UNIVERSITÀ DEGLI STUDI DI MILANO

CORSO DI DOTTORATO ed eventuale curriculum

The human mind and its explanations: language, brain and reasoning

DIPARTIMENTO DI AFFERENZA DEL CORSO

Dipartimento di Filosofia Piero Martinetti

TESI DI DOTTORATO DI RICERCA

Exploring Covert Motor Processes: Behavioral Insights and Innovative Applications of Transcranial Magnetic Stimulation

sigla del settore/i scientifico disciplinare/i
PSIC-01/B

NOME DEL DOTTORANDO Francesca Genovese

Matr. R13676

ORCID n. 0009-0000-4956-0461

NOME E COGNOME DEL TUTOR Corrado Sinigaglia, Guido Barchiesi

NOME E COGNOME DEL COORDINATORE DEL DOTTORATO Andrea Sereni

A.A.

(2024-2025)

Table of contents

Introduction.....4
REFERENCES.....7

Chapter 1: Multidimensional Motor Evoked Potentials (MultiMEP): Digging up buried information from single trials.....9
1. INTRODUCTION.....9
2. METHODS.....13
2.1 Participants.....13
2.2 Procedure.....13
2.2.1 Electrodes Grid Montage.....13
2.2.2 TMS.....15
2.2.3 Stimuli, Task, and Trial Structure.....16
3. ANALYSES.....19
3.1 MultiMEP classification.....19
3.1.1 Data processing.....19
3.1.2 Classification.....20
3.2 Exploratory Analysis: Cross-classification between contraction and imagery pattern.....21
3.2.1 MultiEMG Data Processing.....22
3.2.2 Sanity Check: MultiEMG Classification.....22
3.2.3 Cross-classification between contraction and imagery patterns: Data Pre-processing.....22
3.2.4 Cross-classification between contraction and imagery patterns: Classification.....22
3.3 Exploratory Analysis: Cross-Subjects MultiMEP classification.....23
3.4 Exploratory Analysis: MultiMEP Classification Accuracy for Electrode Grid Subdivisions.....24
4. RESULTS.....25
4.1 MultiMEP classification.....25
4.2 Exploratory Analysis: Cross-classification between contraction and imagery patterns.....26
4.2.1 Sanity Check: MultiEMG Classification.....26

4.2.2 MultiEMG-to-MultiMEP Classification.....27
4.2.3 MultiMEP-to-MultiEMG Classification.....28
4.3 Exploratory Analysis: Cross-Subjects MultiMEP classification.....29
4.4 Exploratory Analysis: MultiMEP Classification Accuracies for Electrodes Grid Subdivisions.....31
5. DISCUSSION.....33
6. CONCLUSIONS.....37
Supplementary Materials.....38
REFERENCES.....49

Chapter 2: Unlocking motor preparation with Multidimensional Motor Evoked Potentials (MultiMEP)
1. INTRODUCTION.....57
2. METHODS.....58
2.1 Participants.....58
2.2 Procedure.....59
2.2.1 Electrodes Grid Montage.....59
2.2.2 TMS.....*61*
2.3 Stimuli, task and trial structure.....*61*
2.4 Training session.....63
3. ANALYSES.....64
3.1 MultiMEP Data processing.....64
3.2 MultiIMEP classification.....65
3.3 Exploratory analysis: MultiEMG data processing.....66
3.3.1 MultiEMG classification.....66
3.3.2 Cross-Classification between contraction and preparatory patterns.....66
4. RESULTS.....66
4.1 MultiMEP classification.....66
4.2 Exploratory analysis: MultiEMG classification.....67
4.2.1 MultiMEP-to- MultiEMG.....68
4.2.2 MultiEMG-to-MultiMEP.....69
5. DISCUSSION.....70
6. CONCLUSIONS.....72
REFERENCES.....73

Chapter 3: Does experience modulate automatic imitation? A new look.....77
1. INTRODUCTION.....77
2. MATERIALS AND METHODS.....80
2.1 Sensitivity Analysis.....80
2.2 Participants.....80
2.3 Stimuli.....81
2.4 Response Detection System.....81
2.4.1 Flex-sensor.....81
2.4.1 Foam Rubber Manipulandum.....82
2.5 General Experimental Structure.....82
2.5.1 Automatic Imitation Time Course (Pre-training and Post-training Sessions).....84
2.5.1.1 Task and Trial Structure.....84
2.5.1.2 Familiarization Session.....86
2.6.2 Counter-imitative Training and Imitative Training Sessions.....88
2.6 Data Analysis.....88
2.6.1 Analysis: Pre and Post-training Sessions.....88
2.6.2 Analysis: Counter and Imitative Training Sessions.....89
3. RESULTS.....90
3.1 Results: Pre and Post-training Sessions.....90
3.2 Results: Counter and Imitative Training Sessions.....96
3.3 Exploratory analysis: The 0 ms Delay.....97
4. DISCUSSION.....100
REFERENCES.....104

Conclusions and future directions.....110

Introduction

As Marc Jeannerod wrote twenty-five years ago:

> "The possibility to experimentally access cognitive or mental states characterized by absence of overt behavior represents a new avenue for neuroscience. In the field of human motor cognition, it is only recently that it was realized that actions involve a covert stage. Covert and overt stages thus represent a continuum, such that every overtly executed action implies the existence of a covert stage, whereas a covert action does not necessarily turn out into an overt action" (Jeannerod, 2001: S103).

