The involvement of articulatory motor cortex in motor imagery of speech

Gwijde Maegherman^{1*}, Helen E. Nuttall², Joseph T. Devlin³, Patti Adank¹

1: Department of Speech, Hearing and Phonetic Sciences Chandler House 2 Wakefield Street, University College London, WC1N 1PF, UK (*: corresponding author)

2: Department of Psychology, Lancaster University, Bailrigg, LA1 4YF, UK

3: Department of Psychology, 26 Bedford Way, University College London, WC1H 0AP, UK

Keywords: Motor imagery; Speech imagery; Motor simulation; Transcranial Magnetic Stimulation; Auditory Verbal Hallucination

Abstract

Motor imagery of speech is thought to involve motor planning and simulated execution of speech actions. However, the precise cortical mechanisms subserving motor imagery are poorly understood. For instance, it is unclear to which extent articulatory motor cortex is involved during motor imagery of speech. We investigated the involvement of motor cortex during articulatory motor imagery through transcranial magnetic stimulation (TMS) to determine its contribution to motor imagery processing. We collected motor-evoked potentials (MEPs) to assess motor cortex excitability in three conditions: speech motor imagery, auditory speech perception, and baseline (no action). MEPs were collected at two time points after trial onset (200ms and 500ms), to measure the relative engagement of tongue primary motor cortex. The results showed that MEPs were larger for the motor imagery than for the baseline condition at 500ms post trial onset only, with no differences between the auditory speech perception and baseline conditions. These results suggest greater facilitation of tongue motor cortex during motor imagery of speech compared to rest, supporting the idea that motor cortex is engaged in motor simulation of complex speech actions.

Highlights

- We investigate primary motor cortex's role in speech motor imagery using TMS
- MEPs show increases 500ms after stimulus onset, suggesting increased processing
- Motor cortex is suggested to be engaged in the process of speech motor imagery
- Motor cortex should be included in cortical models of speech motor imagery

1. Introduction

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 that find 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) 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), 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. 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), 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

76 simulated action), which can be compared with actual auditory and sensory feedback to monitor and correct
77 motor errors in an online and rapid manner.

78
79 In inner speech where no motor output is present, the simulation-estimation mechanism still outputs
80 somatosensory and auditory predictions and perceptual consequences, which results in the inner speech
81 experience. As a result, an internal forward model is a necessary component of both overt and inner speech.
82 Tian, Zarate, and Poeppel (2016) further proposed that in the case of motor imagery of speech, articulatory
83 planning areas produce motor plans that would ordinarily be forwarded to primary motor cortex (M1) are
84 instead sent to parietal regions for simulation and sequential estimation processes, resulting in the inner speech
85 experience. Crucially, this simulation-estimation processing pathway excludes M1. In contrast, Loevenbruck et
86 al., (2018) postulated that even in speech motor imagery, M1 integrates motor programmes computed by
87 auditory and somatosensory regions, but that any action is suppressed via inhibitory commands from anterior
88 prefrontal cortex. The proposed involvement of M1 in speech motor imagery therefore remains unclear. Here,
89 we investigate the involvement of tongue M1 during speech motor imagery, comparing M1 excitability during
90 speech motor imagery, speech hearing and a baseline condition.

91
92 A recent neuro-imaging meta-analysis by Hardwick, Caspers, Eickhoff, and Swinnen (2018) on motor imagery of
93 body movements – excluding facial muscles due to a low number of relevant studies – highlighted the shared
94 neural substrates between action observation, motor imagery and motor execution. This meta-analysis showed
95 overlap between cortical regions involved in action observation, motor imagery, and motor execution, including
96 pre-supplementary motor area (pre-SMA), SMA-proper, dorsal and ventral premotor cortex (PMC), and
97 sensorimotor cortex. It also emphasised regions where there are differences in activation between these
98 conditions. For example, motor imagery was consistently associated with greater activation in premotor regions
99 than motor execution or action observation. In contrast, no evidence of M1 involvement was found in motor
100 imagery or action observation, echoing earlier findings by Héту et al., (2013).

