, hearing of speech and a baseline condition. Both Experiment 5 and Experiment 6 suffered from the same problem, which consisted of greater force on the tongue electrodes during the motor imagery condition relative to the other conditions. Nevertheless, an analysis of the combined data found evidence that there was a difference in M1 excitability between conditions, and that this difference was dependent on the time point at which MEPs were collected: at 500ms post-cue, data showed increased cortical excitability during motor imagery of speech relative to the baseline condition. The results from this chapter support current models of speech motor imagery that suggest involvement of M1 in motor plan simulation and (limited) motor plan execution in M1 followed by inhibition, possibly from other brain areas in the parietal lobes.

The final chapter provides a discussion of all experimental results and places them within the existing literature. The final chapter also collects the insights gained from the studies that make up this thesis.

6 Chapter Six

6.1 General discussion

In this chapter, the work presented in the thesis as a whole will be discussed on the basis of the research aims set out in the introduction and the hypotheses as addressed in each chapter. The degree to which the aims were met are discussed, as are limitations of the current research and directions for future research into the research aims.

6.1.1 Research aims

The overall research aim of the thesis was *to investigate the cognitive mechanisms behind motor imagery of speech, and the neural correlates subserving these mechanisms, and specifically whether motor cortex was involved in motor imagery of speech.* Ultimately, I hoped to provide a better understanding of the cortical mechanisms involved in motor imagery which could in turn be used to influence models of motor imagery of speech. Throughout the thesis, the following aims were addressed:

Aim No. 1:	To replicate motor imagery tasks of visual actions, and to replicate these tasks for auditory actions, on the basis of behavioural paradigms (chapters 2 and 3);
Aim No. 2:	To investigate motor imagery tasks using novel paradigms, specifically suited to speech and language (chapters 2-5);
Aim No. 3:	To investigate the role of motor cortex in motor imagery of speech, through neuroimaging and neurostimulation (chapter 3 - 5);
Aim No. 4:	To investigate the relationship between inner speech experience and motor cortex activation (chapters 2-5).

6.1.2 Impact on the research aim and experimental literature

Aim No. 1: Replication of visual tasks and expansion to auditory modality

The primary aim of the studies presented in chapter 1 was to replicate motor imagery of an existing visual rotation motor imagery task, and to develop a new auditory speech motor imagery task. The behavioural results in experiments 1 – 3 were all suggestive of a correct replication of the visual motor imagery task: time course and accuracy confirmed those from previous literature (Kosslyn, Ganis, et al., 2001; Parsons, 1994; Parsons et al., 1995; Wexler et al., 1998) in which laterally-rotated stimuli showed longer response times and lower accuracy regardless of whether the stimulus was a left or right hand. The reaction time and accuracy findings were consistent across three experiments, irrespective of the number of stimuli used and regardless of whether the task took place in a behavioural lab, on-line or during fMRI scanning– all three of which are very different experimental environments. The visual task showed robust results.

The novel auditory task was specifically designed to elicit a similar implicit motor imagery strategy, and results were expected to follow similar trends: stimuli containing non-native vowels were expected to show longer reaction times and lower accuracy than stimuli containing native vowels as the motor trajectory for non-native vowels was expected to take longer to compute and simulate. This hypothesis was confirmed: non-native vowels consistently showed longer reaction times than native vowels, suggesting that the computation of the lip and tongue trajectories required more time for vowels which were less familiar. However, the results were not as clear as those in the visual task due to a confound with mouth shape: unrounded stimuli showed longer reaction times than rounded stimuli regardless of whether the stimulus was a native or non-native vowel. In addition, the auditory task also consistently showed lower overall accuracy than the visual task, suggesting that the two tasks were not as comparable in terms of difficulty as initially hoped. This was mainly due to responses to non-native stimuli consistently showing far lower accuracy relative to native stimuli. This difference between native and non-native sounds was much greater relative to the same difficulty difference in the visual task (medial vs lateral rotations), making direct comparison of the tasks difficult. However, the auditory task also largely showed the same results regardless of the conditions of the experiment (behavioural lab, on-line or during scanning), except for the effect of mouth shape which was sometimes present (experiments 1 and 2) and sometimes not (experiment 3). Results for the auditory task were generally less robust than for the visual task.

The visual task replications add to the number of replications that have already taken place in the literature (Date et al., 2015; Hyde et al., 2017; Parsons et al., 1995) and further establishes its robustness. The auditory task development should be considered work in progress: the three versions of the experiment presented here show that the task succeeds in showing differences in reaction time and accuracy between native and non-native stimuli in a task that is highly likely to involve motor imagery.

Aim No. 1: Limitations and future research directions

Limitations to each experiment related to this aim individually were previously discussed in section 3.3.2. Here I want to briefly discuss some of the general limitations related to the aim of replicating the visual task and expanding the task to the auditory modality.

One of the primary issues with the auditory task was the use of vowels classified as native and non-native for the easy-difficult distinction. The reasoning for this was related to the reasoning for the easy-difficult distinction in the visual task: according to Parsons (1994), one central reason for the easy-difficult distinction is the motor trajectory required to (mentally) achieve the orientation shown in the stimulus, or in other words, to match the mental percept to the visual percept. Parsons refers to the different joints required for such motion and the increasing complexity of the motor trajectory with an increasing number of joints. However, Parsons also consistently refers to familiarity differences in the stimuli, claiming that one possibility for the differences in reaction time and accuracy is less familiarity with the lateral visual stimuli, and consequently less familiar target postures for which it takes more time (with less accurate results) to generate appropriate motor trajectories. In designing the auditory task, the familiarity difference was exploited, but the increasing complexity due to an increasing number of joints less so. While the stimuli differed primarily in lip rounding, there was also an element of tongue position that could not be avoided. In this sense there is some similarity between the visual and auditory task, although not measured precisely. That is, lip and tongue movement are expected to move in synchrony regardless of the nativeness of the stimulus, so both articulators are always to some degree involved. This is in contrast to the visual stimuli where sometimes one joint may be involved, other times two or even three, with each additional joint increasing reaction time and reducing accuracy. It is not, however, altogether impossible to set up an auditory task that more stringently follows an increase in articulators used per stimulus. Future studies using an auditory task of this kind could attempt to use more complex articulatory stimuli, perhaps even without the native and non-native distinction in order to tease apart whether it is familiarity or increased articulator use that leads to longer reaction times and decreased accuracy. A suggestion would be to use a larger range of vowels than used here, and to include jaw movement as articulatory movement, for instance comparing back and front vowels more directly.

An additional limitation is the use of vowels as stimuli for the auditory task. Consonants were not used due to the strong categorical perception associated with consonant identification, which is less strong for vowels. The stimuli used however do remain simple CVCV sequences. It could be argued that these simple stimuli are insufficient to warrant genuine motor planning and motor simulation. Making the stimuli more complex, such as by repeating the same vowel twice, construction diphtongal stimuli or incorporating two vowels of the same type in one stimulus, could lead to greater certainty regarding use of motor simulation. More consistent motor simulation could result in more consistent results for reaction time and accuracy, as well as increased cortical activation, making the auditory and visual tasks more comparable in terms of motor planning/simulation required.

Aim No. 2: Novel speech tasks in motor imagery

This thesis presents several novel speech tasks to investigate motor imagery of speech: a decision-making task involving implicit articulatory motor imagery (experiments 1-3), an explicit, simple articulatory motor imagery task (experiment 4) and an explicit, complex articulatory motor imagery task (experiments 5 and 6). For these experiments, which focus on a hitherto little-explored process, many task design decisions were made and subsequently questioned. The implicit speech task was successful in many ways, but as previously described, produced results that were less robust than expected. This led to a focus shift towards an explicit motor imagery task. In turn, this explicit task evolved into a complex motor imagery task using a complex, non-native consonant stimulus. All three tasks represent valuable first attempts at auditory speech imagery task designs, with varying degrees of success. The implicit task was largely successful in showing behavioural data to support active use of motor imagery of speech processes, though as described above could benefit from further refinement. The simple explicit task showed that motor imagery of both hand and lip actions was not distinguishable from a baseline condition, in direct contrast to other TMS studies such as Ganis et al., (2000) who do find increased excitability in hand M1 during the hand laterality task. The results for this explicit task are therefore surprising and contribute to data suggesting that M1 is not involved in motor imagery processing when simple tasks are concerned. The complex implicit task successfully showed that motor imagery of speech resulted in increased tongue MEP amplitudes at a specific (late) timepoint. Importantly, this task also revealed that observation of the auditory stimulus (hearing) did not increase MEP amplitude as expected, in contrast with previous studies (Fadiga et al., 1995; Watkins et al., 2003).

Overall these tasks are the first attempts at exploring the research question regarding motor imagery of speech. With the results presented in the thesis they add significantly to the understanding of motor imagery, and speech motor imagery specifically, but can doubtless be improved upon.

Aim No. 2: Limitations and future research directions

There are several improvements that can be made to the tasks for use in future studies. Several improvements for the implicit task have already been suggested, so the focus here is on the explicit tasks, excepting the final point.

With reference to the explicit tasks, one issue was the lack of reaction time and accuracy data. Due to the nature of the imagery task no timed response was required, but in retrospect it would have been useful to have some idea of when imagery action finished. Active responses as in experiment 4 presented good data with regard to when active movement started, but no such data was available for motor imagery in any experiment. One solution to this problem is to have participants press a button when the requested action (or imagined action) is complete. Such data would give at least a hint of when the imagined action was completed and would be comparable to the action condition (within participant). Although there would likely be the differences in reaction time post-completion across participants, lag between finishing the (imagined) action and button press could be rectified by comparing the action condition EMG signal returning to baseline to the button press. This lag difference could then also be applied to the imagined condition for improved understanding of the time course of imagined action.

A major issue remains that even for the explicit tasks, there is no guarantee that participants actually performed motor imagery. This issue permeates the motor imagery literature but is often eschewed from explicit discussion (Parsons et al., 1995; Roosink & Zijdewind, 2010; Zapparoli et al., 2014). The general understanding is that participants will perform the requested actions and follow the correct strategy. One way to circumvent this issue is by using self-report questionnaires after the task. While the VISQ tested general inner speech experience, it did not specifically ask about the task demands and strategies.

With respect to the complex task, there are two issues that warrant further discussion. This task used only one complex, non-native stimulus articulation. While the stimulus was specifically designed to effect tongue motor imagery, an improved design might use at least one other stimulus against which the MEP results could be compared. For instance, an additional stimulus incorporating a bilabial trill could be used to compare tongue MEPs and lip MEPs, or one or more stimuli could be used and made more complex, e.g., /tratratra/ and/or /bʙabʙabʙa/.

Related to this point is the other issue with the complex task, which is the necessity of aMT use. As discussed in the methods to chapter 4 (4.2.1), aMT was used since it is difficult to keep the articulatory muscles fully at rest. However, the results from experiment 5 and 6 suggest that a significant proportion of participants consciously or subconsciously activated their tongue muscles more in specific conditions, rendering a significant subset of data unusable. This flaw is difficult to correct but using rMT may yield improved results even with the aforementioned caveats in mind. This is especially the case since using rMT would facilitate the experimenter's role of notifying the participant when they exert too much pressure (it is easier to note an increase from rest to activation than to note an increase from pre-set activation levels).

The final point relates to training for each task. The implicit tasks were only presented to the participant very briefly prior to beginning the experiment – in fact, only to the extent that a practice task could be used to ensure that participants understood the tasks. This was because of potential learning and task habituation effects, which I aimed to keep to a minimum. While learning over the task time course was explored, it is not clear whether these results reflect on actual learning or simply task habituation. Related to this is the use of complex non-native stimuli in the explicit task in experiments 5 and 6. Over the time course of the experiment, through mental practice (due to the task paradigm involving repetition of the same stimulus) it is probable that participants learned how to mentally form the stimulus better near the end of the study compared to the start of the study. Some training was provided to ensure participants could successfully recreate the stimulus, but training was not explored further. In both cases, future studies could make use of training paradigms over several hours or even several days to investigate how repeated exposure and practice affect performance, and perhaps even MEP amplitude during motor imagery.

Aim No. 3: Neuroimaging and neurostimulatory insights into M1 use in motor imagery

The majority of fMRI literature on inner speech relates to words or sentences spoken covertly, and often use explicit tasks – that is, participants are asked to perform covert speech. The fMRI task presented here focused on implicit motor imagery of speech specifically. The most important finding of the fMRI task is that relative to the well-studied visual motor imagery task, similar behavioural and neuroimaging results for the auditory task can be reported. The fMRI results showed a number of areas which were commonly activated in the visual and auditory imagery tasks, such as bilateral inferior frontal gyrus and premotor areas, pre-supplementary motor areas bilaterally as well as left superior parietal lobule. This network was largely the same yet slightly different from the expected network based on previous findings (Hardwick et al., 2018), though these areas are found across the Tian and Poeppel (2012) and Lœvenbruck et al., (2018) models as they are all involved in generating and interpreting forward model states and sensory consequences. The general network involving the aforementioned cortical areas was present in experiment 3 as in the meta-analysis, but some areas, showed different results. For instance, a small area in IFG which was previously found primarily in action observation rather than motor imagery showed the opposite response in experiment 3, and activation in dorsolateral prefrontal cortex was absent when using the appropriate mask. It is possible that both of these finding are a result of the control task used here, specifically with regard to the role of dorsolateral prefrontal cortex in response inhibition: while motor imagery of speech implies inhibition of overt speech, both tasks involved pressing a button, which may have caused imagery results in this area to be overshadowed by active manual motor processing. Note also that the coordinates for pre-supplementary motor area are relatively deep (e.g. -6, 10, 50) and it was difficult to determine whether activity loci included areas in cingulate gyrus, which was also expected (Lœvenbruck et al., 2018).

Regardless of the larger motor imagery network found, the results showed that M1 was not active in either the visual or auditory imagery task, but only in the overall task analysis where it can be attributed to button-pressing. This is in agreement with previous literature.

The neurostimulation literature, as previously discussed, has shown both evidence for and against the inclusion of M1 in the motor imagery network (Guillot et al., 2012; Hétu et al., 2013; Munzert et al., 2009). The TMS experiments reported in this thesis show evidence towards both points of view: while experiment 4 shows no increase in MEP amplitude during motor imagery, the combined analysis for experiment 5 and 6 shows a specific increase of speech motor imagery MEP amplitude over baseline MEPs at the late (500ms) timepoint.

Two models of inner speech were discussed in the introduction: the Tian and Poeppel model, which excludes M1 recruitment in inner speech, and the Lœvenbruck model, which includes M1 recruitment in inner speech. Both models propose a forward model which includes the elements of efference copy, motor-to-sensory transformation and corollary discharge which lead to an inner speech experience. With regard to M1 use in motor imagery of speech, the fMRI and TMS results show tentative support for the Tian and Poeppel model, although the final TMS results do suggests some role for M1 at a specific timepoint. The results suggest that the complexity of simulation directly influences whether M1 is recruited or not: complex stimuli show M1 recruitment (experiment 5 and 6), but simple stimuli do not (experiment 4). More research is required to definitively state the role of M1 in speech motor imagery, and some suggestions are provided.

Aim No. 3: Limitations and future research directions

The fMRI design used in chapter 3 suffers from two design issues which should be addressed. The number of stimuli, while acceptable for the hand task (previously used, sufficiently robust to show effect in fMRI/PET) may have been too low for effective use with the auditory task. One reason for this assumption is a recent study by W. Zhang et al., (2020) which used similarly short stimuli (CV rather than CVCV) but used a larger number with several repetitions. The authors find activation in IFG as well as parietal areas, and even in MTG. However, activation in MTG (and IFG to some degree) may be explained by activation through the ventral stream (Tian et al., 2016) since the Mandarin Chinese CV stimuli were not semantically vacuous (as discussed with a native Mandarin speaker). Future studies would benefit from an increased number of stimuli and trials. Additionally, the reaction time task, while useful, did not aid the understanding of inner speech processes beyond allowing for a contrast in which a button press also occurred. A better additional task might involve overt speech or silently articulated speech.

Another possible improvement that was not included in this thesis was a predefined region-of-interest (ROI) analysis. Since the hand task revealed several areas related to motor imagery and other studies have shown effects of covert speech, it would have been possible to assign non-stringent ROIs to specific cortical areas, and certainly given the results presented here it would be possible to assign more stringent ROIs for both the visual and auditory task in follow-up studies, such as areas around pre-SMA, cingulate gyrus, inferior frontal gyrus and superior parietal lobule. It would even be possible for another researcher initially blind to the results from experiment 3 to assign ROIs and then assess the experiment 3 data in light of the ROIs concerned. However, the COBIDAS report (Nichols et al., 2016) highlights the difficulties with the same researchers assigning ROIs post-hoc and re-using these regions on the same dataset from which they were originally taken. Since I did not plan predefined ROI analyses and doing so post-hoc would likely lead to circular analyses, such analyses were not included in the thesis. Future studies should assign ROI analyses (based on the research presented in this thesis, amongst other sources) prior to starting the experiment.

Some issues with the TMS task design, such as questions regarding the necessity of muscle contraction, the complexity of the task and pre-MEP baseline statistics have already been discussed (see 4.3.2). However, there are some TMS-specific issues that require discussion. As alluded to in chapter 4, the pause between informing the participant of the upcoming action required and cuing the participant to perform the action required led to unexpected issues with increased baseline EMG in the pre-pulse window. This complication may be avoided by reducing the length of time taken up by the pause, at the potential expense of participants switching to a Go – No Go strategy. Such a strategy involves preparatory motor planning and motor activation, which may result in data that is difficult to analyse should the experimenter have to decide on a trial-by-trial basis whether a certain strategy was used or not.