Since then, a vast body of research employing diverse methodologies has investigated the motor processes and representations underlying covert action. Prominent examples include motor imagery (e.g., Grosprêtre, Ruffino, & Lebon, 2016), motor preparation (e.g., Duque & Ivry, 2009), and action observation (Hannah et al., 2018) .

Building on Jeannerod's insight, the study of covert motor processes can no longer be regarded as a "new avenue," but has matured into a well-established and productive research field. Nonetheless, key challenges persist in identifying and measuring the processes and representations that support covert action. By developing and applying new methodological approaches, this thesis seeks to advance our ability to address these challenges.

A standard method to investigate whether—and how—motor processes are recruited during covert action is to assess their influence at the motoneuron level by measuring corticospinal excitability. Typically, this is done by applying transcranial magnetic stimulation (TMS) to the motor cortex and recording the resulting motor-evoked potentials (MEPs) from one or two target muscles. This 1 muscle × 1 action (1m1a) approach assumes a direct mapping between a muscle's activity and a specific action. For example, increased MEP amplitude in the first dorsal interosseus (FDI) during imagined or observed index-finger abduction is interpreted as facilitation of the corresponding motor representation (e.g. Fadiga et al., 1998). While this approach has yielded important insights, it has three key limitations. First, it reduces complex motor representations to isolated joint displacements. Second, it neglects

intermuscular relationships. Third, it relies on averaging across trials, thereby discarding informative single-trial variability.

To overcome these shortcomings, this thesis develops the MultiMEP approach. Instead of focusing on a single muscle, MultiMEP applies TMS to the motor cortex and records MEPs simultaneously from a large set of electrodes, followed by decoding methods inspired by multivoxel pattern analysis (MVPA) in neuroimaging. By treating the *pattern* of MEP amplitudes across multiple muscles as the unit of analysis, this method captures intermuscular relationships and leverages trial-to-trial variability, revealing the multidimensional richness of corticospinal output. Moving beyond the constraints of the 1m1a framework, MultiMEP enables the study of more ecologically valid actions involving multiple effectors and complex synergies, shifting the focus from isolated muscle responses to distributed motor codes.

The first chapter of this thesis presents an experimental study applying the MultiMEP approach to motor imagery (MI), one of the most investigated forms of covert action. MI refers to the mental simulation of movement without overt motor output and can be distinguished into visual and kinesthetic forms (Jeannerod, 1994; Chong & Stinear, 2017). In this study, participants imagined complex hand actions while TMS was applied to the motor cortex and MEPs were simultaneously recorded from 24 electrodes. Single-trial decoding of these signals achieved classification accuracies above 70%, demonstrating that MEPs convey much richer information about the motor processes underlying covert action than previously assumed.

The second chapter extends the MultiMEP approach to motor preparation, another extensively investigated form of covert action. In this study, MEPs were recorded from 32 electrodes placed on the forearm and hand while participants prepared to perform five different actions. Decoding analysis revealed that each action could be classified during the preparation phase with accuracies above 40% (chance level: 20%). These findings demonstrate that MultiMEP is effective not only for motor imagery but also for preparatory motor processes. Taken together, the first two studies provide proof-of-principle for MultiMEP as a powerful tool to investigate motor representations underlying classical forms of covert action, with potential applications to other covert domains such as action observation.

The third chapter introduces a different methodological approach to the study of covert motor processes: examining the effect of visuomotor training on action observation, which at the behavioral level is reflected in automatic imitation. Automatic imitation is defined as a stimulus–response compatibility effect whereby observing another's action influences the observer's motor output (Stürmer et al., 2000). Typically, this effect is measured behaviorally through reaction times: participants execute an action more quickly when it resembles the observed action. The robustness of this phenomenon has been well documented across different paradigms, with converging evidence that action observation elicits a corresponding motor representation, thereby facilitating the execution of similar actions.

Recent work has shown that this effect can be modulated by experience, using counter-imitative training paradigms. In such paradigms, participants are instructed to perform the opposite action relative to the one observed for a brief training period, after which the automatic imitation effect is reassessed (Heyes et al., 2005; Catmur et al., 2007; Cavallo et al., 2014). It is well established that counter-imitative training influences the automatic imitation effect. According to one of the best-known theoretical accounts, the Associative Sequence Learning (ASL) model, this modulation is explained as the result of sensorimotor contingencies (Heyes et al., 2005). In the study presented here, we examined the detailed time course of this effect to better understand its underlying mechanisms. Our findings demonstrate that the automatic imitation effect cannot be fully explained by mere sensorimotor contingencies, at least not in the way this account has traditionally been interpreted.