101
102 Both reviews do note that whole-brain functional Magnetic Resonance Imaging (fMRI) may be suboptimal for
103 investigating motor imagery, due to the possibly rapid inhibition of M1 from other brain areas, such as pre-SMA.
104 A meta-analysis by Munzert, Lorey, and Zentgraf (2009) including neurostimulation studies showed a more
105 divided picture of evidence for M1 involvement in general motor imagery during fMRI (Alkadhi et al., 2005;
106 Kuhtz-Buschbeck et al., 2003; Stippich et al., 2002), but also showed stronger support for M1 involvement by
107 including the transcranial magnetic stimulation (TMS) literature (Kumru et al., 2008; Tremblay et al., 2001;
108 Yahagi et al., 1996). However, none of these imagery studies investigate speech specifically.

109
110 While Hardwick et al., (2018) show a lack of M1 involvement for general movement, there is a paucity of speech
111 motor imagery studies focusing on M1 involvement during speech imagery in both neuroimaging and
112 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 lip M1, and left hand M1 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 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. 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 with what was proposed by Tian et al., (2016).

The aim of the current study is first 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 (Maegherman et al., 2019), though MEPs are likely useful markers in tracking changes in M1 excitability during speech imagery processes. We previously used TMS and MEPs to investigate whether excitability varied during imagery processes of simple articulatory gestures using lip muscles (Maegherman et al., 2019). 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. We interpreted this result as likely due to the lack of complexity of the action. Although Roosink & Zijdwind (2010) found increased excitability in hand M1 during imagery of complex hand motor tasks, Maegherman et al. (2019) did not find increased excitability in simple hand motor imagery tasks or simple lip motor imagery tasks. In our current study, we used a complex tongue motor task to investigate whether complexity of the task plays an important role in cortical excitability. We did not want to involve deeper linguistic networks in our tongue actions, however, and so focused primarily on articulatory stages.

A second aim of our current study is 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 (Maegherman et al., 2019). As noted by Ganis (2000), 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 & Poeppel (2010) presented a task with a paradigm similar to our own. In their experiment 4, 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. However, in addition to the spacing information provided through the 1s distance between cues, we needed to include a final visual-only cue that would limit variability of responses and would focus response times further: piloting showed that participants often wait for the TMS pulse and then respond, rather than simply responding as normal - a common occurrence as the pulse is a salient auditory and sensory stimulus in itself. This additional visual-only cue allowed us to instruct participants to use the preceding cues for guidance, but to wait until the final visual-only cue was shown to respond. Our previous study 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 (Tian & Poeppel, 2010) by milliseconds, here we included two timepoints for the TMS pulse: 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 study was 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*. Additionally, we aimed to better understand the potential differences in tongue motor cortex excitability during motor imagery of speech and hearing of speech.

2. Materials and methods

2.1 Participants

39 participants (29F, 10M, mean age = 23y 3m, SD = 3y 2m age range = 19-33) took part in the study. Group A consisted of 20 participants (19F, 1M, mean age = 22y 6m, SD = 2y 4m, age range = 19-27) while Group B consisted of 19 participants (10F 9M, mean age = 24y 1m, SD = 3y 10m, age range = 19-33). Handedness was established using the Edinburgh Handedness Inventory (Oldfield, 1971): 35 participants were right-handed and 4 were left-handed (2 in each group). 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). All research was carried out in accordance with the Declaration of Helsinki.

2.2 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 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 study.

2.3 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 graded the performance of the participant on a Likert scale from 1 (very easily performs /tr/) to 5 (performs /tr/ with great difficulty).

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 1 cm 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

Figure 1: detailed example of single trial showing baseline and MEP window relative to EMG trace and prompts.

decrease of $\pm 100 \mu\text{V}$ above or below the pre-established threshold of 20% of maximum voluntary contraction). 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 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 (i.e., imagine saying /tr/, hearing /tr/, or do nothing, all the while maintaining 20% of maximum voluntary 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 study was made up of 12 trials, showing each condition four times, interleaved.