One way in which the recruitment of M1 during motor imagery of speech may be probed in by using modulatory TMS either in conjunction with, or instead of a TMS-MEP design. Short interval (intra)cortical inhibition (SICI) (Hanajima & Ugawa, 2012) involves cortical stimulation in two separate areas, often a non-motor region and M1, to see the effects of a pulse in a cortical region on MEPs, which may show increased or decreased amplitude as a result. In this case, a SICI paradigm, e.g., a combination of left- and right-hemisphere TMS, could be used to measure change in MEPs during motor imagery of speech. Alternatively, a repetitive TMS paradigm could be used to modulate the activation of M1 lip region. Using a suprathreshold stimulation paradigm, it would be possible to inhibit M1 activation and look at reaction time and/or accuracy results to determine if there is a change after M1 inhibition. If there is an increase in reaction time, this would point to a role for M1 in speech motor imagery since it is the inhibition of M1 that, *ceteris paribus*, results in slower reaction times.

Another use of modulatory TMS would be to use a repetitive paradigm with a focus on other nodes of the imagery network, such as those found in the fMRI task (e.g., pre-SMA/SMA, IFG) and larger meta-analyses (e.g., DLPFC). A repetitive TMS design could make use of specific paradigms to test whether stimulation of these areas has a consequent excitatory or inhibitory effect on MEPs or task performance, as measured through reaction time and accuracy (Fitzgerald et al., 2006). Specifically, this type of design could potentially disentangle whether M1 is actively inhibited during speech motor imagery, as proposed by some researchers (Grandchamp et al., 2019; Lœvenbruck et al., 2018). Interestingly, rTMS applied to such areas in order to inhibit their function might lead to increased MEPs, and so show that inhibition of active motor processing does take place (or not). Such a study was designed, and while a number of participants (n = 10) had taken part the project was cut short due to lab closures related to coronavirus restrictions, and not reported in this thesis.

Aim No. 4: Relationship between inner speech experience and M1 activation

Throughout all experiments, one constant was the post-hoc administration of the Varieties of Inner Speech Questionnaire (VISQ). This questionnaire was used to better understand each participant's individual experience of inner speech, and to see if there was a correlation between their inner speech experience and performance in motor imagery of speech tasks. Overall, the results from the questionnaire in each experiment were useful in giving a general indication of the inner speech experience of each sample, but correlations were limited to medium positive correlations between factor 3 (other people in inner speech) and accuracy in experiment 2, a medium negative correlation between factor 3 and reaction time in experiment 3, and small to medium correlations between factor 4 (condensed inner speech) and MEP AUC in experiment 4. These correlations were not particularly indicative of genuine relationships, since they either contradicted each other in direction (factor 3) or were limited to the middle spectrum of the VISQ scale (factor 4). The VISQ therefore does not add much of significance to the results and a different questionnaire, such as one that asks more direct questions related to the task, may be more appropriate for this type of research.

Aim No. 4: Limitations and future research directions

One potentially interesting avenue would be to use descriptive experience sampling (Hurlburt & Akhter, 2006) to better understand participants' experience of inner speech. This method involves asking participants to carry around a beeper device which occasionally beeps, signifying that the participant should write down what (and how) they were thinking about. This has previously provided results with regard to how condensed or expanded participants' inner speech is, which may in turn provide a more naturalistic dataset with which to compared fMRI and MEP results.

Note on relation of research to atypical populations

Finally, this research focused entirely on normal adult populations. However, there is substantial evidence that ASD populations have a very different experience of inner speech, and therefore likely a different experience of speech motor imagery. Patients suffering from frequent auditory hallucinations also form a group with whom this research may yield different and revealing results.

6.2 Overall Conclusion

The research presented in this thesis explored the relation between motor imagery of speech, also referred to as motor simulation of speech, and the recruitment of primary motor cortex (M1). Some models suggest that M1 is actively recruited during inner speech processes (Lœvenbruck et al., 2018) while others suggest that M1 is not recruited during inner speech processes (Tian & Poeppel, 2012). Behavioural results showed that auditory motor imagery tasks designed in similar ways to visual motor imagery tasks show similar trends suggestive of active simulation during motor imagery of speech, although neuroimaging did not find activation of M1 during visual or auditory motor imagery. Neurostimulation methods using TMS - MEP designs show that articulatory motor imagery can result in M1 recruitment, but that this may depend on several design parameters, such as choice of articulator, stimuli and stimulation paradigm. Overall, the research suggests that M1 may only be recruited when involved in simulation of complex speech actions, and not during simple speech act simulation.

6.3 Summary of significant findings of the thesis

The following findings summarise the significant additions to the literature on motor imagery of speech. The replication of visual motor imagery results, behaviourally and using neuroimaging, provides further support for the theory that motor imagery of speech at some level makes active use of motor planning networks. The behavioural results suggest that motor imagery processes follow simulation parameters such as movement limits and effector positioning. The neuroimaging results show evidence of a shared network of cortical activation involved in simulation of visual and auditory imagery, while neurostimulatory findings suggest that active speech performance recruits motor cortex, but simple perception of auditory stimuli does not. However, the neurostimulatory findings also suggest that auditory imagery of simple speech actions does not recruit motor cortex, but auditory imagery of complex actions does recruit motor cortex.

Funding acknowledgements

This work was generously supported by an ESRC studentship to the candidate (award reference 1891294). In addition, the work was supported in part by the BIAL foundation (No. 281/16), and the candidate thanks the Birkbeck-UCL Centre for Neuroimaging (BUCNI) for generously providing scanning privileges and scanning time.

Bibliography

Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of multimovement coordination: Sensorimotor mechansims in Speech motor programming. *Journal of Motor Behavior*, *16*(2), 195–232.

Adams, R. A., Shipp, S., & Friston, K. J. (2012). Predictions not commands: active inference in the motor system. *Brain Structure & Function*, *218*, 611–643.

Adank, P., Kennedy-Higgins, D., Maegherman, G., Hannah, R., & Nuttall, H. E. (2018). Effects of Coil Orientation on Motor Evoked Potentials From Orbicularis Oris. *Frontiers in Neuroscience*, *12*(November), 1–8. https://doi.org/10.3389/fnins.2018.00683

Adank, P., Nuttall, H. E., & Kennedy-Higgins, D. (2017). Transcranial magnetic stimulation and motor evoked potentials in speech perception research. *Language, Cognition and Neuroscience*, *32*(7), 900–909. https://doi.org/10.1080/23273798.2016.1257816

Agnew, Z. K., McGettigan, C., Banks, B., & Scott, S. K. (2013). Articulatory movements modulate auditory responses to speech. *NeuroImage*, *73*, 191–199. https://doi.org/10.1016/j.neuroimage.2012.08.020

Alderson-Day, B., & Fernyhough, C. (2015). *Inner Speech : Development , Cognitive Functions , Phenomenology , and Neurobiology*. *141*(5), 931–965.

Alkadhi, H., Brugger, P., Boendermaker, S. H., Crelier, G., Curt, A., Hepp-Reymond, M. C., & Kollias, S. S. (2005). What disconnection tells about motor imagery: Evidence from paraplegic patients. *Cerebral Cortex*, *15*(2), 131–140. https://doi.org/10.1093/cercor/bhh116

Altmann, C. F., Uesaki, M., Ono, K., Matsuhashi, M., Mima, T., & Fukuyama, H. (2014). Categorical speech perception during active discrimination of consonants and vowels. *Neuropsychologia*, *64*, 13–23. https://doi.org/10.1016/j.neuropsychologia.2014.09.006

Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature*, *568*(7753), 493–498. https://doi.org/10.1038/s41586-019-1119-1

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2018). Gorillas in our Midst: Gorilla.sc, a new web-based Experiment Builder. *BioRxiv*.

https://doi.org/10.1101/438242

Aristotle, (translated by E.M Edgehill). (1928). *On Interpretation*. Original work published 350. B.C.E. http://classics.mit.edu/Aristotle/interpretation.2.2.html

Arthurs, O. J., & Boniface, S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? "… …up to 95% of regional cerebellar blood flow increases might be dependent on postsynaptic activity." *TRENDS in Neurosciences*, *25*(1), 27–31. http://tins.trends.com

Ashburner, J., Barnes, G., Chen, C.-C., Daunizeau, J., Flandin, G., Friston, K. J., Kiebel, S., Kilner, J., Litvak, V., Moran, R., Penny, W., Razi, A., Stephan, K., Tak, S., Zeidman, P., Gitelman, D., Henson, R., Hutton, C., Glauche, V., … Phillips, C. (2016). *SPM12 Manual*.

Asimov, I. (1953). *Second Foundation*. Bantam Spectra.

Baad-Hansen, L., Blicher, J. U., Lapitskaya, N., Nielsen, J. F., & Svensson, P. (2009). Intra-cortical excitability in healthy human subjects after tongue training. *Journal of Oral Rehabilitation*, *36*(6), 427–434. https://doi.org/10.1111/j.1365-2842.2009.01955.x

Baars, B. J. (2003). How Brain Reveals Mind. *Journal of Consciousness Studies*, *9–10*(9), 100–114.

Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working Memory and the Control of Action: Evidence From Task Switching. *Journal of Experimental Psychology*, *130*(4), 641–657.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation*, 260–268.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*, 37–46.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barwood, C. H. S., Murdoch, B. E., Goozee, J. V., & Riek, S. (2013). Investigating the neural basis of stuttering using transcranial magnetic stimulation: Preliminary case discussions. *Speech, Language and Hearing*, *16*(1), 18–27.

https://doi.org/10.1179/2050571X12Z.0000000001

Basho, S., Palmer, E. D., Rubio, M. A., Wulfeck, B., & Müller, R. A. (2007). Effects of generation mode in fMRI adaptations of semantic fluency: Paced production and overt speech. *Neuropsychologia*, *45*(8), 1697–1706. https://doi.org/10.1016/j.neuropsychologia.2007.01.007

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2018). Voice-Selective Areas in Human Auditory Cortex. *Foundations in Social Neuroscience*, 309–312. https://doi.org/10.7551/mitpress/3077.003.0023

Bergen, B. K. (2012). *Louder than words: The new science of how the mind makes meaning*.

Blakemore, S., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self produced tickle sensation. *Nature Neuroscience*, *1*(7), 635–640.

Blasi, G., Goldberg, T. E., Weickert, T., Das, S., Kohn, P., Zoltick, B., Bertolino, A., Callicott, J. H., Weinberger, D. R., & Mattay, V. S. (2006). Brain regions underlying response inhibition and interference monitoring and suppression. *European Journal of Neuroscience*, *23*(6), 1658–1664. https://doi.org/10.1111/j.1460-9568.2006.04680.x

blender.org. (2015). *blender.org - Home of the Blender project - Free and Open 3D Creation Software*. Blender.Org.

Bocci, T., Briscese, L., Giorli, E., Pecori, C., & Sartucci, F. (2012). Tongue’s motor evoked potentials in the diagnosis of Primary Lateral Sclerosis (PLS): Preliminary report. *Journal of the Neurological Sciences*, *316*(1–2), 67–71. https://doi.org/10.1016/j.jns.2012.01.028

Boersma, W., & Weenink, D. (2006). Praat software. In *Amsterdam: University*.

Boisgontier, M. P., & Cheval, B. (2016). The anova to mixed model transition. *Neuroscience and Biobehavioral Reviews*, *68*, 1004–1005. https://doi.org/10.1016/j.neubiorev.2016.05.034

Boudreau, S. A., Lontis, E. R., Caltenco, H., Svensson, P., Sessle, B. J., Struijk, L. N. A., & Arendt-Nielsen, L. (2013). Features of cortical neuroplasticity associated with multidirectional novel motor skill training: A TMS mapping study. *Experimental Brain Research*, *225*(4), 513–526. https://doi.org/10.1007/s00221-012-3391-2

Boudreau, S., Romaniello, A., Wang, K., Svensson, P., Sessle, B. J., & Arendt-Nielsen, L.

(2007). The effects of intra-oral pain on motor cortex neuroplasticity associated with short-term novel tongue-protrusion training in humans. *Pain*, *132*(1–2), 169–178. https://doi.org/10.1016/j.pain.2007.07.019

British Society of Audiology. (2012). Recommended procedure. Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. *British Society of Audiology*, *1*(February), 1–32. https://doi.org/10.1017/CBO9781107415324.004

Busan, P., Del Ben, G., Bernardini, S., Natarelli, G., Bencich, M., Monti, F., Manganotti, P., & Battaglini, P. P. (2016). Altered modulation of silent period in tongue motor cortex of persistent developmental stuttering in relation to stuttering severity. *PLoS ONE*, *11*(10), 1–22. https://doi.org/10.1371/journal.pone.0163959

Cattaneo, L., & Pavesi, G. (2014). The facial motor system. *Neuroscience and Biobehavioral Reviews*, *38*, 135. https://doi.org/10.1016/j.neubiorev.2013.11.002

Chang-Yit, R., Pick, H. L., & Siegel, G. M. (1975). Reliability of sidetone amplification effect in vocal intensity. *Journal of Communication Disorders*, *8*(4), 317–324.

Cheng, I. K. Y., Yiu, E. M. L., & Chan, K. M. K. (2017). Changes in resting motor threshold of the tongue with normal aging and stroke. *Somatosensory and Motor Research*, *34*(4), 242–247. https://doi.org/10.1080/08990220.2017.1421159

Chipchase, L., Schabrun, S., Cohen, L., Hodges, P., Ridding, M., Rothwell, J., Taylor, J., & Ziemann, U. (2012). A checklist for assessing the methodological quality of studies using transcranial magnetic stimulation to study the motor system: An international consensus study. *Clinical Neurophysiology*, *123*(9), 1698–1704. https://doi.org/10.1016/j.clinph.2012.05.003

Cona, G., Marino, G., & Semenza, C. (2017). TMS of supplementary motor area (SMA) facilitates mental rotation performance: Evidence for sequence processing in SMA. *NeuroImage*, *146*(May 2016), 770–777. https://doi.org/10.1016/j.neuroimage.2016.10.032

Conant, R. C., & Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, *1*(2), 89–97.

Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error Biases in Inner and Overt Speech: Evidence From Tongue Twisters. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(1), 162–175. https://doi.org/10.1037/a0021321

Corretge, R. (2012). *Praat Vocal Toolkit*. Http://Www.Praatvocaltoolkit.Com.

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

Darling, W. G., Wolf, S. L., & Butler, A. J. (2006). Variability of motor potentials evoked by transcranial magnetic stimulation depends on muscle activation. *Experimental Brain Research*, *174*(2), 376–385. https://doi.org/10.1007/s00221-006-0468-9

Date, S., Kurumadani, H., Watanabe, T., & Sunagawa, T. (2015). Transcranial direct current stimulation can enhance ability in motor imagery tasks. *NeuroReport*, *26*(11), 613–617. https://doi.org/10.1097/WNR.0000000000000398

de Vries, S., & Mulder, T. (2007). Motor imagery and stroke rehabilitation: A critical discussion. *Journal of Rehabilitation Medicine*, *39*(1), 5–13. https://doi.org/10.2340/16501977-0020

Decety, J. (1996). Do imagined and executed actions share the same neural substrate? *Cognitive Brain Research*, *3*(2), 87–93. https://doi.org/10.1016/0926-6410(95)00033-X

Decety, J., Jeannerod, M., & Prablanc, C. (1989). The timing of mentally represented actions. *Behavioural Brain Research*, *34*(1), 35–42.

Deiber, M. P., Ibañez, V., Honda, M., Sadato, N., Raman, R., & Hallett, M. (1998). Cerebral processes related to visuomotor imagery and generation of simple finger movements studied with positron emission tomography. *NeuroImage*, *7*(2), 73–85. https://doi.org/10.1006/nimg.1997.0314

Dell, G. S., & Oppenheim, G. M. (2015). Insights for Speech Production Planning from Errors in Inner Speech. *The Handbook of Speech Production*, 404–418. https://doi.org/10.1002/9781118584156.ch18

Devlin, J. T., & Watkins, K. E. (2007). Stimulating language: Insights from TMS. *Brain*, *130*(3), 610–622. https://doi.org/10.1093/brain/awl331

Eisenegger, C., Herwig, U., & Jäncke, L. (2007). The involvement of primary motor cortex in mental rotation revealed by transcranial magnetic stimulation. *European Journal of Neuroscience*, *25*(4), 1240–1244. https://doi.org/10.1111/j.1460-9568.2007.05354.x

Ernberg, M., Serra, E., Baad-Hansen, L., & Svensson, P. (2009). Influence of topical anaesthesia on the corticomotor response to tongue training. *Archives of Oral Biology*, *54*(7), 696–704. https://doi.org/10.1016/j.archoralbio.2009.04.004

Fadiga, L., Buccino, G., Craighero, L., Fogassi, L., Gallese, V., & Pavesi, G. (1998).

Corticospinal excitability is specifically modulated by motor imagery: A magnetic stimulation study. *Neuropsychologia*, *37*(2), 147–158. https://doi.org/10.1016/S0028-3932(98)00089-X

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). *Speech listening specifically modulates the excitability of tongue muscles : a TMS study*. *15*(September 2001), 399–402.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611. https://doi.org/10.1152/jn.1995.73.6.2608

Farah, M. J. (1989). The neural basis of mental imagery. *Trends in Neurosciences*, *12*(10), 395–399. https://doi.org/10.1016/0166-2236(89)90079-9

Fernyhough, C. (2004). Alien voices and inner dialogue: Towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology*, *22*(1), 49–68. https://doi.org/10.1016/j.newideapsych.2004.09.001

Fernyhough, C. (2017). The Voices Within: The History and Science of How We Talk to Ourselves. In *Basic Books*.