In summary, this thesis combines methodological innovation with theoretical investigation to advance our understanding of covert motor processes. The MultiMEP approach provides a powerful tool to decode corticospinal activity with greater ecological validity, while the behavioral study of automatic imitation reveals the flexibility of motor representations shaped by experience. Together, these studies deepen our knowledge of the neural and behavioral mechanisms underlying covert action and lay the groundwork for future research into motor cognition and its applications.

REFERENCES

Catmur, C., Walsh, V., & Heyes, C. (2007). Sensorimotor Learning Configures the Human Mirror System. Current Biology, 17(17), 1527–1531. https://doi.org/10.1016/j.cub.2007.08.006

Cavallo, A., Heyes, C., Becchio, C., Bird, G., & Catmur, C. (2014). Timecourse of mirror and counter-mirror effects measured with transcranial magnetic stimulation. Social Cognitive and Affective Neuroscience, 9(8), 1082–1088. https://doi.org/10.1093/scan/nst08

Chong, B. W., & Stinear, C. M. (2017). Modulation of motor cortex inhibition during motor imagery. Journal of neurophysiology, 117(4), 1776–1784. https://doi.org/10.1152/jn.00549.2016

Duque, J., & Ivry, R. B. (2009). Role of corticospinal suppression during motor preparation. Cerebral Cortex, 19, 2013–2024.

Fadiga, L. et al. (1998). Corticospinal excitability is specifically modulated by motor imagery a magnetic stimulation study. 37, 147–158.

Grosprêtre, S., Ruffino, C., & Lebon, F. (2016). Motor imagery and cortico-spinal excitability: A review. European journal of sport science, 16(3), 317–324. https://doi.org/10.1080/17461391.2015.1024756

Hannah, R., Cavanagh, S. E., Tremblay, S., Simeoni, S. & Rothwell, J. C. (2018). Selective Suppression of Local Interneuron Circuits in Human Motor Cortex Contributes to Movement Preparation. J. Neurosci. 38, 1264–1276.

Heyes, C., Bird, G., Johnson, H., & Haggard, P. (2005). Experience modulates automatic imitation. Cognitive Brain Research, 22(2), 233–240. https://doi.org/10.1016/j.cogbrainres.2004.09.009

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. Behavioral and Brain Sciences, 17(2), 187–202. https://doi.org/10.1017/S0140525X00034026

Jeannerod M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. NeuroImage, 14(1 Pt 2), S103–S109. https://doi.org/10.1006/nimg.2001.0832

Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence Effects With Manual Gestures and Postures: A Study of Imitation. Journal of Experimental

Psychology: Human Perception and Performance, 26(6), 1746-1759. https://doi.org/10.1037/0096-1523.26.6.1746

Chapter 1

Multidimensional Motor Evoked Potentials (MultiMEP): Digging up buried information from single trials[1]

1. INTRODUCTION

Research on covert motor processes such as motor imagery (Fadiga et al., 1998; Grosprêtre et al., 2016; Keogh, Bergmann & Pearson, 2020; Meers et al., 2020), action observation (Ubaldi et al., 2015; Fadiga et al., 1995; Cattaneo et al., 2013; Catmur et al., 2007; Craighero et al., 2007; Naish et al., 2014), motor preparation (Derosiere et al., 2020, Duque et al., 2010; Duque et al., 2014; Duque & Ivry 2009; Ficarella & Battelli 2019; Hannah et al., 2018; Tandonnet et al., 2011), inhibition (Aron & Verbruggen, 2008; Claffey et al., 2010; Majid et al., 2012), or other similar covert activities (Barchiesi et al., 2022; Messina et al., 2016), often employs suprathreshold transcranial magnetic stimulation (TMS) of the motor cortex to elicit motor evoked potentials (MEPs) from muscles associated with the processed actions (Bestmann & Krakauer, 2015).

For example, in a widely employed experimental design, researchers examine motor representations underlying covert processing of index and little finger abduction, which depend on the isolated first dorsal interosseous (FDI) and abductor digiti minimi (ADM) contraction, respectively. The design follows a 2 (actions) × 2 (muscles) structure, with MEP amplitudes as the dependent variable. Amplitudes are averaged within each condition, and an interaction between factors is analyzed. FDI MEPs increase when observing or imagining index finger abduction compared to little finger abduction which allows to infer that the former motor representation is facilitated (and vice versa for ADM) (Ubaldi et al., 2015; Catmur et al., 2007;

---

[1]This chapter is based on the article: ***Genovese, F**.,*** Mussini, E**., Zazio, A., Beltrami, F., Bortoletto, M., Cattaneo, L., Rota, P., Negro, F., Fanghella, M., Sinigaglia, C., & Barchiesi, G. (2025). Multidimensional Motor Evoked Potentials (MultiMEP): Digging up buried information from single trials. Brain stimulation, S1935-861X(25)00320-1. Advance online publication. https://doi.org/10.1016/j.brs.2025.09.001 **These authors equally contributed to the research reported in the present work.

