Title: Motor Imagery in Speech Processing
Project number: 281/16
Researchers: Patti Adank, Gwijde Maegherman, Helen Nuttall
Institution: University College London UK, University of Lancaster UK
Project duration: 01-01-2018 - 01-04-2020
Keywords: Imagery, Cognitive processes, Movement, Somatosensory system

Abstract
Activity in motor cortex (M1) increases when we imagine actions. Speech imagery occurs when we silently prepare what we want to say without speaking. However, past research focused on neural bases of imagery of gestures, and research on speech imagery is limited. Theories on the neural organisation of speech attribute a pivotal role to covert motor simulation processes such as imagery. Our work aimed to elucidate the behavioural and neurophysiological substrates supporting motor imagery of speech. We combined Transcranial Magnetic Stimulation (TMS) with behavioural tasks. When TMS is applied to left M1, the relative excitability of the motor system and associated lip and tongue muscles can be measured using Motor Evoked Potentials (MEPs) in lips or tongue. We recorded MEPs from lip (study 1a) and tongue (study 1b) muscles when participants imagined articulating simple (study 1a) or complex (study 1b) speech actions. In study 1b participants also listened to speech to test effects of action observation. Study 2 used TMS in its causal mode by testing how activation in motor cortex is affected after a TMS temporarily disrupted activity in two key brain areas: Dorsolateral Prefrontal Cortex (DLPFC), pre-Supplementary Motor Area (pre-SMA) during imagery and observation. An additional functional Magnetic Resonance Imaging (fMRI) experiment aimed to determine the networks responsible for motor imagery of hand and lip movement. Results showed that lip M1 was not facilitated during motor imagery of simple speech actions and tongue M1 was facilitated during the late phase of motor imagery of complex speech articulations. The results from study 2 and the fMRI experiment are not yet available, as testing was interrupted due to the COVID-19 crisis. The results from this project inform theories on speech production and perception by clarifying the role of motor cortex in speech imagery.

1. Introduction
In the course of the project we made a change from the original intended project. We originally planned to only run two planned TMS studies (1 and 2), but we ended up running one additional TMS study, two additional behavioural studies, plus a functional magnetic resonance imaging (fMRI) study. The aims of these studies were as follows:

1. Behavioural study 1a: aimed to determine (a) the replicability of a classic motor imagery study (the hand laterality task by Parsons (1994)) in our lab, and (b) whether such a motor imagery task can be extended to the articulatory motor domain using the lip muscles and auditory stimuli.
2. Behavioural study 1b: aimed to further test the tasks from Behavioural study 1a online using a browser-based version of the task. This made use of the Gorilla Experiment Builder (www.gorilla.sc).
3. TMS Study 1a: aimed to determine whether (a) it is possible to detect imagery processes in lip motor cortex, (b) whether we can distinguish between imagery processes, action execution processes and a baseline condition using simple lip actions

in terms of lip motor cortex excitability, and (c) whether we can determine the time course of the above processes.

4. TMS Study 1b: aimed to further develop the task from Study 1a and extend findings from the lip motor cortex to the tongue motor cortex.
5. TMS study 2: aimed to investigate the changes in tongue motor cortex excitability as found in Study 1b when modulating the state of other brain areas, such as dorsolateral prefrontal cortex and premotor areas, using inhibitory TMS.
6. fMRI study: aimed to determine the networks responsible for (a) motor imagery of hand movement and (b) motor imagery of lip movement, leading to a further understanding of the shared network of motor imagery. This study makes use of the tasks previously replicated/established in behavioural studies 1a and 1b.

3. Method

Participants

Behavioural studies: Participants in the behavioural studies were all healthy speakers of British English, 18-30 years old. 48 participants completed behavioural study 1, 40 participants completed behavioural study 1b.

TMS studies: Participants in the TMS studies were healthy right-handed speakers of British English, 18-30 years old. 20 participants completed TMS study 1a, 39 participants completed TMS study 1b and 11 participants completed TMS study 2.

fMRI study: Participants in the fMRI study were right-handed speakers of British English, 18-30 years old, without developmental/mental/speech/language/hearing disorders or any contra-indications for fMRI. 17 participants completed the fMRI study thus far.