Fitzgerald, P. B., Fountain, S., & Daskalakis, Z. J. (2006). A comprehensive review of the effects of rTMS on motor cortical excitability and inhibition. *Clinical Neurophysiology*, *117*(12), 2584–2596. https://doi.org/10.1016/j.clinph.2006.06.712

Friston, K. J. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325–1352. https://doi.org/10.1016/j.neunet.2003.06.005

Friston, K. J., Mechelli, A., Turner, R., & Price, C. J. (2000). Nonlinear responses in fMRI: The balloon model, Volterra kernels, and other hemodynamics. *NeuroImage*, *12*(4), 466–477. https://doi.org/10.1006/nimg.2000.0630

Frith, C. D. (1992). The Cognitive Neuropsychology Of Schizophrenia. In *Neuropsychologia*. https://doi.org/10.1016/0028-3932(93)90120-O

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593–609. https://doi.org/10.1093/brain/119.2.593

Ganis, G., Keenan, J. P., Kosslyn, S. M., & Pascual-Leone, A. (2000). Transcranial magnetic stimulation of primary motor cortex affects mental rotation. *Cerebral Cortex*, *10*(2), 175–180. https://doi.org/10.1093/cercor/10.2.175

Gauvin, H. S., & Hartsuiker, R. J. (2020). Towards a new model of verbal monitoring. *Journal of Cognition*, *3*(1), 1–37.

Gentilucci, M., Campione, G. C., Dalla Volta, R., & Bernardis, P. (2009). The observation of manual grasp actions affects the control of speech: A combined behavioral and Transcranial Magnetic Stimulation study. *Neuropsychologia*, *47*(14), 3190–3202. https://doi.org/10.1016/j.neuropsychologia.2009.07.020

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J. C., & Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain*, *134*(10), 3071–3082. https://doi.org/10.1093/brain/awr232

Ghezzi, A., & Baldini, S. (1998). A simple method for recording motor evoked potentials of lingual muscles to transcranial magnetic and peripheral electrical stimulation. *Electroencephalography and Clinical Neurophysiology - Electromyography and Motor Control*, *109*(2), 114–118. https://doi.org/10.1016/S0924-980X(98)00008-3

Grandchamp, R., Rapin, L., Perrone-Bertolotti, M., Pichat, C., Haldin, C., Cousin, E., Lachaux, J. P., Dohen, M., Perrier, P., Garnier, M., Baciu, M., & Lœvenbruck, H. (2019). The ConDialInt Model: Condensation, Dialogality, and Intentionality Dimensions of Inner Speech Within a Hierarchical Predictive Control Framework. *Frontiers in Psychology*, *10*(September). https://doi.org/10.3389/fpsyg.2019.02019

Groppa, S., Oliviero, A., Eisen, A., Quartarone, A., Cohen, L. G., Mall, V., Kaelin-lang, A., Mima, T., Rossi, S., Thickbroom, G. W., Rossini, P. M., Ziemann, U., Valls-solé, J., & Siebner, H. R. (2012). Clinical Neurophysiology A practical guide to diagnostic transcranial magnetic stimulation : Report of an IFCN committee. *Clinical Neurophysiology*, *123*(5), 858–882. https://doi.org/10.1016/j.clinph.2012.01.010

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. https://doi.org/10.1016/j.bandl.2005.06.001

Guillot, A., Di Rienzo, F., MacIntyre, T., Moran, A., & Collet, C. (2012). Imagining is Not Doing but Involves Specific Motor Commands: A Review of Experimental Data Related to Motor Inhibition. *Frontiers in Human Neuroscience*, *6*(September), 1–22. https://doi.org/10.3389/fnhum.2012.00247

Haggard, P., Newman, C., & Magno, E. (1999). On the perceived time of voluntary actions. *British Journal of Psychology*, *90*(2), 291–303. https://doi.org/10.1348/000712699161413

Halkjaer, L., Melsen, B., McMillan, A. S., & Svensson, P. (2006). Influence of sensory deprivation and perturbation of trigeminal afferent fibers on corticomotor control of human tongue musculature. *Experimental Brain Research*, *170*(2), 199–205. https://doi.org/10.1007/s00221-005-0199-3

Hanajima, R., & Ugawa, Y. (2012). Paired-pulse measures. *Oxford Handbook of Transcranial Stimulation*, *July 2016*, 1–27. https://doi.org/10.1093/oxfordhb/9780198568926.013.0011

Hardwick, R. M., Caspers, S., Eickhoff, S. B., & Swinnen, S. P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. *Neuroscience and Biobehavioral Reviews*, *94*(August), 31–44. https://doi.org/10.1016/j.neubiorev.2018.08.003

Hashimoto, R., & Rothwell, J. C. (1999). Dynamic changes in corticospinal excitability during motor imagery. *Experimental Brain Research*, *125*(1), 75–81. https://doi.org/10.1007/s002210050660

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetencephalographic evidence for a precise forward model in speech production. *Neuroreport*, *17*(13), 1375–1379. https://doi.org/10.1097/01.wnr.0000233102.43526.e9.Magnetoencephalographic

Hétu, S., Grégoire, M., Saimpont, A., Coll, M. P., Eugène, F., Michon, P. E., & Jackson, P. L. (2013). The neural network of motor imagery: An ALE meta-analysis. *Neuroscience and Biobehavioral Reviews*, *37*(5), 930–949. https://doi.org/10.1016/j.neubiorev.2013.03.017

Heyes, C. (2011). Automatic Imitation. *Psychological Bulletin*, *137*(3), 463–483. https://doi.org/10.1037/a0022288

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience.*, *13*, 135-145. https://doi.org/10.1038/nrn3158

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review*, *24*(2), 189–233. https://doi.org/10.1016/j.dr.2004.01.001

Holland, L., & Low, J. (2010). Do children with autism use inner speech and visuospatial resources for the service of executive control? Evidence from suppression in dual tasks. *British Journal of Developmental Psychology*, *28*(2), 369–391. https://doi.org/10.1348/026151009X424088

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, *5*(September), 82.

https://doi.org/10.3389/fnhum.2011.00082

Huang, J., Carr, T., & Cao, Y. (2001). Comparing Cortical Activations for Silent and Overt Speech Using Event-Related fMRI. *Human Brain Mapping*, *15*(26), 26–38. https://doi.org/10.1002/hbm.XXXX

Huettel, S. A., Song, A. W., & McCarthy, G. (2014). Functional Magnetic Resonance Imaging, Third Edition. In *Functional Magnetic Resonance Imaging, Third Edition*. https://doi.org/10.1088/1475-7516/2003/08/005

Hurlburt, R. T., Heavey, C. L., & Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Consciousness and Cognition*, *22*(4), 1477–1494. https://doi.org/10.1016/j.concog.2013.10.003

Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., & Turner, R. (2002). Image distortion correction in fMRI: A quantitative evaluation. *NeuroImage*, *16*(1), 217–240. https://doi.org/10.1006/nimg.2001.1054

Hyde, C., Fuelscher, I., Lum, J. A. G., Williams, J., He, J., & Enticott, P. G. (2017). Primary Motor Cortex Excitability Is Modulated During the Mental Simulation of Hand Movement. *Journal of the International Neuropsychological Society*, *23*(02), 185–193. https://doi.org/10.1017/S1355617717000029

Hyde, C., Wilmut, K., Fuelscher, I., & Williams, J. (2013). Does implicit motor imagery ability predict reaching correction efficiency? A test of recent models of human motor control. *Journal of Motor Behavior*, *45*(3), 259–269. https://doi.org/10.1080/00222895.2013.785927

Iacoboni, M. (2005). Neural mechanisms of imitation. *Current Opinion in Neurobiology*, *15*(6), 632–637. https://doi.org/10.1016/j.conb.2005.10.010

Indefrey, P., & Levelt, W. J. M. (2000). The Neural Correlates of Language Production. *The New Cognitive Neurosciences*, 845–865.

Izumi, S., Findley, T., Ikai, T., Andrews, J., Daum, M., & Chino, N. (1995). Izumi et al 1995.pdf. *American Journal of Physical Medicine & Rehabilitation*, *75*(3), 207–213.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention. *Behavioral and Brain Sciences*, *17*(1994), 187–245.

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *NeuroImage*, *14*(1), S103--S109. https://doi.org/10.1006/nimg.2001.0832

Jones, S. R. (2010). Do we need multiple models of auditory verbal hallucinations? examining the phenomenological fit of cognitive and neurological models.

Schizophrenia Bulletin, *36*(3), 566–575. https://doi.org/10.1093/schbul/sbn129

Jones, S. R., & Fernyhough, C. (2007). Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition*, *16*(2), 391–399. https://doi.org/10.1016/j.concog.2005.12.003

Kasess, C. H., Windischberger, C., Cunnington, R., Lanzenberger, R., Pezawas, L., & Moser, E. (2008). The suppressive influence of SMA on M1 in motor imagery revealed by fMRI and dynamic causal modeling. *NeuroImage*, *40*(2), 828–837. https://doi.org/10.1016/j.neuroimage.2007.11.040

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *9*(6), 718–727.

Kell, C. A., Darquea, M., Behrens, M., Cordani, L., Keller, C., & Fuchs, S. (2017). Phonetic detail and lateralization of reading-related inner speech and of auditory and somatosensory feedback processing during overt reading. *Human Brain Mapping*, *38*(1), 493–508. https://doi.org/10.1002/hbm.23398

Kelso, J. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(6), 812.

Khedr, E. M., Abdel-Fadeil, M. R., El-Khilli, F., & Ibrahim, M. Q. (2005). Impaired corticolingual pathways in patients with or without dysarthria after acute monohemispheric stroke. *Neurophysiologie Clinique*, *35*(2–3), 73–80. https://doi.org/10.1016/j.neucli.2005.03.003

Komeilipoor, N., Vicario, C. M., Daffertshofer, A., & Cesari, P. (2014). Talking hands: Tongue motor excitability during observation of hand gestures associated with words. *Frontiers in Human Neuroscience*, *8*(SEP), 1–7. https://doi.org/10.3389/fnhum.2014.00767

Komoda, Y., Iida, T., Kothari, M., Komiyama, O., Baad-Hansen, L., Kawara, M., Sessle, B., & Svensson, P. (2015). Repeated tongue lift movement induces neuroplasticity in corticomotor control of tongue and jaw muscles in humans. *Brain Research*, *1627*, 70–79. https://doi.org/10.1016/j.brainres.2015.09.016

Kosslyn, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology*, *22*(3–4), 333–347. https://doi.org/10.1080/02643290442000130

Kosslyn, S. M., Digirolamo, G. J., Thompson, W. L., & Alpert, N. M. (1998). Mental rotation of objects versus hands: Neural mechanisms revealed by positron emission tomography. *Psychophysiology*, *35*(2), S0048577298001516. https://doi.org/10.1017/S0048577298001516

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635–642.

Kosslyn, S. M., Thompson, W. L., Wraga, M., & Alpert, N. M. (2001). Comparing Cortical Activations for Silent and Overt Speech Using Event-Related fMRI: Distinct neural mechanisms. *NeuroReport*, *12*(11), 2519–2525. https://doi.org/10.1097/00001756-200108080-00046

Kothari, M., Baad-Hansen, L., & Svensson, P. (2016). Bilateral sensory deprivation of trigeminal afferent fibres on corticomotor control of human tongue musculature: a preliminary study. *Journal of Oral Rehabilitation*, *43*(9), 656–661. https://doi.org/10.1111/joor.12414

Kothari, M., Stubbs, P. W., Figlewski, K., Pedersen, A. R., Jensen, J., Baad-Hansen, L., Svensson, P., & Nielsen, J. F. (2017). Effect of transcranial direct current stimulation on neuroplasticity in corticomotor pathways of the tongue muscles. *Journal of Oral Rehabilitation*, *44*(9), 691–701. https://doi.org/10.1111/joor.12529

Kothari, M., Svensson, P., Jensen, J., Kjærsgaard, A., Jeonghee, K., Nielsen, J. F., Ghovanloo, M., & Baad-Hansen, L. (2013). Training-induced cortical plasticity compared between three tongue-training paradigms. *Neuroscience*, *246*, 1–12. https://doi.org/10.1016/j.neuroscience.2013.04.040

Kothari, Mohit, Svensson, P., Nielsen, J. F., & Baad-Hansen, L. (2014). Influence of position and stimulation parameters on intracortical inhibition and facilitation in human tongue motor cortex. *Brain Research*, *1557*, 83–89. https://doi.org/10.1016/j.brainres.2014.02.017

Kuhtz-Buschbeck, J. P., Mahnkopf, C., Holzknecht, C., Siebner, H., Ulmer, S., & Jansen, O. (2003). Effector-independent representations of simple and complex imagined finger movements: a combined fMRI and TMS study. *Federation of European Neuroscience Societies*, *18*, 3375–3387. https://doi.org/10.1046/j.1460-9568.2003.03066.x

Kumru, H., Soto, O., Casanova, J., & Valls-Sole, J. (2008). Motor cortex excitability changes during imagery of simple reaction time. *Experimental Brain Research*, *189*(3), 373–378. https://doi.org/10.1007/s00221-008-1433-6

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*(NOV), 1–12. https://doi.org/10.3389/fpsyg.2013.00863

Lancaster, J. L., & Martinez, M. J. (2007). Mango–Multi-Image Analysis GUI. In *Research Imaging*. University of Texas Health Science Center. http://ric.uthscsa.edu/mango/index.html

Langland-Hassan, P., & Vicente, A. (2018). *Inner speech: New voices* (P. Langland-Hassan & A. Vicente (eds.)). Oxford University Press.

Lee, B. S. (1950). Effects of delayed speech feedback. *The Journal of the Acoustical Society of America*, *22*(6), 824–826.

Léonard, G., & Tremblay, F. (2007). Corticomotor facilitation associated with observation, imagery and imitation of hand actions: A comparative study in young and old adults. *Experimental Brain Research*, *177*(2), 167–175. https://doi.org/10.1007/s00221-006-0657-6

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). Multiple perspectives on word production. *Behavioral and Brain Sciences*, *22*(01), 61–69.

Li, S., Zhu, H., & Tian, X. (2020). Corollary Discharge Versus Efference Copy: Distinct Neural Signals in Speech Preparation Differentially Modulate Auditory Responses. *Cerebral Cortex*, 1–15. https://doi.org/10.1093/cercor/bhaa154

Lidstone, J. S. M., Fernyhough, C., Meins, E., & Whitehouse, A. J. O. (2009). Brief report: Inner speech impairment in children with autism is associated with greater nonverbal than verbal skills. *Journal of Autism and Developmental Disorders*, *39*(8), 1222–1225. https://doi.org/10.1007/s10803-009-0731-6

Locke, J. (1690). An Essay Concerning Human Understanding Book II -- Ideas. *Early Modern Texts*.

Lœvenbruck, H., Grandchamp, R., Rapin, L., Dohen, M., Perrier, P., & Baciu, M. (2018). A cognitive neuroscience view of inner language : to predict and to hear , see , feel. In P. Langland-Hassan & A. Vicente (Eds.), *Inner Speech: New Voices* (pp. 131–167). Oxford University Press. https://global.oup.com/academic/product/inner-speech-9780198796640?cc=fr%0Alang=en . hal-01898992%0AHAL

Lombard, E. (1911). Le signe de l'elevation de la voix. *Larynx, Nez, Pharynx*, *37*, 101–119.

Marr, D. (1985). Vision. A Computational Investigation into the Human Representation and Processing of Visual Information. In *The Philosophical Review* (Vol. 94, Issue 3). https://doi.org/10.2307/2185011

Mars, R. B., & Grol, M. J. (2007). Dorsolateral prefrontal cortex, working memory, and prospective coding for action. *Journal of Neuroscience*, *27*(8), 1801–1802. https://doi.org/10.1523/JNEUROSCI.5344-06.2007

Martínez-Manrique, F., & Vicente, A. (2015). The activity view of inner speech. *Frontiers in Psychology*, *6*(MAR), 1–13. https://doi.org/10.3389/fpsyg.2015.00232

McCarthy-Jones, S., & Fernyhough, C. (2011). The varieties of inner speech: Links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and Cognition*, *20*(4), 1586–1593. https://doi.org/10.1016/j.concog.2011.08.005

McGuigan, F. J., & Dollins, A. B. (1989). Patterns of covert speech behavior and phonetic coding. *The Pavlovian Journal of Biological Science*, *24*(1), 19–26. https://doi.org/10.1007/BF02964527

McGuire, P. K., Silbersweig, D. A., Wright, I., Murray, R. M., Frackowiak, R. S. J., & Frith, C. D. (1996). The Neural Correlates of Inner Speech and Auditory Verbal Imagery in Schizophrenia: Relationship to Auditory Verbal Hallucinations. *British Journal of Psychiatry*, *169*(2), 148–159. https://doi.org/10.1192/bjp.169.2.148

Meers, R., Nuttall, H. E., & Vogt, S. (2020). Motor imagery alone drives corticospinal excitability during concurrent action observation and motor imagery. *Cortex*, *126*, 322–333. https://doi.org/10.1016/j.cortex.2020.01.012

Meyer, B. U., Liebsch, R., & Röricht, S. (1997). Tongue motor responses following transcranial magnetic stimulation of the motor cortex and proximal hypoglossal nerve in man. *Electroencephalography and Clinical Neurophysiology - Electromyography and Motor Control*, *105*(1), 15–23. https://doi.org/10.1016/S0924-980X(96)96598-4

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, *9*(8), 1265–1279.