Cattaneo et al., 2009; Cavallo et al., 2014; Fourkas et al., 2006; Naish & Obhi, 2015; Urgesi et al., 2006; Urgesi et al., 2010).

Most of these studies share methodological features that simplify testing but at the cost of limiting the information extracted from MEPs, which reduces the possibility of a thorough understanding of how motor content is represented during covert processes.

They implicitly adopt a "one-muscle one-action" (1m1a) approach assuming that each studied action depends solely on the isolated contraction of one of the recorded muscles. This limits the range of testable actions, as identifying additional one-to-one muscle-action pairs becomes increasingly challenging, and it reduces the ecological validity of the results as the studied actions are often single-joint movements rarely occurring in isolation in real-life[2]. Additionally, 1m1a studies average MEP amplitudes within design cells, improving the signal-to-noise ratio but discarding valuable single-trial information. Lastly, the adopted univariate analyses treat the recorded muscles as independent levels of a factor, disregarding inter-muscle relationships and thus losing valuable information. As a result, much of the potential insights from single-trial MEP data remains substantially unexplored.

Here, we introduce Multidimensional Motor Evoked Potentials (MultiMEP) to address these limitations. Inspired by Multivoxel Pattern Analysis (MVPA) (Haxby et al., 2014; Hebart & Baker, 2018; Jimura & Poldrack, 2012; Weaverdyck et al., 2020), which classifies condition-specific information from voxels activity patterns, in the MultiMEP decoding framework we aim at discriminating covertly processed actions based on the multidimensional pattern of MEPs recorded from different muscles, rather than treating them as independent units of analysis. Just as MVPA enhances information discrimination in neuroimaging studies, we expected a similar advantage in MultiMEP, given that different locations on an effector are differentially influenced by distinct upstream corticospinal input.

In a proof-of-concept experiment, we recorded MEPs from 24 forearm electrodes while participants imagined three complex hand actions that, if produced, would involve the contraction of multiple muscles. We then used supervised learning classification to train an algorithm to distinguish between imagined actions based on

[2] This problem has been mitigated by a minority of studies using inertial sensors (Barchiesi & Cattaneo 2013; Classen et al., 1998; McMillan et al., 2006; Ruffino et al., 2019) or ultrasound tissue Doppler imaging (D'Ausilio et al., 2014) to investigate more naturalistic motor representations.

MultiMEP patterns. The classifier estimates hyperplanes in a multidimensional space to separate activity patterns associated with each action. Hyperplanes decoding generalizability was tested on new trials, with classification accuracy reflecting their ability to decode imagined actions from MultiMEP amplitude patterns (Figure 1).

This decoding approach overcomes the information limitations imposed by the 1m1a strategy by enabling the study of complex actions, leveraging activity patterns across multiple electrodes, and fully exploiting relational information between MEPs. Once trained, the classifier model categorizes imagined actions based on single-subject and single-trial MultiMEP data.

In an exploratory analysis, we demonstrated that MultiMEP can support inferences similar to 1m1a experiments within a multidimensional decoding framework. We tested a classic claim deriving from early 1m1a studies supporting the idea that motor imagery and action production share motor representations (Fadiga et al., 1998; Rossi et al., 1998; Stinear et al., 2006). To test this we also extracted contraction patterns and trained two classifiers in two directions: one has been trained on the contraction domain and tested on the MultiMEPs evoked during imagery, while the other has been trained on MultiMEPs and tested on contraction patterns. If motor representations are shared, cross-classification pattern discriminability should exceed chance, mirroring classical 1m1a findings. Eventually, we conducted a cross-subjects classification analysis to explore the similarity between decoding spaces across participants, and an electrode selection analysis, to explore the contribution of different arm areas to the classification outcome.