Materials

Behavioural study 1a: The study made use of 16 visual stimuli showing a hand (adapted from [1]). The initial stimulus was created using the Blender 3D computer graphics program, which was rotated in 45° increments along the z-axis, as well as mirrored to provide left and right hand stimuli. The study also used 16 auditory stimuli. Twenty-five tokens of each target vowel for the AIT were recorded in a sound-proof anechoic chamber, by a native 24-year old female SSBE speaker. Selection and subsequent modification of the final vowel stimuli was performed using Praat [2]. 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. Two vowels were moderately pitch-corrected to ensure all vowels had the same pitch. The vowel stimuli are embedded in a sonorant consonantal frame, e.g., /sazə/. Contexts and vowels were concatenated with an overlap of 15ms. In-house Praat scripts were used to create KlattGrid-synthesised vowel stimuli with a pass Hann filter (8000Hz) applied. Participants used left and right foot switches to respond in each trial. Visual stimuli were presented on a 21.5” computer monitor, while auditory stimuli were presented on a pair of Sennheiser HD-25SP headphones.

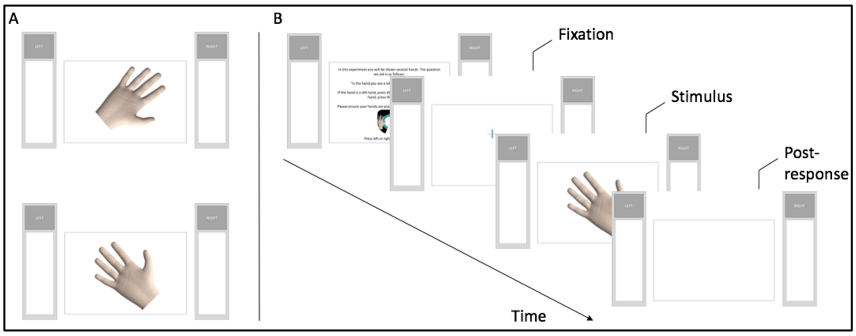

Figure 1. A. Examples of trials from the hand laterality task. B. The progression of events in the hand laterality task.

Behavioural study 1b: The study used a subset of stimuli from Study 1a (left and right hand stimuli) as before but only at rotations of 0°, 90°, 180° and 270°. The study used a modified subset of auditory stimuli collected during Study 1a, namely the natural stimuli rather than synthesised stimuli, using native or near-native vowels (/i/,/ɔ/, /ʉ/,/e/), or non-native vowels (/ɤ/, /o/, /ɯ/, /ø/). This study was run online using the Gorilla Experiment builder at www.gorilla.sc and so stimuli were presented on screens of different sizes and at a volume comfortable for the participant. Participants responded using the left and right arrow keys.

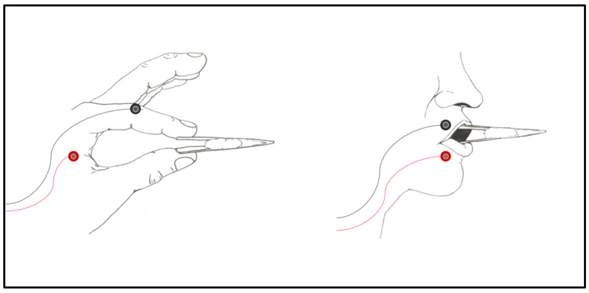

Figure 2. Illustration of the use of the tweezer tool in the hand (left) and lip conditions (right), including the electrode montage (black and red circles) in TMS experiment 1a.

TMS study 1a: The study used visual stimulus prompts displayed on a 21.5” computer monitor (Figure 1). The prompts were a combination of symbols (font size 24) denoting conditions and muscle used: “%%” represented use of hand muscles and “&&” represented use of lip muscles. Colour was used to indicate whether the action should be executed (red), imagined (blue) or whether no action should be taken (black). The study used a tweezer-like tool to effect constant muscle contraction as well as to provide a tool for the motor execution condition in both the lip and hand muscles (Figure 2). Electromyographical data was collected through electrodes (Ag/AgCL, diameter 10mm) on the first dorsal interosseous hand muscle (tendon-belly montage), and on the orbicularis oris lip muscle (belly-belly montage). Transcranial magnetic stimulation was delivered using a Magstim BiStim unit. Testing threshold was set to 120% of active motor threshold (aMT).