Morin, A. (2009). Self-awareness deficits following loss of inner speech: Dr. Jill Bolte Taylor’s case study. *Consciousness and Cognition*, *18*(2), 524–529. https://doi.org/10.1016/j.concog.2008.09.008

Möttönen, R., Farmer, H., & Watkins, K. E. (2010). Lateralization of motor excitability during observation of bimanual signs. *Neuropsychologia*, *48*(10), 3173–3177.

https://doi.org/10.1016/j.neuropsychologia.2010.06.033

Möttönen, R., & Watkins, K. E. (2012). Using TMS to study the role of the articulatory motor system in speech perception. *Aphasiology*, *26*(9), 1103–1118. https://doi.org/10.1080/02687038.2011.619515

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions : mental imagery as mental emulation. *Philosophical Transactions of the Royal Society*, *364*, 1273–1280. https://doi.org/10.1098/rstb.2008.0314

Muellbacher, W., Artner, C., & Mamoli, B. (1998). Motor evoked potentials in unilateral lingual paralysis after monohemispheric ischaemia. *Journal of Neurology Neurosurgery and Psychiatry*, *65*(5), 755–761. https://doi.org/10.1136/jnnp.65.5.755

Muellbacher, W., Boroojerdi, B., Ziemann, U., & Hallett, M. (2001). Analogous corticocortical inhibition and facilitation in ipsilateral and contralateral human motor cortex representations of the tongue. *Journal of Clinical Neurophysiology*, *18*(6), 550–558. https://doi.org/10.1097/00004691-200111000-00005

Muellbacher, W., Mathis, J., & Hess, C. W. (1994). Electrophysiological assessment of central and peripheral motor routes to the lingual muscles. *Journal of Neurology Neurosurgery and Psychiatry*, *57*(3), 309–315. https://doi.org/10.1136/jnnp.57.3.309

Mulder, T. (2007). Motor imagery and action observation: Cognitive tools for rehabilitation. *Journal of Neural Transmission*, *114*(10), 1265–1278. https://doi.org/10.1007/s00702-007-0763-z

Munzert, J., Lorey, B., & Zentgraf, K. (2009). Cognitive motor processes: The role of motor imagery in the study of motor representations. *Brain Research Reviews*, *60*(2), 306–326. https://doi.org/10.1016/j.brainresrev.2008.12.024

Murakami, T., Restle, J., & Ziemann, U. (2011). Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. *Neuropsychologia*, *49*(7), 2045–2054. https://doi.org/10.1016/j.neuropsychologia.2011.03.034

Murakami, T., Ugawa, Y., & Ziemann, U. (2013). Utility of TMS to understand the neurobiology of speech. *Frontiers in Psychology*, *4*(JUL), 1–7. https://doi.org/10.3389/fpsyg.2013.00446

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Neef, N. E., Linh Hoang, T. N., Neef, A., Paulus, W., & Sommer, M. (2015). Speech dynamics are coded in the left motor cortex in fluent speakers but not in adults who stutter. *Brain*, *138*(3), 712–725. https://doi.org/10.1093/brain/awu390

Neef, N. E., Paulus, W., Neef, A., von Gudenberg, A. W., & Sommer, M. (2011). Reduced intracortical inhibition and facilitation in the primary motor tongue representation of adults who stutter. *Clinical Neurophysiology*, *122*(9), 1802–1811. https://doi.org/10.1016/j.clinph.2011.02.003

Nichols, T., Das, S., Eickhoff, S., Evans, A., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M., Poldrack, R., Poline, J.-B., Proal, E., Thirion, B., Van Essen, D., White, T., & Yeo, B. T. T. (2016). Best Practices in Data Analysis and Sharing in Neuroimaging using MRI. *BioRxiv*. https://doi.org/10.1101/054262

Nooteboom, S. G. (2005). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication*, *47*(1–2), 43–58. https://doi.org/10.1016/j.specom.2005.02.003

Nozari, N. (2020). *A comprehension- or a production-based monitor? Response to Roelofs (2019)*. *3*, 1–21. https://doi.org/10.31234/osf.io/b4wpu

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*(1), 1–33.

Nuttall, H. E., Kennedy-Higgins, D., Devlin, J. T., & Adank, P. (2017). The role of hearing ability and speech distortion in the facilitation of articulatory motor cortex. *Neuropsychologia*, *94*(November 2016), 13–22. https://doi.org/10.1016/j.neuropsychologia.2016.11.016

Nuttall, H. E., Kennedy-Higgins, D., Hogan, J., Devlin, J. T., & Adank, P. (2016). The effect of speech distortion on the excitability of articulatory motor cortex. *NeuroImage*, *128*, 218–226. https://doi.org/10.1016/j.neuroimage.2015.12.038

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(24), 9868–9872. https://doi.org/10.1073/pnas.87.24.9868

Oishi, K., Kasai, T., & Maeshima, T. (2000). Autonomic response specificity during motor imagery. *Journal of Physiological Anthropology and Applied Human Science*, *19*(6), 255–261. https://doi.org/10.2114/jpa.19.255

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*(1), 528–537. https://doi.org/10.1016/j.cognition.2007.02.006

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory and Cognition*, *38*(8), 1147–1160. https://doi.org/10.3758/MC.38.8.1147

Page, S. J., & Harnish, S. (2012). Thinking about better speech: Mental practice for stroke-induced motor speech impairments. *Aphasiology*, *26*(2), 127–142. https://doi.org/10.1080/02687038.2011.636027

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

Parker, D., Liu, X., & Razlighi, Q. R. (2017). Optimal slice timing correction and its interaction with fMRI parameters and artifacts. *Medical Image Analysis*, *35*, 434–445. https://doi.org/10.1016/j.media.2016.08.006

Parsons, L. M. (1994). Temporal and kinematic properties of motor behavior reflected in mentally simulated action. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(4), 709–730. https://doi.org/10.1037/0096-1523.20.4.709

Parsons, L. M., Fox, P. T., Downs, J. H., Glass, T., Hirsch, T. B., Martin, C. C., Jerabek, P. A., & Lancaster, J. L. (1995). Use of implicit motor imagery for visual shape discrimination as revealed by PET. *Nature*, *375*(May), 4–8.

Penfield, W., & Rasmussen, T. (1950). Somatomotor responses. In *The cerebral cortex of man* (pp. 46–57). MacMilan.

Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciu, M., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, *261*, 220–239. https://doi.org/10.1016/j.bbr.2013.12.034

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, *36*(4), 329–347.

https://doi.org/10.1017/S0140525X12001495

Pickering, M. J., & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Frontiers in Human Neuroscience*, *8*(March), 132. https://doi.org/10.3389/fnhum.2014.00132

Poldrack, R. A., Nichols, T., & Mumford, J. (2011). Handbook of Functional MRI Data Analysis. In *Handbook of Functional MRI Data Analysis*. https://doi.org/10.1017/cbo9780511895029

Porro, C. A., Francescato, M. P., Cettolo, V., Diamond, M. E., Baraldi, P., Zuiani, C., Bazzocchi, M., & di Prampero, P. E. (1996). Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *16*(23), 7688–7698. https://doi.org/8922425

Pylyshyn, Z. W. (1984). *Computation and cognition*.

R Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.r-project.org/

Raichle, M. E. (2015). The Brain's Default Mode Network. *Annual Review of Neuroscience*, *38*, 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rapin, L., Dohen, M., Polosan, M., Perrier, P., & Lœvenbruck, H. (2013). An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia. *Journal of Speech, Language, and Hearing Research*, *56*(6), S1882--S1893.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. https://doi.org/10.1038/nn.2331

Richter, W., Somorjai, R., Summers, R., Jamarsz, M., Menon, R. S., Gati, J. S., Georgopoulos, A. P., Tegeler, C., Ugurbil, K., & Kim, S.-G. (2000). Motor Area Activity During Mental Rotation Studied by Time-Resolved Single-Trial fMRI. *Journal of Cognitive Neuroscience*, *12*(2), 310–320. https://doi.org/10.1162/089892900562129

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). *Neurophysiological mechanisms and imitation of action*. *2*(September), 1–10.

Rödel, R. M., Laskawi, R., & Markus, H. (2003). Tongue representation in the lateral cortical motor region of the human brain as assessed by transcranial magnetic stimulation. *Annals of Otology, Rhinology & Laryngology*, *112*(1), 71–76.

Roelofs, A. (2020a). *On (Correctly Representing) Comprehension-Based Monitoring in Speaking : Rejoinder to Nozari (2020)*. *3*, 1–7.

Roelofs, A. (2020b). *Self-Monitoring in Speaking : In Defense of a Comprehension-Based Account*. *3*(1), 1–13.

Rogasch, N. C., Daskalakis, Z. J., & Fitzgerald, P. B. (2015). Cortical inhibition of distinct mechanisms in the dorsolateral prefrontal cortex is related to working memory performance: A TMS-EEG study. *Cortex*, *64*, 68–77. https://doi.org/10.1016/j.cortex.2014.10.003

Roosink, M., & Zijdewind, I. (2010). Corticospinal excitability during observation and imagery of simple and complex hand tasks: Implications for motor rehabilitation. *Behavioural Brain Research*, *213*(1), 35–41. https://doi.org/10.1016/j.bbr.2010.04.027

Rösler, K. M. (2001). Transcranial magnetic brain stimulation: a tool to investigate central motor pathways. *News in Physiological Sciences : An International Journal of Physiology Produced Jointly by the International Union of Physiological Sciences and the American Physiological Society*, *16*(December), 297–302.

Rösler, K. M., & Magistris, M. R. (2012). The size of motor-evoked potentials: Influencing parameters and quantification. *Oxford Handbook of Transcranial Stimulation*, 1–13. https://doi.org/10.1093/oxfordhb/9780198568926.013.0009

Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., Avanzini, G., Bestmann, S., Berardelli, A., Brewer, C., Canli, T., Cantello, R., Chen, R., Classen, J., Demitrack, M., Di Lazzaro, V., Epstein, C. M., George, M. S., Fregni, F., Ilmoniemi, R., Jalinous, R., … Ziemann, U. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology*, *120*(12), 2008–2039. https://doi.org/10.1016/j.clinph.2009.08.016

Rossini, P. M., Barker, A., Berardelli, A., Caramia, M., Caruso, G., Cracco, R., Dimitrijevic, M., Hallett, M., Katayama, Y., Luecking, C., Maertens de Noordhout, A., Marsden, C., Murray, N., Rothwell, J., Swash, M., & Tomberg, C. (1994). Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an

IFCN. *Electroencephalography and Clinical Neurophysiology*, *91*, 79–92.

Rossini, P. M., Burke, D., Chen, R., Cohen, L. G., Daskalakis, Z., Di Iorio, R., Di Lazzaro, V., Ferreri, F., Fitzgerald, P. B., George, M. S., Hallett, M., Lefaucheur, J. P., Langguth, B., Matsumoto, H., Miniussi, C., Nitsche, M. A., Pascual-Leone, A., Paulus, W., Rossi, S., … Ziemann, U. (2015). Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: Basic principles and procedures for routine clinical and research application: An updated report from an I.F.C.N. Committee. *Clinical Neurophysiology*, *126*(6), 1071–1107. https://doi.org/10.1016/j.clinph.2015.02.001

Roth, M., Decety, J., Raybaudi, M., Massarelli, R., Delon-Martin, C., Segebarth, C., Morand, S., Gemignani, A., Decorps, M., & Jeannerod, M. (1996). Possible involvement of primary motor cortex in mentally simulated movement: a functional magnetic resonance imaging study. *Neuroreport*, *7*(7), 1280–1284.

Rothwell, J. C., Hallett, M., Berardelli, A., Eisen, A., Rossini, P. M., & Paulus, W. (1999). Magnetic stimulation: motor evoked potentials. *Electroencephalography and Clinical Neurophysiology Supplement*, *52*, 97–103.

Roy, A. C., Craighero, L., Fabbri-destro, M., & Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. *Journal of Physiology Paris*, *102*(1–3), 101–105. https://doi.org/10.1016/j.jphysparis.2008.03.006

Ruohonen, J. (2003). *Background physics for magnetic stimulation Jarmo Ruohonen* (Vol. 56). Elsevier B.V. https://doi.org/10.1016/S1567-424X(09)70204-1

Russell-Smith, S. N., Comerford, B. J. E., Maybery, M. T., & Whitehouse, A. J. O. (2014). Brief report: Further evidence for a link between inner speech limitations and executive function in high-functioning children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(5), 1236–1243. https://doi.org/10.1007/s10803-013-1975-8

Salmelin, R., Service, E., Kiesilä, P., Uutela, K., & Salonen, O. (1996). Impaired visual word processing in dyslexia revealed with magnetoencephalography. *Annals of Neurology*, *40*(2), 157–162. https://doi.org/10.1002/ana.410400206

Sato, M., Buccino, G., Gentilucci, M., & Cattaneo, L. (2010). On the tip of the tongue: Modulation of the primary motor cortex during audiovisual speech perception. *Speech Communication*, *52*(6), 533–541. https://doi.org/10.1016/j.specom.2009.12.004

Scott, M. (2013). Corollary discharge provides the sensory content of inner speech. *Psychological Science*, 0956797613478614.

Seal, M. L., Aleman, A., & McGuire, P. K. (2004). Compelling imagery, unanticipated speech and deceptive memory: Neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cognitive Neuropsychiatry*, *9*(1–2), 43–72. https://doi.org/10.1080/13546800344000156

Sharma, N., Pomeroy, V. M., & Baron, J. C. (2006). Motor imagery: A backdoor to the motor system after stroke? *Stroke*, *37*(7), 1941–1952. https://doi.org/10.1161/01.STR.0000226902.43357.fc

Shepard, R. N., & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects Abstract . The time required to recognize that two perspective drawings portray. *Science (New York, N.Y.)*, *171*(FEBRUARY), 701–703. https://doi.org/10.1126/science.171.3972.701

Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language*, *93*(1), 20–31. https://doi.org/10.1016/j.bandl.2004.07.007

Siegrist, M. (1995). Inner speech as a cognitive process mediating self-consciousness and inhibiting self-deception. *Psychological Reports*, *76*, 259–265.

Sokolov, A. (1972). *Inner speech and thought*. Plenum/Rosetta Press.

Solodkin, A., Hlustik, P., Chen, E. E., & Small, S. N. (2004). Fine Modulation in Network Activation during Motor Execution and Motor Imagery. *Cerebral Cortex*, *14*(11), 1246–1255. https://doi.org/10.1093/cercor/bhh086

Sperry, R. W. (1950). *Neural basis of the spontaneous optokinetic response produced by visual inversion*. 482–489.

Stamatakis, E. A., Orfanidou, E., & Papanicolaou, A. C. (2017). Physiological Processes Underlying the fMRI Signal. In A. C. Papanicolaou (Ed.), *The Oxford Handbook of Functional Brain Imaging in Neuropsychology and Cognitive Neurosciences* (pp. 1–26). https://doi.org/10.1093/oxfordhb/9780199764228.013.7

Stamenov, M. (2003). Language and self-consciousness: Modes of self-presentation in language structure. In *The self in neuroscience and psychiatry* (pp. 76–104).

Stippich, C., Ochmann, H., & Sartor, K. (2002). Somatotopic mapping of the human primary sensorimotor cortex during motor imagery and motor execution by functional magnetic resonance imaging. *Neuroscience Letters*, *331*(1), 50–54.

papers3://publication/uuid/0432B5E0-D0FF-4033-8BF7-D82E0AA4ECD4

Strange, W., & Jenkins, J. J. (2013). Dynamic Specification of Coarticulated Vowels. In G. S. Morrison & P. F. Assmann (Eds.), *Vowel Inherent Spectral Change* (pp. 87–115). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14209-3_5

Strange, W., Jenkins, J. J., & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*, *74*(3), 695–705. https://doi.org/10.1121/1.389855

Svensson, P., Romaniello, A., Wang, K., Arendt-Nielsen, L., & Sessle, B. J. (2006). One hour of tongue-task training is associated with plasticity in corticomotor control of the human tongue musculature. *Experimental Brain Research*, *173*(1), 165–173. https://doi.org/10.1007/s00221-006-0380-3

Svensson, Peter, Romaniello, A., Arendt-Nieisen, L., & Sessle, B. J. (2003). Plasticity in corticomotor control of the human tongue musculature induced by tongue-task training. *Experimental Brain Research*, *152*(1), 42–51. https://doi.org/10.1007/s00221-003-1517-2

Szenkovits, G., Peelle, J. E., Norris, D., & Davis, M. H. (2012). Individual differences in premotor and motor recruitment during speech perception. *Neuropsychologia*, *50*(7), 1380–1392. https://doi.org/10.1016/j.neuropsychologia.2012.02.023

Takaso, H., Eisner, F., Wise, R. J. S., & Scott, S. K. (2010). The effect of delayed auditory feedback on activity in the temporal lobe while speaking: a positron emission tomography study. *Journal of Speech, Language, and Hearing Research*, *53*(2), 226–236.

Théoret, H., Halligan, E., Kobayashi, M., Fregni, F., Tager-Flusberg, H., & Pascual-Leone, A. (2005). Impaired motor facilitation during action observation in individuals with autism spectrum disorder [2]. *Current Biology*, *15*(3), 84–85. https://doi.org/10.1016/j.cub.2005.01.022

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, *1*(October), 1–23. https://doi.org/10.3389/fpsyg.2010.00166

Tian, X., & Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Mental Imagery*, 152.

Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *Journal of Cognitive*

Neuroscience, *25*(7), 1020–1036.

Tian, X., & Poeppel, D. (2014). Dynamics of Self-monitoring and Error Detection in Speech Production: Evidence from Mental Imagery and MEG. *Journal of Cognitive Neuroscience*, *27*(2), 352–364. https://doi.org/10.1162/jocn_a_00692

Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*, *77*, 1–12. https://doi.org/10.1016/j.cortex.2016.01.002

Tomasino, B., Borroni, P., Isaja, A., & Rumiati, R. I. (2005). The role of the primary motor cortex in mental rotation: A TMS study. *Cognitive Neuropsychology*, *22*(3–4), 348–363. https://doi.org/10.1080/02643290442000185

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Tremblay, F., Tremblay, L. E., & Colcer, D. E. (2001). Modulation of corticospinal excitability during imagined knee movements. *Journal of Rehabilitation Medicine*, *33*(5), 230–234. https://doi.org/10.1080/165019701750419635

Tremblay, P., Deschamps, I., & Gracco, V. L. (2015). Neurobiology of Speech Production: A Motor Control Perspective. *Neurobiology of Language*, 741–750. https://doi.org/10.1016/B978-0-12-407794-2.00059-6

Tremblay, P., & Gracco, V. L. (2009). Contribution of the pre-SMA to the production of words and non-speech oral motor gestures, as revealed by repetitive transcranial magnetic stimulation (rTMS). *Brain Research*, *1268*, 112–124. https://doi.org/10.1016/j.brainres.2009.02.076

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, *423*(6942), 866–869.

Vicario, C. M., Komeilipoor, N., Cesari, P., Rafal, R. D., & Nitsche, M. A. (2014). Enhanced corticobulbar excitability in chronic smokers during visual exposure to cigarette smoking cues. *Journal of Psychiatry and Neuroscience*, *39*(4), 232–238. https://doi.org/10.1503/jpn.130086

Vicario, C. M., Rafal, R. D., Borgomaneri, S., Paracampo, R., Kritikos, A., & Avenanti, A. (2016). Pictures of disgusting foods and disgusted facial expressions suppress the tongue motor cortex. *Social Cognitive and Affective Neuroscience*, *12*(2), 352–362. https://doi.org/10.1093/scan/nsw129

Vingerhoets, G., De Lange, F. P., Vandemaele, P., Deblaere, K., & Achten, E. (2002). Motor imagery in mental rotation: an fMRI study. *Neuroimage*, *17*(3), 1623–1633.

Vingerhoets, G., Santens, P., Van Laere, K., Lahorte, P., Dierckx, R. A., & De Reuck, J. (2001). Regional brain activity during different paradigms of mental rotation in healthy volunteers: A positron emission tomography study. *NeuroImage*, *13*(2), 381–391. https://doi.org/10.1006/nimg.2000.0690

Vogt, S., Di Rienzo, F., Collet, C., Collins, A. A., & Guillot, A. (2013). Multiple roles of motor imagery during action observation. *Frontiers in Human Neuroscience*, *7*(November), 807. https://doi.org/10.3389/fnhum.2013.00807

von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip - Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Die Naturwissenschaften*, *37*(20), 464–476. https://doi.org/10.1007/BF00622503

Vygotsky, L. (1986). Thought and Language - Revised Edition. In *The Journal of Applied Behavioral Science*.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989–994. https://doi.org/10.1016/S0028-3932(02)00316-0

Weiss, C., Nettekoven, C., Rehme, A. K., Neuschmelting, V., Eisenbeis, A., Goldbrunner, R., & Grefkes, C. (2013). Mapping the hand, foot and face representations in the primary motor cortex - Retest reliability of neuronavigated TMS versus functional MRI. *NeuroImage*, *66*, 531–542. https://doi.org/10.1016/j.neuroimage.2012.10.046

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, *68*(1), 77–94. https://doi.org/10.1016/S0010-0277(98)00032-8

Whillier, A., Hommel, S., Neef, N. E., Von Gudenberg, A. W., Paulus, W., & Sommer, M. (2018). Adults who stutter lack the specialised prespeech facilitation found in non-stutterers. *PLoS ONE*, *13*(10), 1–26. https://doi.org/10.1371/journal.pone.0202634

Whitehouse, A. J. O., Maybery, M. T., & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *47*(8), 857–865. https://doi.org/10.1111/j.1469-7610.2006.01624.x

Williams, D. M., & Jarrold, C. (2010). Brief report: Predicting inner speech use amongst children with autism spectrum disorder (ASD): The roles of verbal ability and cognitive profile. *Journal of Autism and Developmental Disorders*, *40*(7), 907–913. https://doi.org/10.1007/s10803-010-0936-8

Wilson, M., & Knoblich, G. (2005). The Case for Motor Involvement in Perceiving Conspecifics. *Psychological Bulletin*, *131*(3), 460–473. https://doi.org/10.1037/0033-2909.131.3.460

Wolpert, D. M., Ghahramani, Z., & Jordan, M. (1995). An Internal Model for Sensorimotor Integration. *Science*, *269*(5232), 1880–1882.

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

Wraga, M., Thompson, W. L., Alpert, N. M., & Kosslyn, S. M. (2003). Implicit transfer of motor strategies in mental rotation. *Brain and Cognition*, *52*(2), 135–143. https://doi.org/10.1016/S0278-2626(03)00033-2

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0068910

Yahagi, S., Shimura, K., & Kasai, T. (1996). An increase in cortical excitability with no change in spinal excitability during motor imagery. *Perceptual and Motor Skills*, *83*(1), 288–290. https://doi.org/10.2466/pms.1996.83.1.288

Zapparoli, L., Invernizzi, P., Gandola, M., Berlingeri, M., De Santis, A., Zerbi, A., Banfi, G., & Paulesu, E. (2014). Like the back of the (right) hand? A new fMRI look on the hand laterality task. *Experimental Brain Research*, *232*(12), 3873–3895. https://doi.org/10.1007/s00221-014-4065-z

Zhang, W., Liu, Y., Wang, X., & Tian, X. (2020). The dynamic and task-dependent representational transformation between the motor and sensory systems during speech production. *Cognitive Neuroscience*, *00*(00), 1–11. https://doi.org/10.1080/17588928.2020.1792868

Zhang, Y., Boudreau, S., Wang, M., Wang, K., Sessle, B., Arendt-Nielsen, L., & Svensson, P. (2010). Effects of periodontal afferent inputs on corticomotor excitability in humans. *Journal of Oral Rehabilitation*, *37*(1), 39–47. https://doi.org/10.1111/j.1365-2842.2009.02016.x

List of appendices

Appendix A

Appendix A contains the learning and correlation analyses for each experiment.

Experiment 1

Learning within tasks

To investigate whether learning took place within each task as participants became more exposed to the stimuli, a 2 x 4 repeated-measures ANOVA was performed with factors Task and Block, where block refers to the nth block for a task rather than overall block. 48 trials were presented per block, for 192 trials in total. The term 'learning' is used as a general moniker for improved performance: this could refer to improved performance due to improved ability to imagine but could also refer to general task habituation.

Learning: Reaction time

There was a significant main effect of Task ($F(1,43) = 6.44$, $p = .015$, $\eta_p^2 = .130$) once again showing that the visual task showed significantly shorter RTs (M = 1150ms, SD = 223ms) than the auditory task (M = 1217ms, SD = 222ms). There was also a significant effect of Block ($F(1.726,74.238) = 46.14$, $p < .001$, $\eta_p^2 = .518$) showing clearly that learning took place across tasks, with post-hoc comparisons showing significantly shorter RTs from one block to the next (all comparisons p <.001), except between the third and fourth (final) block (see Figure A.1 and Table A.1)

Table A.1: Estimate statistics for main effect of Block in experiment 1

Estimate statistics			
Block	*Mean (ms)*	*SD(ms)*	*95% CI(ms)*
1	1290	206	1128,1353
2	1199	213	1135,1264
3	1131	190	1074,1189
4	1114	199	1053,1174

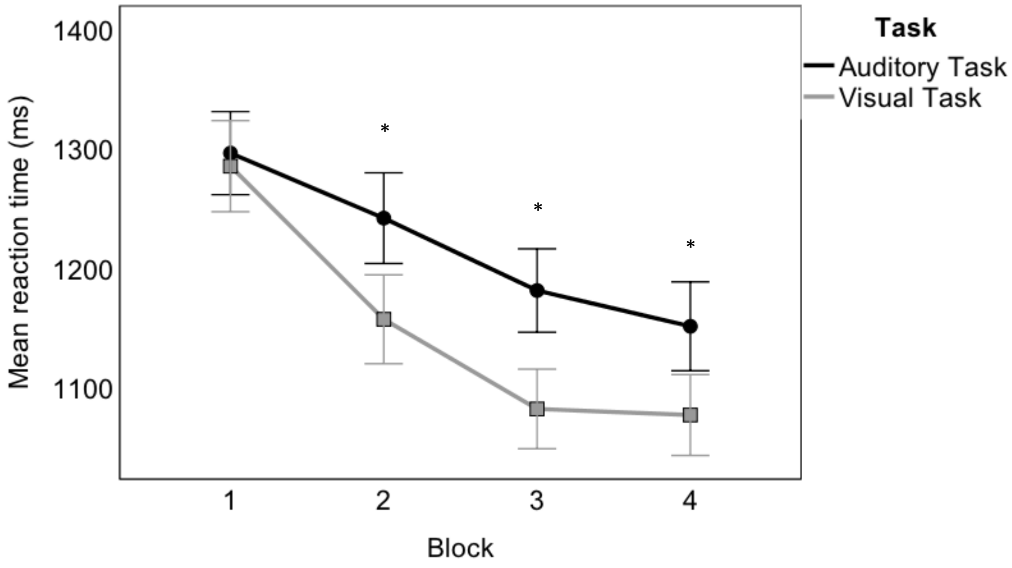


Figure A.2: Differences per block across task, showcasing the interaction effect in experiment 1. Error bars indicate +-1SEM. * = significant at p = .05

Additionally, there was a significant interaction Task x Block (F(2.301, 98.925) = 4.94, p = .006, η_p^2 = .103), which revealed a difference in learning between tasks relative to block. Follow-up comparisons showed that this was due to a difference in rate of change in both tasks: while reaction times were similar in block 1 (near the start of the experiment), by block 2 reaction times in the visual task had been reduced significantly more than in the auditory task, a trend which continued into blocks 3 and 4 (see Figure A.2). For each

Table A.2: Selected post-hoc pairwise comparisons in each task showing differences from block one through to block 4 in experiment 1.

Pairwise Comparisons per task

Task	*Block vs*	*Block*	*MD (ms)*	*SEM (ms)*	*p*	*95% CI for difference (ms)*
Auditory	1	2	55	21	0.030	-4,113
		3	115	28	0.000	38,192
		4	145	30	0.000	61,229
	2	3	61	18	0.008	12,110
		4	91	22	0.001	30,151
	3	4	30	16	0.236	-15,75
Visual	1	2	128	19	0.000	75,182
		3	204	25	0.000	134,273
		4	209	27	0.000	133,284
	2	3	75	16	0.000	31,119
		4	80	20	0.000	26,135
	3	4	5	14	1.000	-33,43

Based on estimated marginal means

p = 0.05 and 95% CI adjusted for multiple comparisons (Bonferroni).

task, each block was significantly faster than the previous block, except for the final block which was non-significantly different to the third block (see Table A.2).

This combination of findings explains the interaction, showing (a) faster learning rate in the visual task relative to the auditory task, and (b) differences from one block to the following block, save for the penultimate and final block.

Learning: Accuracy

An ANOVA was also performed on the accuracy data. Similar to the reaction time analysis, there was again a main effect of Task ($F(1,43)=38.99$, $p < .001$, $\eta_p^2 = .476$) due to the aforementioned differences in accuracy between the visual ($M = 89.3\%$, $SD = 8.6\%$) and the auditory task ($M = 96.7\%$, $SD = 3.3\%$), but here there was no effect of Block ($F(2.288, 98.386) = 1.66$, $p = .193$, $\eta_p^2 = .037$), showing no evidence of a learning effect between blocks. Accuracy remained constant across blocks, regardless of task. There was also no interaction ($F(3, 129) = 0.11$, $p = .956$, $\eta_p^2 = .002$), showing that changes in accuracy (or rather the lack of significant change) showed a similar pattern in both tasks.

Correlational analyses

Correlation analyses were performed to investigate whether there was a relationship between MOCA & VISQ scores, and reaction time or accuracy in either task. No correlations were found between reaction time and any of the scores, and no significant correlations were found between accuracy scores and any of the MOCA or VISQ scores.

Experiment 2

Learning within tasks

To investigate whether learning took place within each task as participants became more exposed to the stimuli, a 2 x 3 repeated-measures ANOVA was performed with factors Task and Block.

Learning: Reaction time

There was a significant main effect of Task (*F*(1,33) = 14.13, *p* = .001, η_p^2 = .300) once again showing that the visual task showed significantly shorter RTs (*M* = 1328ms, *SD* = 366ms) than the auditory task (*M* = 1586ms, *SD* = 504ms). There was also a significant effect of Block (*F*(2,66) = 4.10, *p* = .021, η_p^2 = .111) however unlike experiment 1, this appears to be driven only by the difference between block 1 and 3 (*MD* = 136ms, *SD* = 271ms, *p* =.039), as all other comparisons are non-significant (block 1 vs 2: *MD* = 101ms, *SD* = 221ms, *p* = .335; block 2 vs 3: *MD* = 35ms, *SD* = 222ms, *p* = .556) (see Figure A.3).

Table A.3: Post-hoc comparisons showing differences between blocks across tasks, showcasing potential learning effects in experiment 2.

Pairwise Comparisons of Block across tasks

Task	*Block vs Block*	*MD(ms)*	*SEM(ms)*	*p*	*95% CI (ms)*
Auditory					
	1 vs 2	169	55	0.049	29, 308
	1 vs 3	220	84	0.013	7, 433
	2 vs 3	51	57	0.265	-96, 199
Visual					
	1 vs 2	33	47	1.000	-85, 151
	1 vs 3	52	48	1.000	-70, 174
	2 vs 3	19	40	1.000	-83, 121

Based on estimated marginal means

p = 0.05 and 95% CI adjusted for multiple comparisons (Bonferroni).

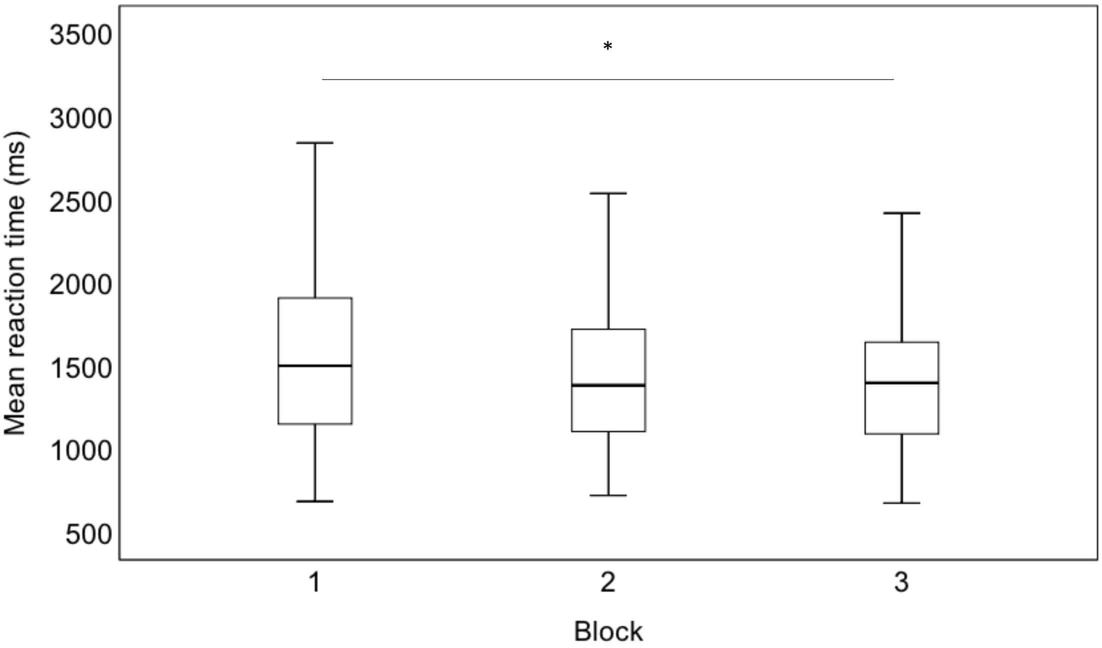


Figure A.3: Main effect of block in reaction times for experiment 2. * = significant at *p* = .05.

There was also a significant interaction Task x Block (*F*(1.697,56.004) = 4.25, *p* = .024, η_p^2 = .114). Follow-up comparisons showed that some learning appeared to have taken place in the auditory task, whereas no learning whatsoever took place in the visual task (see Table A.3). The visual task always showed faster reaction times: the difference between the two tasks was significant in the first (*MD* = 370ms, *SD* = 533ms, *p* < .001), second (*MD* = 234ms, *SD* = 484ms, *p* = .004), and third block (*MD* = 202ms, *SD* = 518ms, *p* = .028, all Bonferroni -corrected for multiple comparisons), see Figure A.4. Given these results, it is likely that the main effect of Block was driven primarily by the changes from one block to the next in the auditory task, since no changes were found for the visual task.