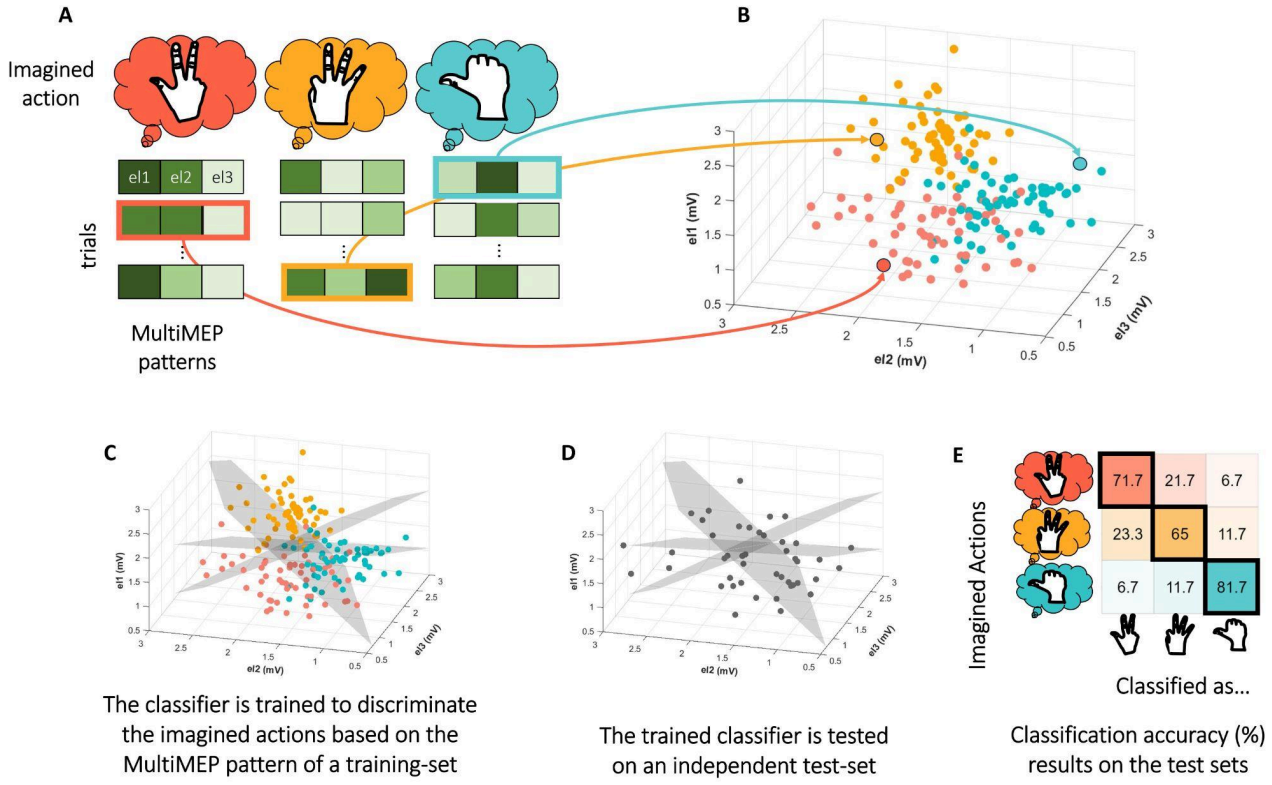

Figure 1. MultiMEP decoding rationale

A. *MultiMEP decoding approach: White-to-green arrays represent a vector of the MEPs amplitudes recorded from each recorded electrode (el1, el2, el3, for visualization purposes, only 3 electrodes are represented) during the imagery of complex actions (i.e., actions that, if produced, would involve the contraction of several muscles).*

B. *A 3D representation of an ideal MultiMEP-space where each dot represents a single trial MultiMEP pattern; the dot's position in the MultiMEP-space is determined by the amplitude of MEPs recorded on each electrode. The dot color represents the action imagined in a specific trial.*

C. *Using a supervised learning approach, the classifier is trained to estimate hyperplanes that best distinguish between imagined actions based on MultiMEP patterns from a so-called "training-set" of trials.*

D. *The estimated hyperplanes are employed to test their classification accuracy on an independent set of trials called test-set.*

E. *The estimated hyperplanes are employed to test their classification accuracy on an independent set of trials called test-set.*

2. METHODS

2.1 Participants

Twenty-five right-handed (Xygonakis et al., 2018) participants (15 females, mean age: 22.6 years), took part in the experimental session. Ad-hoc exclusion criteria corresponded to the safety criteria for TMS (Rossi et al., 2009; Rossi et al., 2021). For this purpose, all potential participants were screened for neurological and psychiatric conditions using a TMS exclusion criteria questionnaire prior to participation. Of the 25 individuals initially recruited, 22 were included in the final analysis: two participants were excluded due to difficulty relaxing the target muscles, and one withdrew due to subjective discomfort. No major adverse effects were reported by any participant. The study was approved by the ethical committee of the University of Milan (“MultiMEP: Potenziali Evocati Motori Multidimensionali”) and conducted in accordance with the revised Helsinki Declaration (World Medical Association, 2013), with written informed consent obtained from all participants.
We based our sample estimation around a classical motor imagery task employed in Rossi et al. 1998 (Rossi et al., 1998). There, we calculated the effect size of their outcomes resulting in a Cohen's d = 1.43. We then calculated, using the same effect-size, the sample size for a paired-sample and a one-sample Wilcoxon-test (two-tails), with alpha = 0.05 and power = 0.8. We obtained a sample of 5 and 7 participants. In order to increase the generalizability of the results we arbitrarily incremented the participants number to 25.