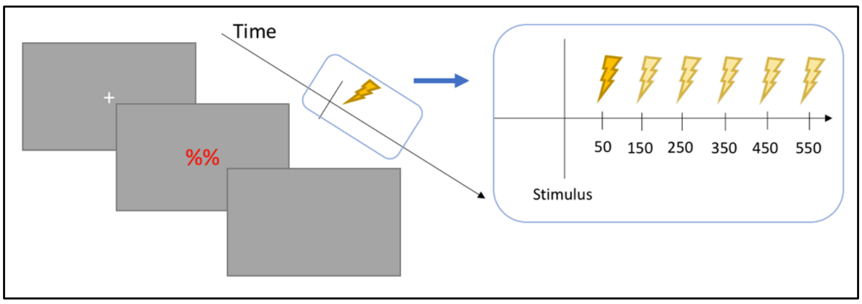

Figure 3. Timing of TMS pulses in TMS study 1a.

TMS study 1b: Prompts were displayed on a 21.5" computer monitor while participants sat ~70cm away from the screen. Prompts used included graphical 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 4). 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. ER1 Etymotic earphones were used to deliver auditory stimuli. Electromyographical data was collected through two electrodes attached to a tongue depressor to form a mouthpiece against which participants pressed their tongue. The ground electrode was placed on the right temple. The signal was sampled as in Study 1a. Frameless stereotaxy and transcranial magnetic stimulation were effected using the same setup. Frameless stereotaxy was used to localize the area of stimulation for each effector. Testing threshold varied by group: one group received stimulation at 100% of active motor threshold, while the other group received stimulation at 120% of active motor threshold.

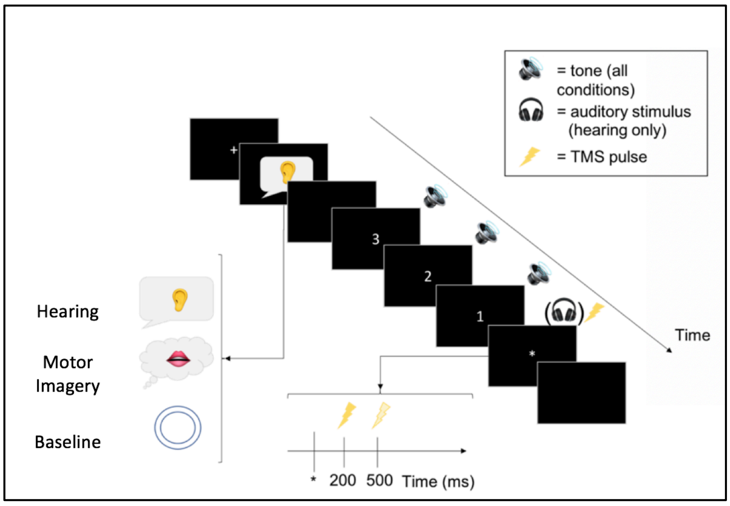

Figure 4. Illustration of the sequence of events in a single trial in TMS experiment 1b.

TMS Study 2: Materials and setup were as in Study 1b, with the addition of repetitive TMS and use of participant-specific structural scans (acquired during the fMRI study). Repetitive TMS was administered to three brain areas (DLPFC, pre-SMA, and parieto-occipital sulcus) using a Magstim Rapid sending pulses at a rate of 1Hz for 15m, at 120% of aMT, as defined per participant.

FMRI study: The study used the same visual and auditory stimuli as Study 1b. Participants were scanned in Siemens Prisma 3T scanner with stimuli displayed on an in-bore screen, equipped with gradient-echo echo-planar imaging (multiband acceleration factor = 4 (interleaved), TE = 35.20ms, TR = 1500ms). ER1 Etymotic earphones were used to deliver auditory stimuli, while a custom-made button box was used to collect responses.