Learning: Accuracy

An ANOVA was also performed on the accuracy data. There was again a main effect of Task (*F*(1,33) = 37.95, *p* < .001, η_p^2 = .535) due to the differences in accuracy between the visual (*M* = 92.3%, *SD* = 8.7%) and the auditory task (*M* = 82.0%, *SD* = 7.6%), but here there was no effect of Block (*F*(2,66) = .05, *p* = .950, η_p^2 = .002), showing no evidence of a learning effect between the first (*M* = 87.2%, *SD* = 6.9%), second (*M* = 87.4%, *SD* = 8.2%), and third block (*M* = 86.9%, *SD* = 9.9%). Unlike reaction time, accuracy remained constant across blocks, regardless of task. There was also no interaction Task x Block (*F*(2, 68) = 0.45, *p* = .641, η_p^2 = .013), showing that the lack of accuracy changes was the same in both tasks.

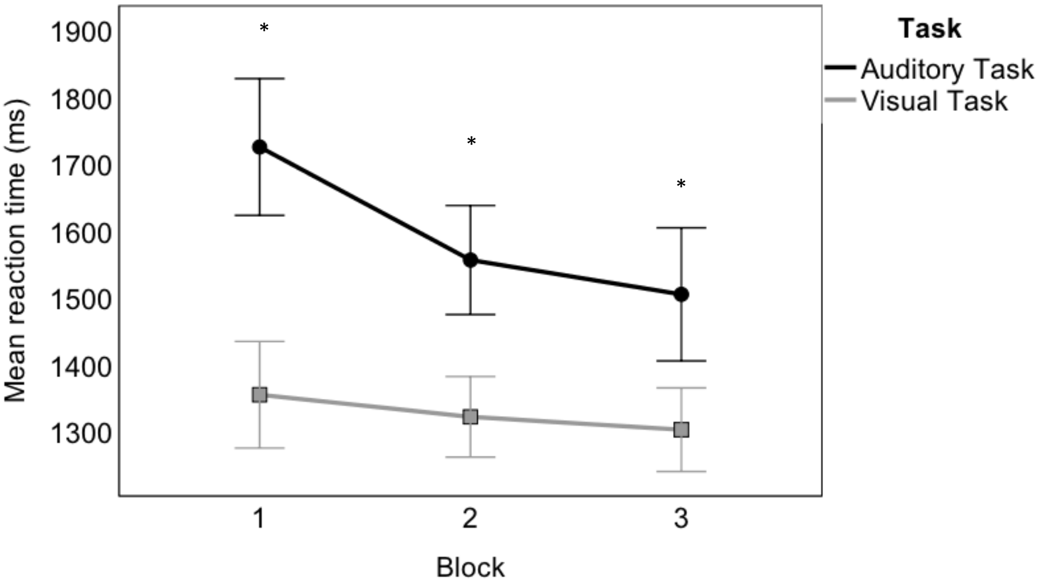


Figure A.4: Differences in reaction time per block across task, showcasing the interaction effect in experiment 2. Error bars indicate +- 1SEM. * = significant at *p* = .05.

Correlational analyses

Correlation analyses were performed to investigate whether there was a relationship between VISQ scores and reaction time or accuracy in either task. No correlations were found between reaction time and any of the scores, however there was a correlation between some accuracy scores and factor 3 (other people). There was a significant medium negative correlation between accuracy in the auditory task and the score reporting how often other people were heard in other speech: this was true for both the easy stimuli (Spearman's *rho* (34) = –.540, *p* = .001) and for hard stimuli (Spearman's *rho* (34) = –.574, *p* < .001, both significant after Bonferroni correction, see Figure A.5).

As previously noted, a high score on factor 3 (other people) suggests many experiences with other people in inner speech, where a low score suggests few experiences – a lower score does not suggest an intrinsically higher amount of inner speech *per se* but suggests that any inner speech involves only oneself. The correlations show that those who experience others' voices in their inner speech are likely to be less accurate in the auditory task than those who report hearing no voices (or no other voices than their own).

Experiment 3

Learning within tasks

Learning: Reaction Time Tasks

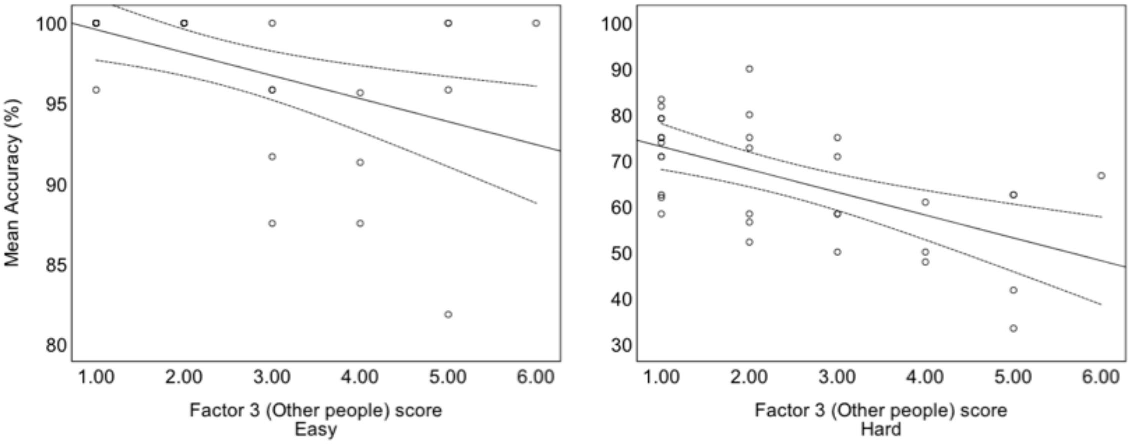


Figure A.5: Correlations between accuracy scores and VISQ factors in experiment 2. Lines shown are linear best fit and 95% confidence interval.

To investigate whether learning took place in the simple reaction time task, a 2 x 4 repeated-measures ANOVA was performed with factors Task (visual, auditory) and Block (1-4). Changes in reaction time from block 1 to block 4 here would reflect a simple change in task performance, not a change in decision-making time as the reaction time task does not require decision-making. One participant showed no responses for the reaction time task in the 4th block due to equipment issues. Their data was excluded, leading to an *n* of 16.

The ANOVA showed that there was a significant main effect of Task ($F(1,15) = 9.51$, $p = .008$, $\eta_p^2 = .388$) which was due to consistently longer reaction times in the auditory task (*M* = 583ms, *SD* = 408ms) compared to the visual task (*M* = 385ms, *SD* = 124ms). In this case, the effect can be attributed to the slightly slower onset of audible sound stimuli while in the scanner undergoing an fMRI protocol, compared to the faster onset of visual stimuli with which the scanner operations would not interfere. There was no significant main effect of Block ($F(1.848,27.721) = 1.15$, $p = .327$, $\eta_p^2 = .071$), showing that across tasks, learning did not occur. However, there was an interaction ($F(3,45) = 4.345$, $p = .009$, $\eta_p^2 = .225$, see Figure A.6).

Follow-up comparisons showed that no learning took place across blocks for the visual task (all differences $p = 1.000$), but learning did take place in the auditory task from block

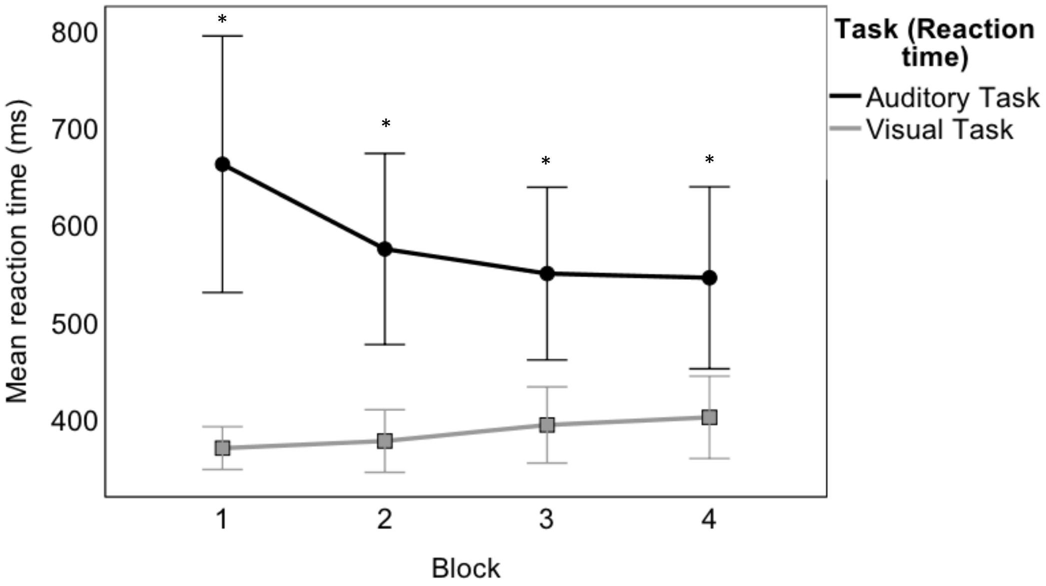


Figure A.6: Differences across blocks in the reaction time tasks in experiment 3. Error bars indicate +- 1SEM. * = significant at $p = .0.5$.

1 to 2 (MD = 87ms, SD = 156ms, p = .043, Hedges' g = -0.160) and from block 1 to 3 (MD = 113ms, SD = 189ms, p = .009, Hedges' g = -0.235) but no further (see also Table A.4). This interaction is likely due to the different trends in the auditory task. This is likely due to habituation of reacting as fast as possible while learning to ignore scanner noise.

Learning: Imagery Tasks - Reaction time

A 2 x 4 repeated-measures ANOVA was performed to compare reaction times per Task (auditory vs imagery task) across Blocks (1 to 4) as a potential measure of learning. Surprisingly, there was no main effect of Task ($F(1,16) = 2.28$, $p = .150$, $\eta_p^2 = .125$). This represents a difference with respect to both experiment 1 and 2, where an effect of Task was found. There was however, once again, a significant main effect of Block ($F(3,48) = 8.11$, $p < .001$, $\eta_p^2 = .336$). This effect was driven by the decrease in reaction time between

Table A.4: Post-hoc comparisons showing differences between tasks per block in experiment 3 (reaction time task), showing improved performance

Pairwise Comparisons of Block across tasks (log10 reaction time)

Block	*Task comparison*	*MD(ms)*	*SEM(ms)*	*p*	*95% CI (ms)*
1	Auditory vs Visual	293	123	,004	30, 555
2	Auditory vs Visual	198	87	.009	12, 384
3	Auditory vs Visual	156	70		6, 306
4	Auditory vs Visual	144	81		-29, 318

Based on estimated marginal means

p = 0.05 and 95% CI adjusted for multiple comparisons (Bonferroni).

Table A.5: Post-hoc comparisons for effect of Block for reaction times in experiment 3.

Block vs	Block	MD(ms)	SEM(ms)	*p*	95% CI (ms)
1	2	163	68	0.181	-42, 368
	3	148	74	0.136	-74, 369
	4	238	65	0.004	42, 433
2	1	-163	68	0.181	-368, 42
	3	16	54	1.000	-179, 148
	4	75	40	0.077	-47, 196
3	1	-148	74	0.136	-369, 74
	2	16	54	1.000	-148, 179
	4	90	42	0.143	-35, 215
4	1	-238	65	0.004	-433, -43
	2	-75	40	0.077	-196, 47
	3	-90	42	0.143	-215, 35

Based on estimated marginal means

p = 0.05 and 95% CI adjusted for multiple comparisons (Bonferroni).

blocks 1 and 4, with other comparisons not surviving Bonferroni corrections (see Table A.5). Unlike experiments 1 and 2, there is no significant interaction Task x Block (*F*(1.893,30.291) = .198, *p* = .897, η_p^2 = .012), showing that in experiment 3, learning took place across tasks at a similar rate.

Learning: Imagery Tasks - Accuracy

A 2 x 4 repeated-measures ANOVA was also performed to compare accuracy per Task (auditory vs imagery task) across Blocks (1 to 4).

The results from the accuracy data showed a very different pattern to the reaction time data. There was a main effect of Task (*F*(1,16) = 15.58, *p* = .001, η_p^2 = .493) as also found in experiments 1 and 2, but not in the reaction time data for experiment 3. This result showed that accuracy was significantly lower in the auditory task (*M* = 78.6%, *SD* = 9.5%) than in the visual task (*M* = 90.4%, *SD* = 10.7%). Similar to experiment 1 and 2, there was no effect of Block (*F*(3,48) = 0.82, *p* = .487, η_p^2 = .049), as there was no difference in accuracy between block 1 and 2 (*M* < 0.1, *SEM* < 0.1, *p* = 1.000), block 2 and 3 (*M* < 0.1, *SEM* < 0.1, *p* = 1.000) and block 3 and 4 (*M* < 0.1, *SEM* < 0.1, *p* = 1.000), or any comparisons (all *p* = 1.000). There was no interaction effect Task x Block (*F*(3,48) = 1.09, *p* = .359, η_p^2 = .064, n.s.), similar to the lack of such an effect in experiments 1 and 2, showing that the pattern was similar regardless of task.

Correlational analyses

Correlation analyses were performed to investigate whether there was a relationship between VISQ scores and reaction time or accuracy in either task. As previously mentioned, only 11 sets of data were available. Therefore these results are presented for completeness but are interpreted with caution.

One significant positive correlation was found between reaction time of easy vowels and VISQ factor 3 (Other people) (Spearman's *rho* (11) = .866, *p* = .001, see Figure A.7). No other correlations survived Bonferroni correction.

This result suggests there is some positive relationship between the extent to which inner speech takes the form of a dialogue (with oneself) and accuracy in the auditory task, specifically the difficult stimuli. However, the figure makes it clear that more data is needed before such a conclusion is drawn, as there are no factor 3 scores of 3, 4 or 6.

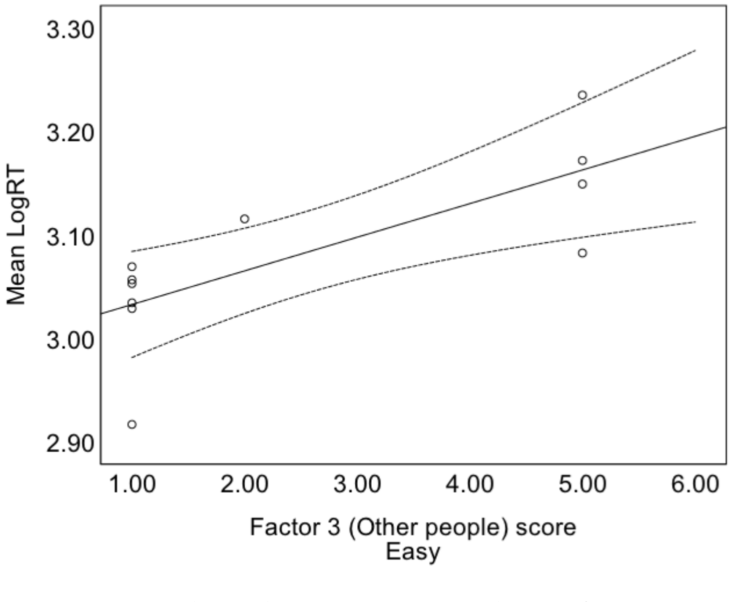


Figure A.7: Correlation between LogRT and VISQ factor 3 in experiment 3. Lines shown are linear best fit and 95% confidence interval.

Experiment 4

Correlational analyses

Analyses pertaining to correlations between scores on the VISQ, MOCA, average pre-pulse baseline size per condition and average MEP size per condition were performed. Expected were correlations between pre-pulse baseline sizes and MEP sizes, as the amount of pressure exerted by muscles is known to correlate positively with the size of MEP regardless of task (i.e., if participants exert more force using the relevant muscle, a larger MEP is expected regardless of task).The focus of this analysis was to see whether there was a relationship between experiences of inner speech (along VISQ factors 1 (dialogic inner speech), 2 (condensed inner speech), 3 (other people in inner speech), 4 (evaluative / motivational inner speech) or a combined score) and MEP changes, for instance, whether if participants report a large amount of condensed inner speech, their lip imagery MEP size increases.

Overall, there were no correlations relating to MOCA, indicating that MOCA score had no relation to MEP size or indeed VISQ scores. There was an expected correlation between pre-pulse baseline results and lip MEP size, but surprisingly this was not the case for hand MEPs (see Table A.6 and Figure A.8). That is, there was a positive correlation between how hard participants pressed to maintain pressure, and the size of their MEP in all conditions) but this was only the case for the lip data, regardless of condition (see Figure A.9). The correlation analysis also returned another significant correlation only for VISQ factor 'Condensed inner speech' and lip MEP data. This was further investigated by examining not only the lip conditions, but also the lip timepoints.