2.2 Procedure

The entire procedure consisted of the electrode grid montage, hotspot-intensity hunting procedure, instructions, practice, and finally, the experimental block. The whole session lasted approximately 2-2.5 hours.

2.2.1 Electrodes Grid Montage

Experimenters began electrode placement by scrubbing the skin of the right forearm with scrubbing paste (Merletti et al., 2020). Electromyographic (EMG) recordings were obtained using 24 passive sintered Ag/AgCl ring electrodes (model: B10,

length: 150 cm, EasyCap GmbH, Germany) arranged in four strips of six electrodes each, spaced 2.5 cm apart. On the dorsal forearm, one strip was aligned along the radial side, with the gap between the fourth and fifth electrodes positioned midway down the forearm, and a second strip was placed medially with a 1.25 cm posterior shift; a similar setup was used on the volar surface, with the third strip aligned along the ulnar side and the fourth anteriorly shifted (Figure 2). The advantage of a grid of evenly distributed electrodes covering a substantial portion of the forearm consists in reducing potential a-priori electrode placement problem (Barchiesi et al., 2020). Disposable ground and reference electrodes were positioned on the proximal volar or dorsal forearm, and the electrode grid was secured with adhesive tape. Conductive gel was applied inside each ring via syringe to achieve conductance levels below 40 kΩ.

Monopolar recordings of EMG were achieved by connecting the electrodes to 24 pins on a BePlusPro Advanced amplifier (EBNeuro, Italy), with data sampled at 4096 Hz. A 10 Hz high-pass filter was applied solely for visualization purposes, not affecting the recorded data.

To increase the spatial specificity, we computed, for visualization purposes, bipolar derivations of the electrodes by taking the difference between pairs of monopolar recordings (Farina et al., 2004; Guerrero et al., 2017) along the longitudinal dimension of the forearm for each of the four strips (1-2 to 5-6, 7-8 to 11-12, 13-14 to 17-18, 19-20 to 23-24) and transversely from the dorsal surface (1-7, 6-12, 13-19,18-24), for a total of 24 bipolar derivations.

To check the signal during the experimental session, we employed Lab Streaming Layer functions (https://labstreaminglayer.org/#/ (Kothe et al., 2025 (Preprint)), extracting the recorded signal stream from the amplifier and producing our own customized “MultiMEP Viewer” displayed on the amplifier-PC (MATLAB2021b, MathWorks). The Viewer was composed of two windows: the “continuous stream window,” where the continuous native monopolar data streams were plotted, and the “multi-epochs window”, where MEPs calculated on the 24 bipolar derivations, were displayed between -150 to 400 ms from the TMS trigger, and whose peak-to-peak amplitude was automatically calculated between 15 ms to 65 ms from the TMS trigger.