Procedure

Common procedures across experiments include behavioural consent forms and screening forms (prior to task), background screening and background information sheets (prior to task), audiometry test (prior to or following task), Montreal Cognitive Assessment [3], Variety of Inner Speech Questionnaire [4], payment/crediting of participant (following task or upon discontinuation of study). For TMS studies only, the following additional procedure applied: TMS consent forms and screening forms (prior to task, TMS Studies 1) and determination of active motor threshold. Active motor threshold is determined as follows: the relevant area was localised using the hot-spot technique: starting from previously-established coordinates and moving the coil by 5-10mm to ascertain the precise location at which the most robust MEPs were elicited. A figure-of-eight coil (wing diameter 70mm) was placed at a 45° angle relative to the sagittal plane. Thresholding was performed using a standard descending-ascending thresholding procedure in which 5 out of 10 MEPs must be elicited. Once active motor threshold was established, testing intensity was set to a relevant testing threshold (usually 120% of active motor threshold) [5].

Behavioural study 1a: Participants positioned their hands on the edge of a table in front of them, palms down, with the left foot on the left foot switch and the right foot on the right foot switch. The experimenter explained that stimuli would be shown on-screen and that a foot press was required in each trial. The participant was presented with a practice task and was told that there were four tasks: two requiring judgements based on the stimulus shown (hand laterality task and auditory form task) and two requiring them to press the foot pedal as soon as the perceived the stimulus. The visual task (Figure 1) and auditory task were explained through on-screen instructions. Oral instructions were limited to examples of spread and rounded vowels, and to answer any questions. Participants were told to answer as quickly and as accurately as possible, but were not told how to make judgements (that is, a strategy was not implied). Participants were monitored via a webcam. Each stimulus was presented until a foot switch was pressed. 192 stimuli were presented per task, in eight blocks of 48 stimuli with a short break in between blocks (the participant pressed either footswitch to continue).

Behavioural study 1b: The procedure was similar to that in Study 1a, with the main difference the use of the Gorilla Experiment builder to run the study in a browser session rather than laboratory conditions. The participants received instructions to use left and right arrow buttons to respond during the tasks. Participants were told they needed to wear headphones to take part, and that the experiment needed to be performed on a laptop or desktop computer. The instructions were identical to those in Study 1a, except the reaction time task was removed. The participant was presented with a practice task during which they were told if their responses were correct. The task itself followed in 3 blocks of 16 stimuli (total of 48 stimuli, each stimulus shown six times).

TMS study 1a: 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 task showed each trial type four times allowing the experimenter to make comments on the actions performed. Training lasted 2 minutes and all were able to successfully perform the actions. Each trial began by displaying a white fixation cross. The prompt was shown for 2000ms, with TMS pulses administered at 50ms, 150ms, 250ms, 350ms, 450ms, or 550ms post-stimulus (counterbalanced). Blocks consisted of 25 trials, with breaks in between each block. Each break was a minimum of 1m, 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 h (45m of TMS).

TMS study 1b: 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/). 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. The experimenter gave advice on how to pronounce the sound when requested. Participants were given a maximum of 15m to train. Next, the tongue electrodes 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, and exerted pressure on the electrodes so that they could comfortably hold tension in the tongue muscle. 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. Each trial began by displaying a white fixation cross. The condition prompt was then shown for 1000ms. Next, participants were shown a countdown visually displaying the numbers 3,2,1 for 1000ms consecutively, accompanied by a tone. After another 1000ms, a white asterisk was shown, which was the 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 (Figure 4). 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 4 times, interleaved. The main study involved eight blocks of 30 trials for a total of 240 trials. The entire session lasted 2h, with the TMS experiment lasting 40m.

TMS Study 2: Procedure for TMS Study 2 was identical to TMS Study 1b, with some alterations. Participants were made aware that the task would be repeated four times, with repetitive TMS applied to a specific brain area (counterbalanced across participants) prior to the task. After the practice task, participants were told they would be watching a documentary (without visual language, auditory stream and showing no anthropoid movement) for 15 minutes prior to the task. During this time repetitive TMS was administered to a specific brain region (dorsolateral prefrontal cortex, pre-SMA, or parieto-occipital sulcus). One additional condition involved performing the task without the preceding repetitive TMS. Each condition lasted a total of 30m (15m for the documentary + repetitive TMS, 15 mins for the TMS), yielding a TMS time of 1h 45m. The study lasted a total of 3h 30m.