Table A.6: Correlations between pre-pulse activity and MEP size

Comparison	*Condition*	*Correlation Coefficient (rho)*	*p*
Lip pre-pulse window - Lip MEP	Execution	.755	<.001
	Imagery	.782	<.001
	Baseline	.731	<.001
Hand pre-pulse window - Hand MEP	Execution	-0.044	0.855
	Imagery	-0.102	0.688
	Baseline	-0.71	0.767

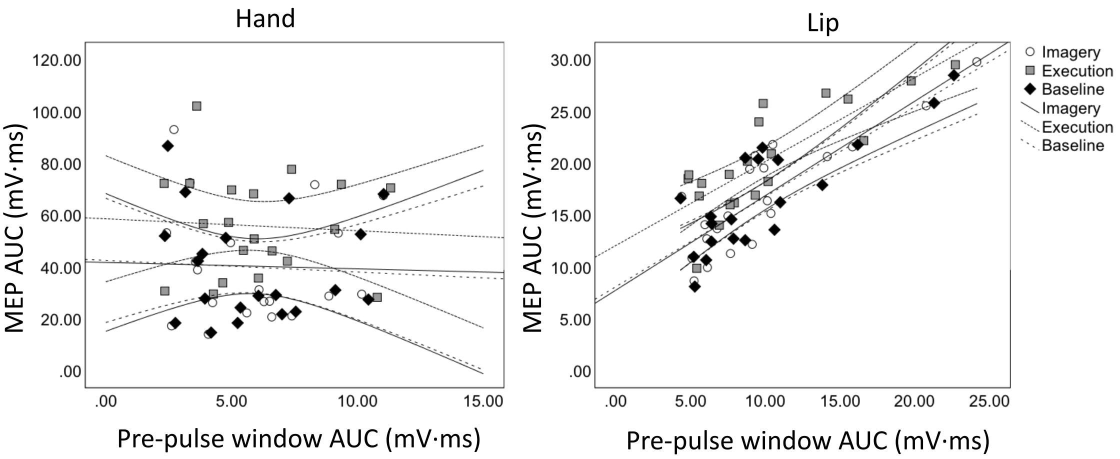


Figure A.8: Correlations between pre-pulse baseline and MEP size. Lines shown are linear best fit and 95% confidence interval per condition.

The results indicate a moderately positive correlation between the score on questions constituting dialogic inner speech and the size of lip MEPs at each time point (see Figure A.10 and Table A.7). In other words, as the (self-reported) use of dialogic inner speech increases, so does MEP size. This effect does not appear to be condition-specific however, as similarly strong correlations exist in all conditions.

Table A.7: Correlation results for experiment 4 (lip data only). Note that significance is not Bonferroni-corrected due to the increased Type-II error rate associated with correcting large amounts of comparisons

VISQ Factor	*Condition*	*Timepoint*	*Correlation Coefficient (rho)*	*p*
Factor 2 (Condensed inner speech score)	Motor execution	50	0.392	0.087
		150	0.601	0.005
		250	0.515	0.020
		350	0.540	0.014
		450	0.515	0.020
		550	0.491	0.028
	Motor imagery	50	0.540	0.014
		150	0.564	0.010
		250	0.515	0.020
		350	0.503	0.024
		450	0.466	0.038
		550	0.515	0.020
	Baseline	50	0.650	0.002
		150	0.515	0.020
		250	0.466	0.038
		350	0.491	0.028
		450	0.405	0.077
		550	0.454	0.044

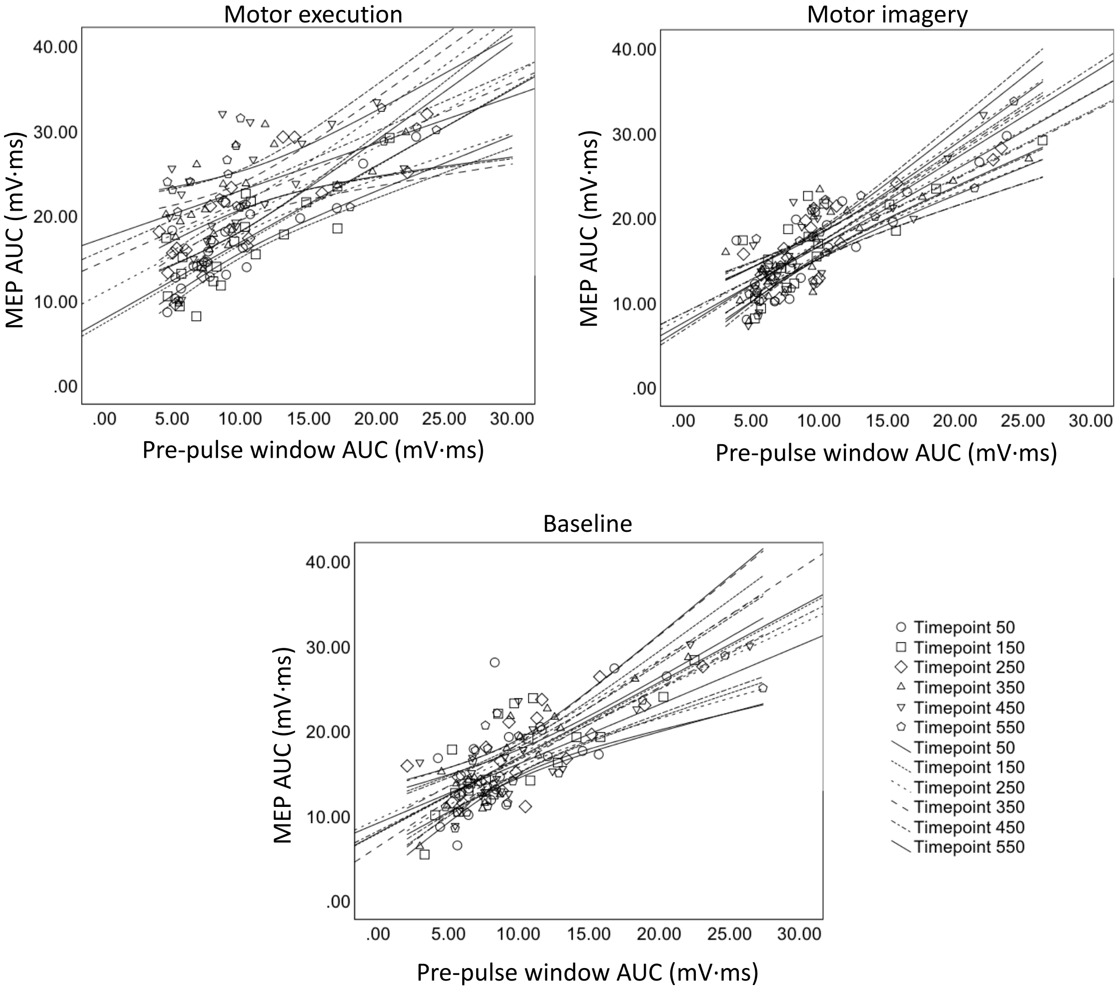


Figure A.9: Correlations of lip pre-pulse window AUC and MEP AUC. Lines shown are linear best fit and 95% confidence interval per timepoint.

Results were not Bonferroni-corrected as the number of comparisons is large and reduces the significance threshold to the point where it may not be useful (Cabin & Mitchell, 2000). The results should be seen as indicative only.

Since there was no relation between VISQ Factor 2 and pre-pulse window AUC, but there was a strong correlation between pre-pulse window AUC and MEP AUC, the correlation analysis of VISQ Factor 2 and MEP AUC was performed as a partial correlation, controlling for pre-pule window baseline AUC. Table A.8 shows that in a majority of timepoint and conditions the correlation coefficient decreased as a result of controlling for pre-pulse window baseline AUC, although many low to moderately strong, if not always significant. In sum, one factor from the VISQ showed a positive correlation with lip MEP size, but due to the marginal nature of this effect and the range of the condensed inner speech score, the correlation should not be overinterpreted.

Table A.8: Partial correlation results for experiment 4 (lip data only) controlling for baseline AUC levels. Note that significance is not Bonferroni-corrected due to the increased Type-II error rate associated with correcting large amounts of comparisons

VISQ Factor	*Condition*	*Time point*	*Correlation Coefficient ($rho_{partial}$)*	*Coefficient difference ($rho_{partial}$ - rho)*	*p*
Factor 2 (Condensed inner speech score)	Motor execution	50	0.297	-0.095	0.217
		150	0.532	-0.069	0.019
		250	0.512	-0.003	0.025
		350	0.545	0.005	0.016
		450	0.398	-0.117	0.092
		550	0.336	-0.155	0.160
	Motor imagery	50	0.475	-0.065	0.040
		150	0.533	-0.031	0.019
		250	0.472	-0.043	0.041
		350	0.343	-0.16	0.150
		450	0.370	-0.096	0.119
		550	0.479	-0.036	0.038
	Baseline	50	0.629	-0.021	0.004
		150	0.410	-0.105	0.081
		250	0.521	0.055	0.022
		350	0.176	-0.315	0.470
		450	0.434	0.029	0.063
		550	0.442	-0.012	0.058

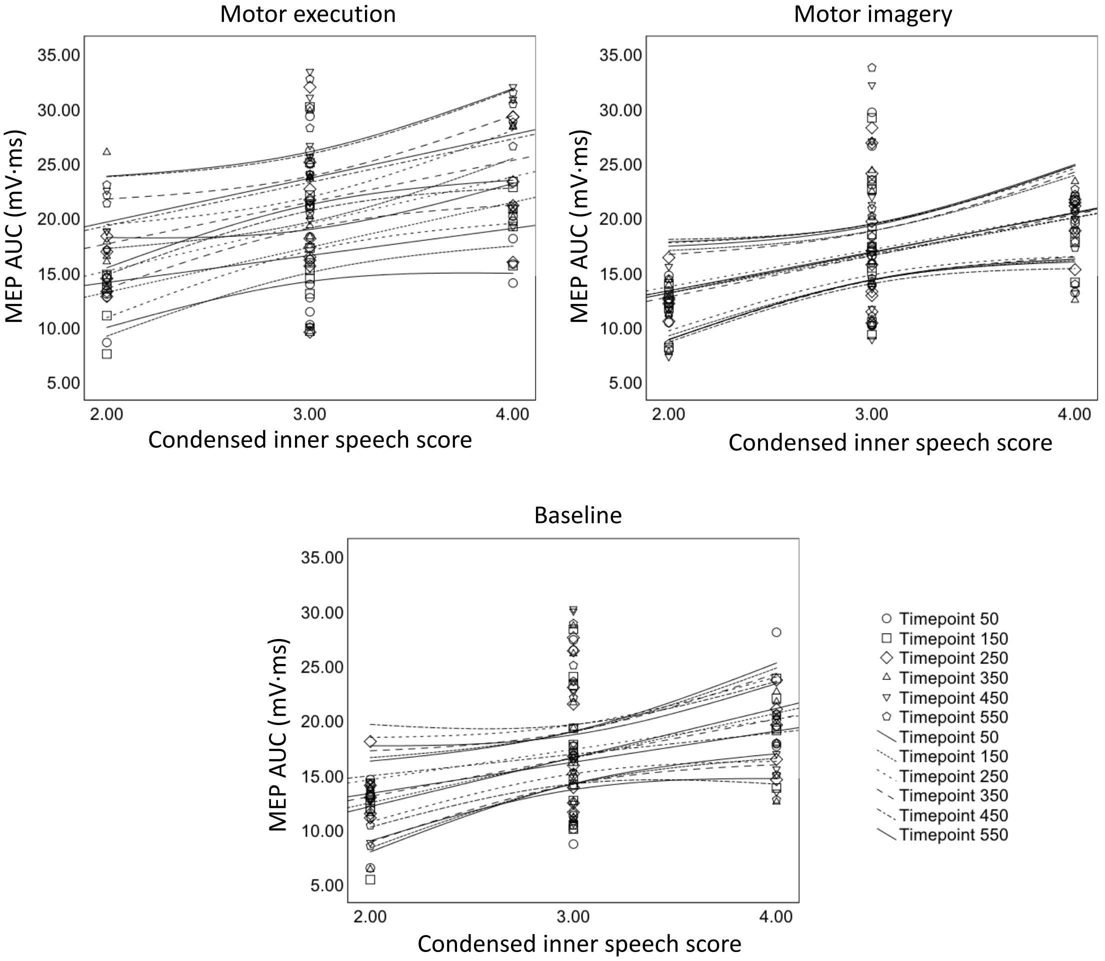


Figure A.10: Correlations of lip MEP AUC and dialogic inner speech score. Lines shown are linear best fit and 95% confidence interval per timepoint.

Experiment 5 and 6

Correlational analyses

No correlations were found between MEP, MOCA or VISQ data points. The only correlations were found between the pre-MEP EMG data and MEP AUC in both timepoints, with correlations that were very similar: at 200 ms $r^2(61)=.624$, $p < .001$, while at 500ms $r^2(61)=.657$, $p < .001$, see Figure A.11.

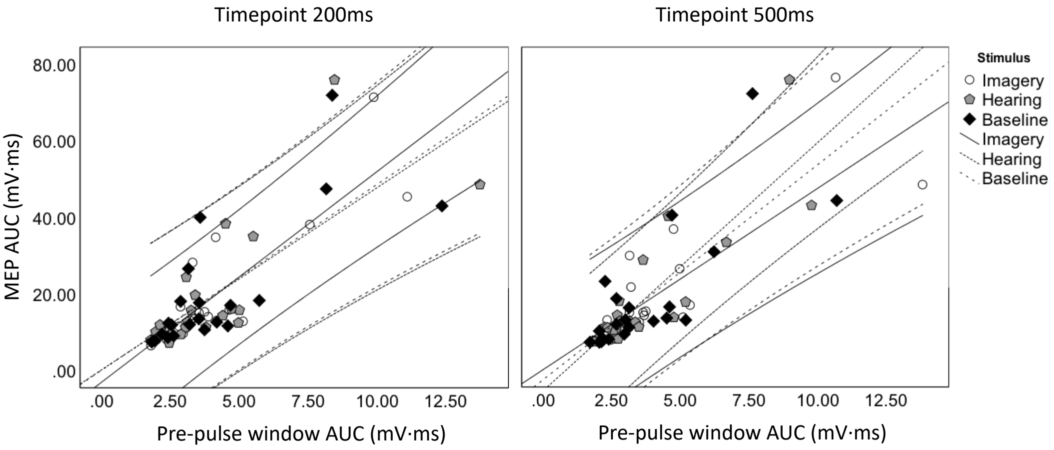


Figure A.11: Correlations between pre-MEP EMG and MEP AUC in experiments 5 and 6 (included participants only). Lines shown are linear best fit and 95% confidence interval per condition.

Appendix B

Figure removed to avoid copyright infringement.

See Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc,53(4),695–699

Figure A.12: Version 7.1A of the the Montreal Cognitive Assessment as used in experiments 1,3-6. See also Nasreddine et al., (2005).

Appendix C

Figure removed to avoid copyright infringement.

See McCarthy-Jones, S., & Fernyhough, C. (2011). The varieties of inner speech: Links between quality of inner speech and psychopathological variables in a sample of young adults. Consciousness and Cognition, 20(4), 1586–1593.

Figure A.13: The Varieties of Inner Speech Questionnaire incorporating all questions from McCarthy-Jones & Fernyhough (2011). Questions were scored from 1 (=Certainly does not apply to me) to 6 (= Certainly does apply to me) and amalgamated per factor (Dialogic inner speech, Condensed inner speech, Other people in inner speech, Evaluative/motivational inner speech)

.

Appendix D

Figure removed to avoid copyright infringement.

See British Society of Audiology. (2012). Recommended procedure. Pure-tone air- conduction and bone-conduction threshold audiometry with and without masking. British Society of Audiology, 1(February), 1–32.

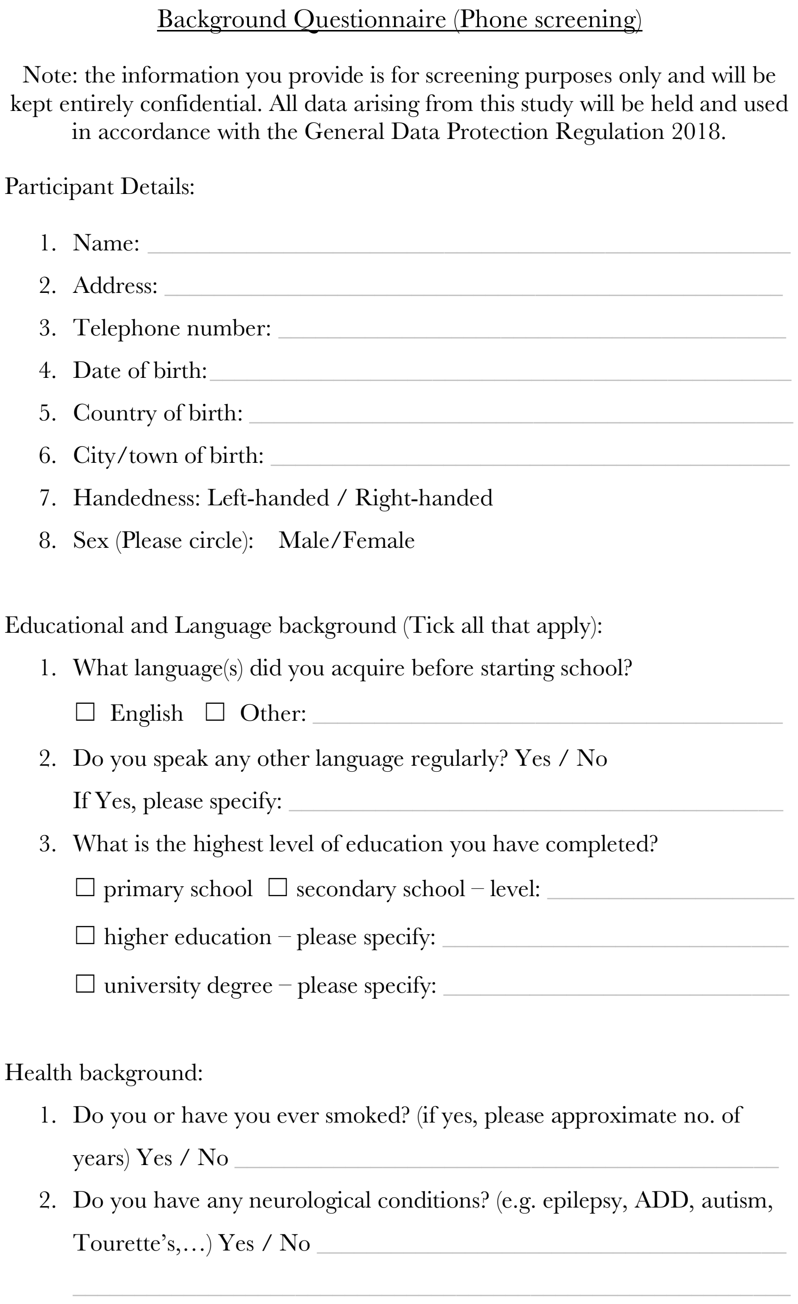
and

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia, 9(1), 97–113.