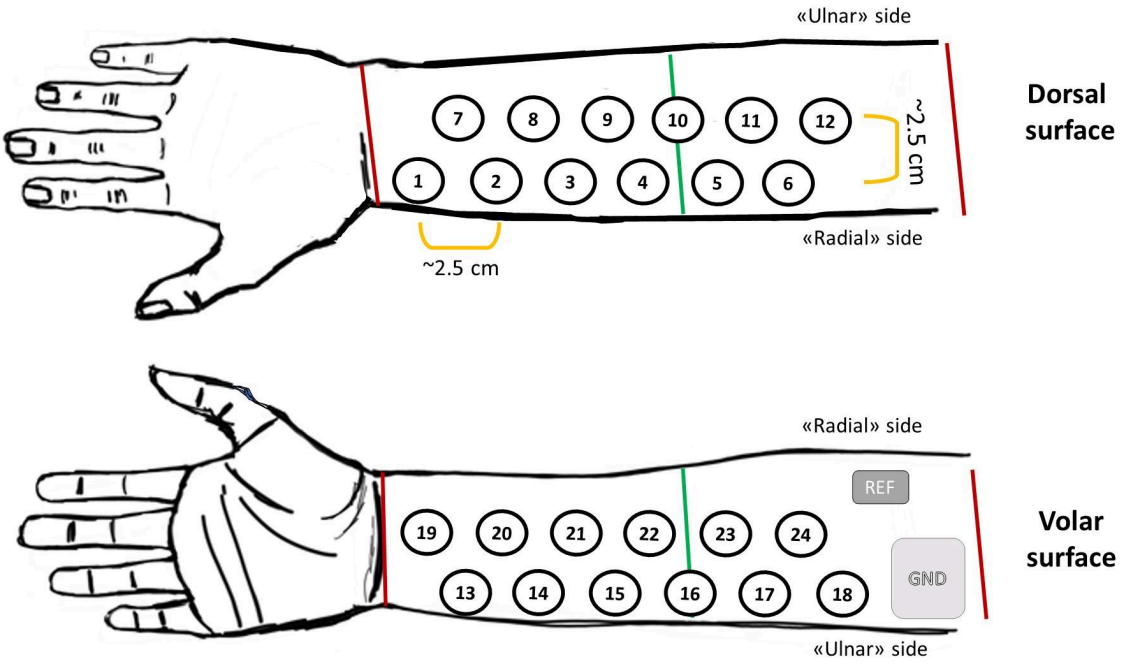

Figure 2. Electrode Grid montage

Four strips, each containing six passive sintered ring electrodes (typically used for electroencephalographic recordings) spaced 2.5 cm apart (24 electrodes total), were used for electromyographic recordings. On the dorsal surface of the forearm, the first strip was aligned along and above the radial bone ("radial" side), with the gap between the fourth and fifth electrode positioned midway down the forearm (green lines), measured from the styloid process to the lateral epicondyle (red lines). The second strip was placed 2.5 cm medially and posteriorly shifted to approximately 1.25 cm. The same arrangement was applied to the volar surface of the forearm, except the third strip was aligned along the "ulnar" side, and the second strip was shifted 1.25 cm forward. Ground and reference electrodes were disposable electrodes, typically placed on the proximal area of the forearm, depending on available space.

2.2.2 TMS

Biphasic TMS pulses were delivered via a 70-mm butterfly coil connected to an STM9000 stimulator (inducing anterior-to-posterior currents) and guided by the NETBRAIN9000 neuronavigation system (EBNeuro, Italy). Participants' heads and the coil were coregistered to a standard 3D MNI brain mesh.

Seated on an armless chair, participants rested their heads on a chin rest and kept their right arm relaxed, dangling with the hand pointing downward. This posture, maintained throughout intensity calibration and the experiment, minimized pre-TMS muscle contractions while allowing unobstructed execution of the displayed actions.

If uncomfortable, participants could lean back to further reduce background muscle activity.

A grid-search procedure was carried out, with the coil held tangential to the scalp approximately 45 degrees with respect the anterior-posterior line, in steps of 5% maximal stimulator output (starting from 20%) to find the "MultiMEP hotspot" on the left hemisphere: the hotspot was defined as the scalp location capable of evoking, at rest, MEPs with amplitudes peak-to-peak ≥ 50 uV in at least 18 of 24 displayed bipolar derivations in five consecutive single pulses (MEPs peaks search time-window: 15 - 65 ms). The lowest intensity meeting this criterion was used as the stimulation intensity for the experimental session (average stimulation intensity: 56.8%+/- 8%). The rationale for the intensity-hunting strategy is described in more detail in the Supplementary Materials.

2.2.3 Stimuli, Task, and Trial Structure

The stimuli consisted of drawings of a right hand representing three different gestures (“three,” “ok,” and “hitchhiking (hh)”). The fingers pointed downwards to resemble the participant’s hand position (Figure 3).

Participants’ underwent a “Move-then-Imagine” task, where they initially had to perform an action displayed on the screen after a Go-signal as fast as possible, relax, and finally engage in sensorimotor imagery of the action just performed: “...imagine the action you have just executed by recalling the muscular tension, the sensation on the skin and within the joints, that the production of that action evoked, but in doing so, do not contract the muscles” (Stinear et al., 2006; Fukumoto et al., 2022). During the imagery phase a single TMS pulse was delivered on the MultiMEP hotspot (Figure 3). A familiarization phase was conducted before the onset of the experimental block (Supplementary Materials).

Each trial started with a blank screen for 1 s, followed by the presentation of one right-hand action drawing. After 1 s, a “Go-signal” high-pitch sound was produced through speakers on each side of the presentation PC monitor. After participants had executed the depicted action, they were required to keep the muscular contraction until, after 1 s, a low-pitched sound was played along with the visually presented “stop” label (1 s), commanding complete muscular relaxation. A feedback screen followed for 0.5 s: If participants anticipated the execution of the action, a red “already contracted” label was displayed as feedback; if participants

produced the action after the deadline, a red "no response" label was shown instead. If they correctly produced the action within the response window, their reaction time (RT, deadline < 350 ms from the Go-signal) was displayed in green (see Supplementary Materials for RT calculation). After the feedback, a 1 s blank screen was presented; next, the same drawing shown at the onset of the trial was shown along with a blue label "*imagine to perform the drawn gesture*"; after 2 s, a TMS pulse was delivered on the participant's MultiMEP hotspot. Following the TMS pulse, they were asked to stop imaging and prepare for the next trial. If pre-TMS muscular contraction was detected, which was defined as a peak-to-peak amplitude greater than 75 uV in the 50 ms before the TMS pulse in at least one of the bipolar derivations, a red "already contracted" label was displayed for 1 s, followed by a "relax" label lasting 1 s. In the latter case or in the case no response was provided after the Go-signal, the trial was marked as non-valid and rescheduled randomly (Figure 4 Experimental Session panel). This strategy allowed us to record 180 clean trials, 60 trials for each of the three imagined actions (see Supplementary Materials Table S1 for a summary of rescheduled trials number).