FMRI study: Participant were asked to read the information sheet and fill in the fMRI consent and safety screening forms. Participants were shown the practice task outside the scanner. Once in the scanner, they were asked if they were comfortable and were made familiar with the safety features of the scanning environment. The study used the same tasks and stimuli as in behavioural study 1b, with the addition of a simple reaction time task using the same stimuli (adapted from study 1a, reducing the number of trials for effective use of the fMRI paradigm). Trials and conditions were counterbalanced across and within participants. Each participant underwent four experimental runs (1 per task), with 32 trials per task coming to a total of 128 trials per participant. This was followed by a field map scan and structural scan for use in subsequent analysis. Once these scans were complete the participant was removed from the scanner and thanked.

3. Results and Discussion

Behavioural study 1a: The study was successful in replicating the visual task: we found significantly longer reaction times for those orientations which were lateral (i.e. orientations which required more arm and hand movement) than for those which were medial (i.e. orientations which required less arm and hand movement), as predicted. We found that there was no difference between left and right hand stimuli. A similar pattern was observed for the novel auditory task: non-native vowels (i.e. vowels which require on-line motor planning for production) had significantly longer reaction times than native vowels (i.e. those vowels which have readily-accessible motor plans). However, we also found that there was a difference in reaction time between round and spread stimuli regardless of their nativeness, with rounded vowels showing shorter reaction times than spread vowels. We also found that there were differences in accuracy along similar lines: we found significantly greater accuracy for medial orientations and lower accuracy for those orientations which were lateral, as predicted. There was no difference in accuracy between left and right hand stimuli. A similar trend was found for the auditory task: native vowels were judged with greater accuracy than non-native vowels. However, we also found that there was a difference in accuracy between rounded and spread stimuli, regardless of nativeness with rounded vowels showing better accuracy judgements than spread vowels.

Behavioural study 1b: The study found the same results for the visual task as found in study 1a, further replicating the findings with a more limited set of stimuli and outside a laboratory environment. In the auditory task we once again found a difference between native (shorter reaction time) and non-native (longer reaction time), but we did not find a difference in reaction time between round and spread stimuli. We did find an interaction effect, due to little difference in reaction time in the rounded condition but a much larger difference in reaction time in the spread condition. Accuracy data for the visual task showed a similar pattern as previously found in study 1a, though we found an interaction between the two factors as the right hand showed a larger difference in accuracy scores between medial and lateral orientations than did the left hand. The auditory task showed that regardless of whether the stimulus was rounded or spread, accuracy was similar, but that there was a difference between native vowels, which were judged more accurately, than non-native vowels, which were judged less accurately.

TMS study 1a: The results showed that primary motor cortex was facilitated (i.e., showed a higher value for Area Under the Curve (AUC), Figure 5.) during motor execution for both effectors, but we could not find evidence supporting the prediction that mental imagery involves M1, and we therefore did not replicate earlier similar studies. The pattern of increasing motor-evoked potentials in the results for lip and hand during action execution followed a comparable time course, but differences in area-under-the-curve were more pronounced for hand muscles. The effect of action execution in both lip and hand muscles showed that we successfully captured M1 facilitation during execution of actions, with the time course showing the expected increase between 150 and 350ms, plateauing thereafter. The results showed no evidence of cortical facilitation for the imagery condition compared to the baseline condition for either effector. The results from this study were published in [6].