The main study 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 participant. 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 main study had finished, we ran a battery of tests including the Montreal Cognitive Assessment (Nasreddine et al., 2005) to ascertain any 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 (tested at 0.25, 0.50, 1,2,4 and 8kHz in accordance with the British Society of Audiology (2012)) to ensure participants had hearing within the normal range. The TMS experiment had a duration of 40min and the total testing session ran for about two hours.

Transcranial magnetic stimulation

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; 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 100 μ V (Rossini et al., 2015) or 200 μ V (Groppa et al., 2012) should be used when determining aMT, although certain recent studies increase this even further, using a criterion as high as 1000 μ V for hand muscles (Möttönen et al., 2010, rMT) or up to 250-450 μ V for tongue muscles (Barwood et al., 2013, aMT). However, it should also be noted that 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 this study. 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 the Supplementary Materials.

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 400 μ V for tongue MEPs.

The literature search also showed a variety of testing thresholds used: most studies, whether using rMT or aMT, used testing thresholds 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 with two groups, with two different testing intensities: Group A proceeded to use a testing intensity of 100% aMT, while Group B 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 (wing 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.

Analysis

A 1040ms sweep was extracted from Spike² (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 4.1). 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 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 2: Flow chart of data analysis preconditions. 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 and slopes for individual participants. This LME approach differs from a more traditional MEP analysis, in this case a 3x2x2 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. Data used in this study is publicly available via [dataset] github.com/Gwiide/2019_Submission (Dataset cited as Maegherman, 2020).

3. Results

Frameless stereotaxy tracking allowed us to extract in-session localisation data for 36 out of 39 participants (3 sets of tracking data were lost due to tracking hardware issues). Average aMT was for Group A was 53.6% (range 41-73) of stimulator output, which was also the testing output for this group. For Group B, average aMT was

49.4% (range 39-62) of stimulator output, with average testing output 59.2% (range 47-74%) of stimulator output.

To investigate whether we could collapse the two studies' datasets to improve our sample size, we performed simple *t*-tests on the MEP AUC data between groups A (100%) and B (120%). An overall *t*-test showed that there was no difference between the two groups when all data were collapsed over condition as well as timepoint: group A ($M=.7978 \mu V \cdot ms$, $SD = .485 \mu V \cdot ms$) vs group B ($M=1.172 \mu V \cdot ms$, $SD = 1.026 \mu V \cdot ms$); $t(17) = -.1330$, $p = .201$. To ensure there were no effects even within the conditions, we also ran an independent samples *t*-tests using per-condition data. These data show that there was no difference in any of the conditions, across groups (see Supplementary Table 2). Participants from both groups were therefore combined from this point forward and analyses conducted on this pooled dataset.

3. 1 Pre-MEP window

A first check analysis involved running a linear mixed-effects analysis on the pre-MEP data to ensure that pre-pulse EMG activity was the same across conditions. 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 slopes and 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 slopes and intercepts for participants yielded an improved model fit for the pre-MEP EMG data ($\chi^2(2) = 54.614$, $p < .001$), 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 ($t(36) = 5.863$, $p < .001$) and between the motor imagery and baseline conditions ($t(36.534) = 6.849$, $p < .001$), but not between the hearing and baseline conditions ($t(36.534) = 0.980$, $p = .327$). Participants provided more contraction prior to the motor imagery condition relative to the other conditions, but that there was no difference in contraction force prior to the hearing and baseline conditions.

To account for the differences in muscle contraction across condition, we ran one-way ANOVA 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. We excluded participants who showed a pattern of significantly different amounts of pressure on the electrodes between conditions prior to the TMS pulse from further analysis. Eight participants in group A and seven participants in group B showed significant differences in their pre-MEP EMG across conditions. These 15 participants were excluded from further analysis, with the mixed design now including 21 participants. We reran the linear mixed-effects analysis on the pre-MEP window EMG data as before, on the reduced participant set. The addition of the fixed factor condition to a baseline model with random slopes and intercepts for participants did not yield an improved model fit for the pre-MEP EMG data ($\chi^2(2) = 3.3707$, $p = .185$), showing that there was no longer a difference between the conditions in terms of pre-MEP EMG data.

Figure 3: 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 show ± 1 SEM.