Figure A.14: Pure-Tone Audiometry (PTA) tests were performed along the guidelines of the British Society of Audiology (British Society of Audiology, 2012) and marked on the PTA grid as below in experiments 1, 3-6. The Handedness test incorporates questions from Oldfield, (1971) and was used in experiments 1, 3-6.. In order to be classified as right- or left-handed, participants answered 8 out of 10 questions with a definitive left or right answer.

Appendix E

Background questions for experiments 1-6.



Background Questionnaire (Phone screening)

Note: the information you provide is for screening purposes only and will be kept entirely confidential. All data arising from this study will be held and used in accordance with the General Data Protection Regulation 2018.

Participant Details:

1. Name: ______________________________
2. Address: ______________________________
3. Telephone number: ______________________________
4. Date of birth: ______________________________
5. Country of birth: ______________________________
6. City/town of birth: ______________________________
7. Handedness: Left-handed / Right-handed
8. Sex (Please circle): Male/Female

Educational and Language background (Tick all that apply):

1. What language(s) did you acquire before starting school?
   ☐ English ☐ Other: ______________________________
2. Do you speak any other language regularly? Yes / No
   If Yes, please specify: ______________________________
3. What is the highest level of education you have completed?
   ☐ primary school ☐ secondary school – level: ____________________
   ☐ higher education – please specify: ____________________
   ☐ university degree – please specify: ____________________

Health background:

1. Do you or have you ever smoked? (if yes, please approximate no. of years) Yes / No ______________________________
2. Do you have any neurological conditions? (e.g. epilepsy, ADD, autism, Tourette's,…) Yes / No ______________________________
   ____

Figure A.15: Background questions for experiments 1-6, page 1.

3. Have you ever had a stroke that you are aware of? Yes / No
   If yes, please provide details: ____________________

4. Have you ever received speech and language therapy? Yes / No
   ____

5. Do you have any hearing problems? Yes / No
   ____

6. Do you have a hearing aid? Yes / No
7. Do you have any problems with your vision? Yes / No
   If Yes please indicate correction: glasses / contacts / other: __________
8. Have you undergone any operations in the past 2 years that involve the head or neck? Yes / No
   If yes, please briefly describe / name the procedure:
   ____
   ____

9. Please list all long- and short-term medication that you use:
- ____________________
- ____________________
- ____________________
- ____________________
- ____________________

10. Do you have any disabilities we should be aware of? Yes / No
    ____
11. Do you play any musical instruments? If so, please list:
    ____

Contact preferences

1. How would you prefer to be contacted by us in the future?
   ☐ Email ☐ Phone ☐Post

Figure A.16: Background questions for experiments 1-6, page 2.

Appendix F

Table A.9: Results from the SCOPUS search (see section 4.2.1.4). Studies using active motor threshold (aMT) are shaded, unlike studies using resting MT (rMT). Ipsi = ipsilateral, Contra = contralateral; coil type ○ (round) or ∞ (figure-of-eight); LH = left hemisphere, RH = right hemisphere

Study	*aMT/ rMT*	*No. of test MEPs*	*Amplitude Criterion*	*Testing Intensity*	*Mean MEP (± SD)*	*Coil type*	*Notes*
(Muellbacher et al., 1994)	aMT	/	/	/	1.8mV ± 1.1mV	80mm (○)	Use tongue depressor/mouthpiece
(Meyer et al., 1997)	rMT	5 responses averaged	/	130%	Ipsi: 1.3mV ± 0.7mV Contra: 1.7,V ± 0.8mV	85mm (∞)	Amplitude criterion: Fig. 2 suggests 0.2mV
(Ghezzi & Baldini, 1998)	aMT	/	/	70% of max output	LH ipsi: 0.5±0.4mV RH ipsi: 0.5mV±0.3mV LH contra: 3.3mV ± 1.1mV RH contra: 1.2mV±0.7mV	(∞)	Data are LH ipsi, LH contra, RH ipsi, RH contra. Tongue was slightly protruded => aMT. Electrodes = clip, 1.5-2cm inter-electrode distance
(Muellbacher et al., 1998)	amT	4 responses averaged	/	/	Ipsi Left: 2.0mV ± 0.84mV Ipsi Right: 2.0mV ± 1.0mV Contra Left: 3.2mV ± 1.49mV Contra Right: 2.8mV ± 1.30mV	70mm (○)	Uses mouthpiece - Data from 40 controls

(Muellbacher et al., 2001)	aMT	5 out of 10	100 μV	90% - 200% rMT	Contra - rMT: 0.1mV Contra - 150%: 1.5mV Contra - 200%: 2.75mV Ipsi - rMT: 0.1mV Ipsi - 150%: 1mV Ipsi - 200%: 2.4mV	45mm (∞)	10% rMT steps - Data approximate from their Fig. 3.
(Fadiga et al., 2002)	rMT(?)	/	/	120%	Reports z-scores only	/	
(Rödel et al., 2003)	aMT			120%	1.4mV ± 0.9mV 2.6mV ± 1.5mV	70mm (∞)	Use tongue depressor and slight MVC - data reported is their rest and moderate activity
Svensson, Romaniello, Arendt-Nielsen, & Sessle (2003)	rMT	5 out of 10	5 μV	120%		50mm (∞)	
(Khedr et al., 2005)	aMT	/	"maximal"	/	Ipsi Left: 1.4mV ± 1.14mV Ipsi Right: 1.27mV ± 1.01mV Contra Left: 1.92mV ± 1.17mV Contra Right: 1.73mV ±1.06mV	90mm (∞)	Mouthpiece - pressing up - 10% MVC

Svensson, Romaniello, Wang, Arendt-Nielsen, & Sessle (2006)	rMT	5 (presumed out of 10)	5 µV	30-100% of stimulator output	40%: 0mV 60%: 0.05mV 80%: 0.175mV 100%: 0.2mV	50mm (∞)	Data are baseline only, not post-training
(Halkjaer et al., 2006)	rMT	5 (presumed out of 10)	5 µV	10% steps of motor threshold, rMT-10% to rMT+60%	rMT: 0.05mV rMT+ 20%: 0.1mV rMT+ 40%: 0.2mV rMT+ 60%: 0.27mV	50mm (∞)	Data show 'Before' conditions only
(S. Boudreau et al., 2007)	rMT	5 out of 10	10 µV	1T,1.2T, 1.5T	40mV, 70mV, 110mV/ms	50mm (∞)	Data reported in Tesla and AUC (approx. from their Fig.4)
(Roy et al., 2008)	rMT(?)	5 out of 10	/	120%	Reports z-scores only	70mm (based on picture) (∞)	
(Gentilucci et al., 2009)	rMT	5 out of 10	50 µV	120%	Large object: 37.8mV Small object: 35.7mV Power grasp: 27.9mV Precision graps:25.6mV	90m(∞)	Electrodes on metal clip device - dorsal surface and ventral aspect, 2cm caudal to tongue apex. 50µV assumed

							("apparatus was the same as in experiment 2").
(Ernberg et al., 2009)	rMT	5 out of 10	5 µV	100% and 180%	Contra 100%: 0.1mV Contra 180%: 0.4mV Ipsi 100%: 0.05mV Ipsi 180%: 0.2mV	70mm (∞)	Data is approximate from their Fig.2, not including genioglossus data.
(Baad-Hansen et al., 2009)	rMT	5 out of 10	5 µV	120%, 150%	Contra 120%: 0.1mV Contra 150%: 0.35mV Ipsi 120%: 0.1mV Ipsi 150%: 0.3mV	(∞)	Data is approximate from their Fig. 2, not including FDI data. Data combined over before and after conditions since no significant difference was found between them.
(Y. Zhang et al., 2010)	rMT	5 out of 10	10 µV	100%, 120%, 180%	100% rMT: 0.05mV 120% rMT: 0.175mV - 0.300mV 180% rMT: 0.425mV - 0.550mV	50mm (∞)	2-3mm from midline, 10mm from tongue tip, inter-electrode distance 10mm. Data approx from graph, baseline (2 conditions) only
(Sato et al., 2010)	rMT (of FDI)	5 out of 10	50 µV	120%	Reports change in size	/	Midline recording (p535)
(Neef et al., 2011)	aMT	3 out of 6	100 µV	90% - 140% of aMT	10% MVC - 100%aMT: 0.25mV 60% MVC - 100%aMT: 0.6mV	70mm (∞)	Silicon mouthpiece - Data show only fluent speakers (Experiment 2)

					10% MVC - 120%aMT: 0.5mV 60% MVC - 120%aMT: 1.0mV 10% MVC - 140%aMT: 0.8mV 60% MVC - 140%aMT: 1.1mV		
(Bocci et al., 2012)	rMT	5 out of 10	50 µV	120%	1.7 ± 0.2mV	90mm (○)	Tongue depressor, inter-electrode distance of 18mm, lateral
(Barwood et al., 2013)	amT	6 out of 10	250 - 450 µV	100%	Report input(VC level) - output (P2P amplitude) linear regression	70mm (∞)	Changes pressure level of tongue 10%-70% of MVC
(M. Kothari et al., 2013)	rMT	5 out of 10	5 µV	80-160%	0.1mV, 0.2mV, 0.3mV	50mm (∞)	Data reported is rMT+20%, rMT+ 40%, rMT + 60% (approx. from their Fig2a, baseline)
Boudreau et al. (2013)	rMT	5 out of 10	10 µV	100%, 140%, 160%	0.2mV, 1.1mV, 1.6mV	50mm (∞)	Data reported is rMT, rMT+ 40%, rMT + 60% (approx. from their Fig. 4)
(Weiss et al., 2013)	rMT	5 out of 10	50 µV	110%	/	(∞)	MEPs only used to reconstruct stimulation mapping
(Vicario et al., 2014)	rMT	5 out of 10	50 µV	120%	Report z-scores only	70mm (∞)	Iron zinc coil

Kothari, Svensson, Nielsen, & Baad-Hansen (2014)	rMT	5 out of 10	5 μV	120%, 140% & 160%	0.05mV, 0.15mV, 0.2mV	50mm (∞)	Amplitude criterion assumed through reference to Svensson et al (2003)
(Komeilipoor et al., 2014)	rMT(FDI?)	5 out of 10	50 μV	120%	Range of 5 only: 0.25 - 1.25mV	(∞)	Electrodes pasted on plstic buttons and fixed on a spring of iron zinc
Neef, Linh Hoang, Neef, Paulus, & Sommer (2015)	aMT	3 out of 6	100 μV	90% - 140% of aMT	100% rMT: 0.4mV 120% rMT: 0.7mV 140% rMT: 1.2mV	70mm (∞)	Silicon mouthpiece - Data shows all speakers including stutterers, contra only
(Komoda et al., 2015)	rMT	5 out of 10	5 μV	5kPA	1.05mV - 1.75mV	∞ (no size)	Data is from 5kPA (2√3 x RMS = 3.5x RMS = 3.5 x 0.3 = 1.05mV) and 10kPA (2√3 x RMS = 3.5x RMS = 3.5 x 0.5 = 1.75mV)
(Busan et al., 2016)	MT / aMT*	Half of stimulations	200 μV	110	110% MT - LH: 0.152mV 110% MT - RH: 0.147mV 125% MT - LH:0.205mV 125% MT: RH: 0.248mV 140% MT - LH: 0.362mV 140% MT - RH: 0.436mV	75mm (∞)	4 electrodes, all dorsal - one left and right 0.5c from midline, close to tip, others 2cm posterior Data shown is for fluent speakers LH and RH MT: defined as MEPS of at least 50 μV in half of stimulation. AMT as half

							over 200μV. MT = 10-20% of MVC, aMT = 60-70% MVC
(Vicario et al., 2016)	rMT	5 out of 10	50 μV	120%	0.182mV ± 0.158mV	70mm (∞)	Midline recording (p355) - Data reported is from baseline trials
(M. Kothari et al., 2016)	rMT	5 out of 10	5 μV	10% steps of motor threshold, rMT-10% to rMT+60%	100% rMT: 0.02mV 120% rMT: 0.1mV 140% rMT: 0.15mV 160% rMT: 0.15mV	50mm (∞)	Inter-electrode distance of 2cm, 10mm from tongue tip, dorsal
Kothari et al. (2017)	rMT	5 out of 10	15 μV	120%, 140% & 160%	0.1mV ± 0.1mV 0.25mV ± 0.3mV 0.35mV ± 0.5mV	50mm (∞)	Data reported is rMT+20%, rMT+ 40%, rMT + 60% (approx. from their Fig. 1b, pre-tDCS)
(Cheng et al., 2017)	rMT	at least 5 MEPs	50 μV	100%	Not reported	70mm (∞)	Testing rMT only - mouthpiece but still relaxed (?)
(Whillier et al., 2018)	aMT	3 out of 6	100 μV	120%	Report ratios only	70mm (∞)	Use tongue depressor and slight MVC
(Panouillères & Möttönen, 2018)	rMT	/	"reliable"	100%	Report z-score only	70mm (∞)	Testing intensity assumed:" intensity of the stimulation was set as the lowest intensity consistently eliciting reliable MEPs in the resting muscle."

Appendix G

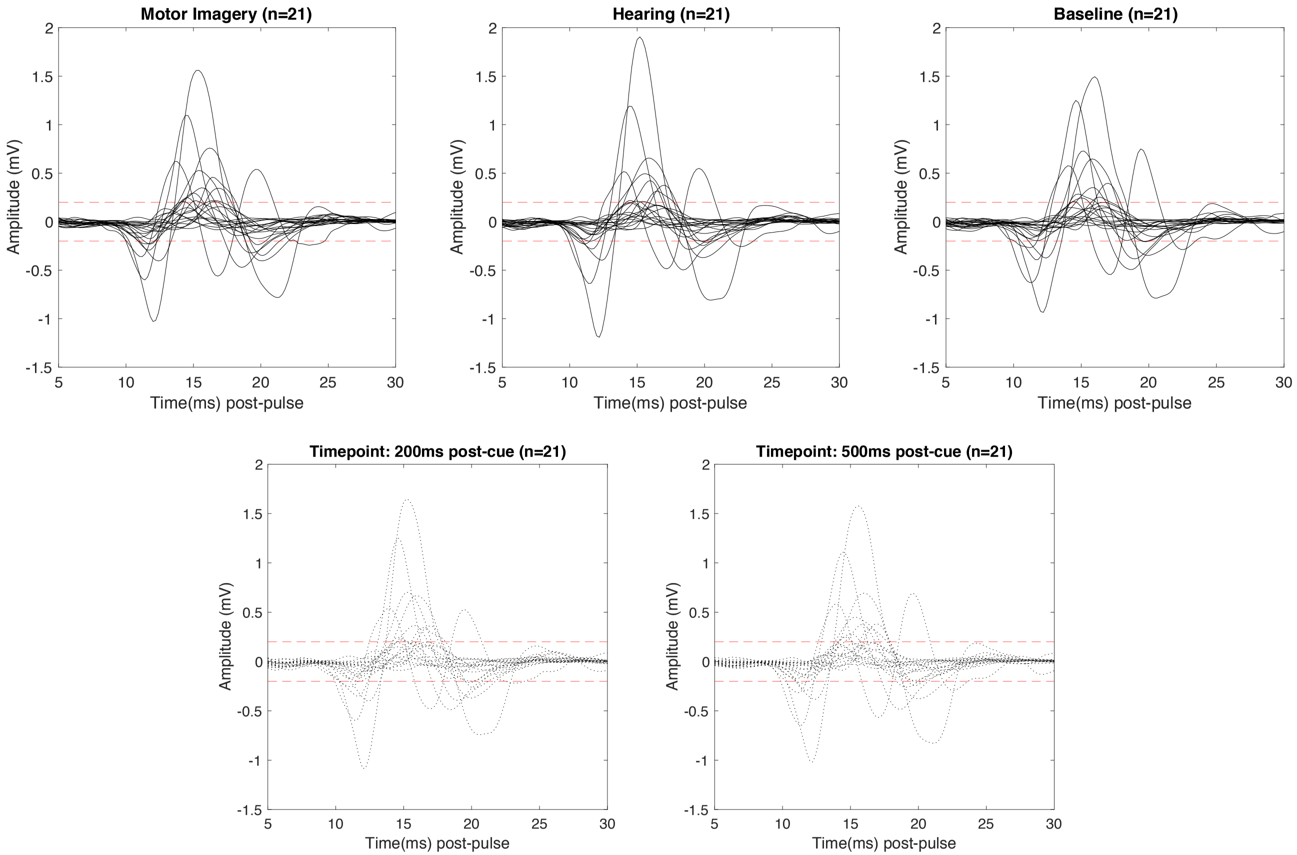


Figure A.17: Examples of average MEP waveforms per participant included in overall analysis (experiment 5 & 6). Note the variability in number of peaks/troughs and latency post-pulse. Such inter-individual differences can also commonly be found in lip MEPs. Note the MEP threshold during aMT was 400 μV, shown as dotted (red) line.