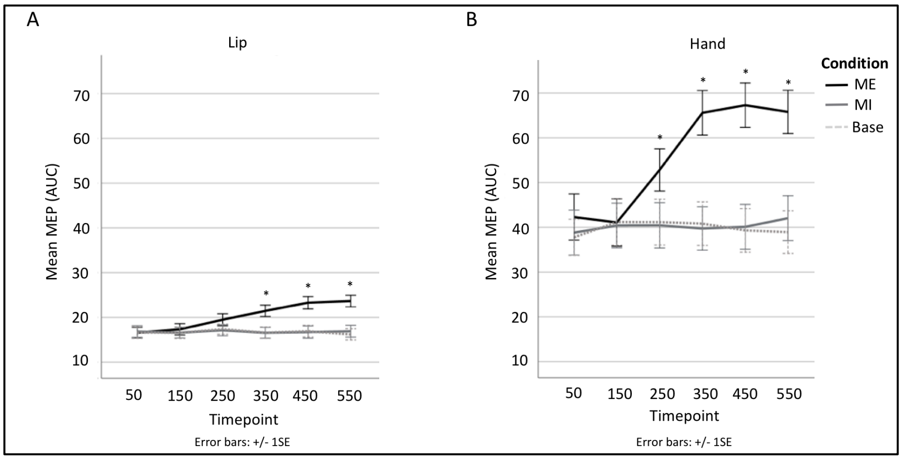

Figure 5. Mean Area Under the Curve (MEP amplitude) for lip/hand muscles for Motor Imagery (MI), Motor Execution (ME), and Baseline (Base).

TMS study 1b: 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 (AUCs, Figure 6). In addition, in contrast to previous studies reporting increased motor cortex engagement during listening to speech we did not find an increase in MEPs for the hearing condition relative to the baseline condition at either time point. Overall, we found evidence for motor cortex involvement during imagery of speech, but not during hearing. This last point is surprising as some previous studies have reported finding motor cortex involvement during hearing, however these studies used a variety of different stimuli and paradigms and cannot be directly compared to our own. The results of this study are currently being prepared to journal submission [7].

TMS study 2: Due to the shutdown of laboratory spaces in response to governmental guidelines regarding COVID-19, we have not at this point collected a sufficient number of participants for analysis.

fMRI study: Due to the shutdown of laboratory spaces in response to governmental guidelines regarding COVID-19, we have not at this point collected a sufficient number of participants for analysis.

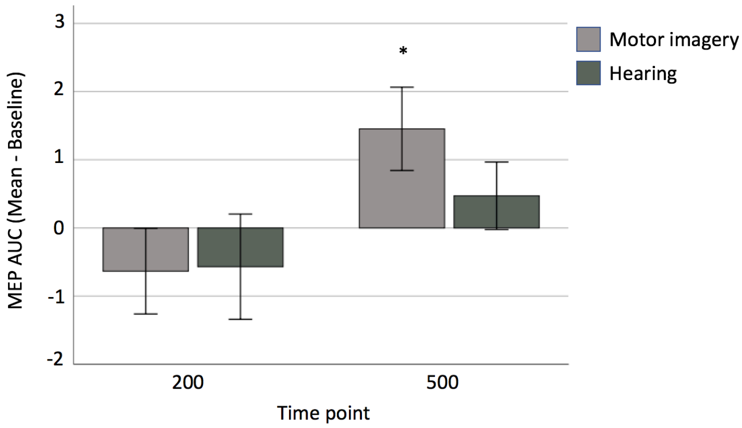

Figure 6. Mean Area Under the Curve (MEP amplitude) for lip/hand muscles for Motor Imagery and Hearing (action observation).

4. Conclusions and Recommendations

Our work aimed to elucidate the behavioural and neurophysiological substrates supporting motor imagery of speech. We conducted three behavioural experiments, two TMS experiments, and one fMRI experiment. The results from the behavioural experiments demonstrated slower performance for the task conditions that were designed to be more difficult to imagine performing. Therefore we succeeded in replicating published results showing analogies between executed and imagined manual actions [1] for speech stimuli. The results from the two TMS experiments showed that M1 only engages during M1 when the speech action to be imagined is complex enough, which suggests a role in M1 in covert simulation processes during complex articulatory processes.

In conclusion, the work contributed to theory development in the field of speech imagery and on the neurobiology of speech perception and speech production, especially with respect to the neural substrates supporting covert simulation processing [8, 9]. The results also augment theories on the underlying mechanisms in auditory verbal hallucinations [10, 11] In terms of publications, the results from TMS experiment 1a were published last year in Frontiers in Human Neuroscience, and we are presently in the last stages before submitting the results from TMS experiment 1b to a journal for cognitive neuroscience. We will inform the BIAL Foundation about our progress with this paper. In case we will be able to complete testing on TMS study 2 and the fMRI study (not included in the initial grant proposal), we will keep the BIAL foundation in the loop and will acknowledge in case this work is published. Overall this grant was very beneficial for our research and has furthered our way of thinking about imagery of speech substantially.