In other words, any changes in MEP cannot be attributed to a different amount of force on the electrodes between conditions.

3.2 Linear mixed-effects analysis

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

Random slopes and intercepts were included per participant and fixed effects included condition (motor imagery, hearing, baseline), time point (200, 500ms) and group (Group A and B). The addition of group to the model including random effects did not improve model fit ($\chi^2(1) = 1.227$, $p = .268$) indicating again that performance did not differ between the two groups – we therefore abolished group as a factor henceforth. The addition of condition to the model including random effects did not improve model fit ($\chi^2(2) = 1.061$, $p = .588$). The addition of time point to the model including random effects did not improve model fit ($\chi^2(1) = 2.873$, $p = .090$). The addition of condition and timepoint to the model including random effects 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 to the model did significantly improve model fit ($\chi^2(2) = 7.001$, $p = .030$). Fixed effects follow-up show that there was a significant difference between motor imagery and baseline at 500ms ($t(20) =$

Figure 4: 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.

2.645, $p=.008$) but not between motor imagery and hearing ($t(20) = 1.305$, $p=.1920$) or hearing and baseline ($t(20) = 1.337$, $p=.181$, see Figures 3 and 4).

The LME with the same model procedure was also carried out on the pre-MEP EMG data, which showed no effects (see Supplementary table 3a). This analysis was also carried out on the excluded group of participants and their pre-MEP EMG data (see Supplementary table 3b), 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 we know are due to pre-existing differences in the pre-MEP EMG data. 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.

4. Discussion

4.1 General discussion

This study aimed to investigate cortical excitability changes in the tongue motor cortex during motor imagery of speech, hearing of speech, and a baseline condition. We predicted that cortical excitability, as measured through MEPs, would be greater in the motor imagery condition than in a baseline condition. Moreover, we predicted that cortical excitability in a hearing condition would be greater than that in a baseline condition. Neither prediction was met outright. Our 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), we did not find an increase in MEPs for the hearing condition relative to the baseline condition at either time point. We discuss the implications of our results below.

Our results show an increase in cortical excitability at the later timepoint during imagining production of a complex consonant cluster (see Figure 3). Models of speech production such as Indefrey & 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 these experiments and our experiment are the wait time (i.e. the condition was known well before a response was required) and the non-semantic nature of our task. In our 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, & 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 an MEG study. Since in our case the condition and task were already known to the participants, we can assume a time period roughly between 100-300ms, not accounting for any additional reaction time, in which motor activation might have occurred. Our 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 our paradigm and that in experiment 4 in Tian & 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 & Poeppel (2010) suggest a mean reaction time of 1047ms from the final auditory cue (their table 2, p. 11), or around 150ms prior to our 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), we have no way of knowing whether participants performed the imagined execution for the same duration when performing motor imagery of speech. Our 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 & Levelt (2000), MEP differences should be observable during at least one of our time points if motor cortex is involved in motor imagery.

Our 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 coupled with the length and robustness of our 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 & Poeppel (2010). /da/ 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 our expectations, we did not find an increase in the hearing condition at either timepoint. This null effect might be due to the fact that our 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 which were confirmed by Murakami, Restle, & 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, & 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 our study 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, with no significant difference in MEP size found. 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.

Another possibility is that attentional demands set this study 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, our study 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 we 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, 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, our 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). We attempted to circumvent this issue by specifically eliciting a motor strategy in our instructions (“Say the sound in your head”), but we cannot be certain 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. Our results at the later time point suggest that, while we cannot be certain of every participant’s strategy, a sufficient number of participants may have used the egocentric articulatory imagery strategy over the allocentric strategy.

4.2 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 μ V, and during subsequent

mental execution [...] it increased still further, not infrequently being accompanied by ‘bursts’ or ‘volleys’ of impulses which reached 50 μ V or more in intensity.” It is this difficulty in achieving complete relaxation of the tongue that made us decide 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.

Another potential issue with the design is that we presented the upcoming condition in each trial to the participant well before they had to perform what the condition requires. 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 any issues from Go - NoGo strategies, in which immediate decision-making was thought to potentially cause increased muscle movement, thereby skewing the results. However, the pattern in our 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.

Finally, the combining of participants with different stimulus intensities should be addressed. Our initial analysis showed that there were no significant differences between groups A and B, which is unexpected: higher stimulation intensities should lead to larger MEPs. It may be the case that our use of a large MEP criterion (400 μ V) had the effect of flattening the response curve of our MEPs to some degree. The use of a large MEP criterion is not in itself an issue as long as the criterion remains consistent across participants, but we did not employ input-output curves in this study to investigate the effect of the amplitude criterion per se. Further studies using this design could use a smaller MEP criterion, e.g. 200 μ V, to investigate if these results are maintained.

4.3 Implications

This study provides 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. Løevenbruck 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 do not 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. 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), our results suggest a role for articulatory M1 in inner speech and to some degree, executive processing. Our 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. Tian & Poeppel (2012) predict there to be sequential nature to simulation estimation, whereas Løevenbruck 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 we would expect smaller MEPs after rTMS, however if they are involved in the inhibition of overt responses then we would predict expect larger MEPs as M1 would be disinhibited.

5. Conclusions

This study investigated cortical excitability changes in the tongue motor cortex during motor imagery of speech, hearing of speech and a baseline condition. We found that the difference between conditions was dependent on the time point at which we collected MEPs: at 500ms post-cue, we found increased cortical excitability during motor imagery of speech relative to the baseline condition. These results 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.

Acknowledgments

This study was funded by the BIAL foundation (Fundação BIAL grant no. 281/16) and further by an ESRC studentship grant to the corresponding author (grant no. ES/P000592/1). The authors declare no conflict of interest.

Author contributions

Gwijde Maegherman: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Software; Validation; Visualization; Writing - original draft; Writing - editing.

Helen E. Nuttall: Conceptualization; Funding acquisition; Methodology; Supervision; Validation; Writing - review & editing.

Joseph T. Devlin: Conceptualization; Methodology; Supervision; Writing - review & editing.

Patti Adank: Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Validation; Writing - review & editing.

Bibliography

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>

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>

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>

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

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>

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>

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>
- 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>
- British Society of Audiology. (2011). *Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking*. Reading: British Society of Audiology.
- 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>
- 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>
- 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>
- 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](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>
- 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>
- 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>
- 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](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., Baciú, 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>

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>

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>

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>

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>

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

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>
- 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>
- 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>
- 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>
- Løevenbruck, H., Grandchamp, R., Rapin, L., Dohen, M., Perrier, P., Baciú, M., Loevenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., & Dohen, M. (2018). *A cognitive neuroscience view of inner language : to predict and to hear , see , feel*.
- Maegherman, G. (2020). *Dataset via Github*. https://github.com/Gwijde/2019_submission
- Maegherman, G., Nuttall, H. E., Devlin, J. T., & Adank, P. (2019). Motor imagery of speech: The involvement of primary motor cortex in manual and articulatory motor imagery. *Frontiers in Human Neuroscience*, 13(June), 1–10. <https://doi.org/10.3389/fnhum.2019.00195>
- 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>
- 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](https://doi.org/10.1016/S0924-980X(96)96598-4)
- 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>
- 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>
- 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., 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>
- 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>
- 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>
- 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](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>
- 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>
- 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.
- Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciú, 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>

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

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>

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.

Roosink, M., & Zijdwind, 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>

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. *And Clinical*, 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>

Rothwell, J. C., Hallett, M., Berardelli, A., Eisen, A., Rossini, P. M., & Paulus, W. (1999). Magnetic stimulation: motor evoked potentials. *Electroencephalogr Clin Neurophysiol Suppl*, 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>

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>

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>

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>

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

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](https://doi.org/10.1016/S0304-3940(02)00001-5)

Svensson, P., Romaniello, A., Arendt-Nielsen, 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>

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(OCT), 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., 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., & Rumati, 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>

- 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>
- 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>
- 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](https://doi.org/10.1016/S0028-3932(02)00316-0)
- 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>
- 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](https://doi.org/10.1016/S0278-2626(03)00033-2)
- 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>