The action production during the first part of each trial had the double purpose of providing a fresh "haptic-proprioceptive" model that was putatively easier to retrieve from memory during the motor imagery part of the trials (Cattaneo et al., 2015; Maule et al., 2015) and keeping participants engaged, preventing them from getting bored or falling asleep.

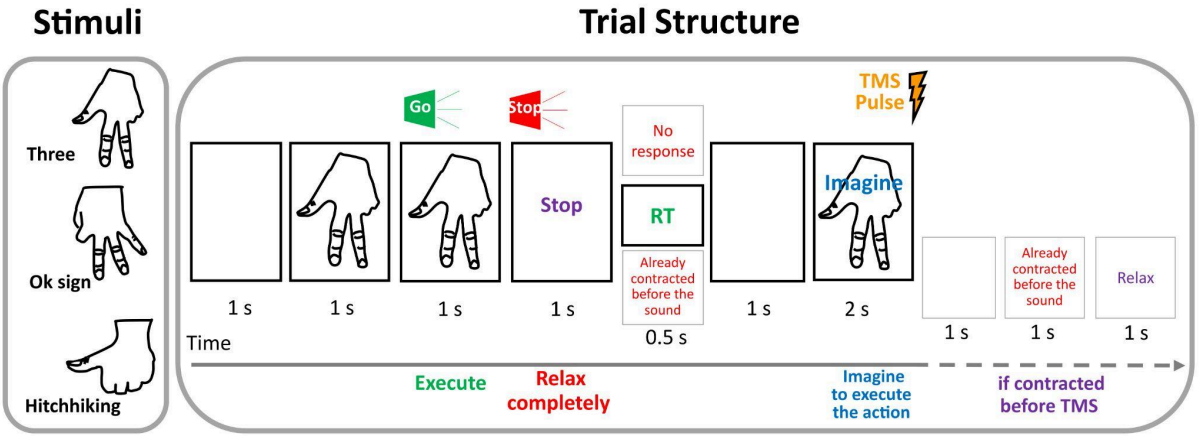

Figure 3 Stimuli and Trial Structure

Stimuli. *Action drawings presented during the experiment: a "three" sign with thumb, index, and middle fingers extended and ring and little fingers flexed; an "ok" sign with middle, ring, and little*

fingers extended while the tips of the index touched the thumb, and a “hitchhiking (hh)” gesture with the thumb extended, and all the remaining fingers flexed. The downward orientation of the drawings matched the participants' hand position.

Trial structure. *Each trial started with an action drawing. After a second, a Go-signal commanded to execute the drawn action as fast as possible and keep the contraction until the stop sound signal, which instructed the muscles to relax completely. Next, feedback was displayed relative to the performance. After a 1 second blank screen, the same action drawing and an “imagine” label were presented again. After 2 seconds, a TMS was delivered on the MultiMEP hotspot. If pre-stimulus contraction was detected (defined as > 75 uV in the 50 ms before the TMS pulse in at least one of the visualized bipolar derivations), feedback was displayed, followed by a relax label.*

3. ANALYSES

3.1 MultiMEP classification

3.1.1 Data Preprocessing

We employed a decoding approach (Figure 1) to discriminate the imaged actions based on the MEP amplitude patterns. All analyses have been produced using MATLAB, including Fieldtrip toolbox (Oostenveld et al., 2011), and custom Python scripts, including Scikit-learn libraries for the classification algorithms (Pedregosa et al., 2011).

Each participant's dataset has been epoched around the TMS trigger between -300 and 500 ms. Trials contaminated by pre-TMS contractions, as defined in the “Stimuli, Task, and Trial Structure” paragraph, were removed, ensuring that exactly 60 valid trials were analyzed for each imagined action.

Forward 4th order Butterworth 10 Hz high-pass and 500 Hz low-pass filters have been applied to padded epochs. Next, the filtered epochs have been downsampled to 1000 Hz to reduce the computational load.

To fully exploit the spatial information embedded in the EMG signals, we implemented a procedure we refer to as “massive bipolarization”. This involved computing all possible bipolar derivations by subtracting the signal of one electrode from another for every unique pair of the 24 monopolar channels. Mathematically, this corresponds to all pairwise combinations of the 24 channels, resulting in 276 bipolar derivations (virtual channels). Thus, each virtual channel reflects the voltage difference between two monopolar electrodes.

For each virtual channel, we calculated the peak-to-peak amplitude of the MEPs within the 15 - 60 ms time window after the TMS pulse.

Thus, a matrix of MEP amplitudes (MultiMEP matrix) constituted by 180 trials (rows) × 276 MEPs virtual channels (columns) has been computed, along with a vector containing the imagined action labels for each trial (Figure 4 Preprocessing panel).

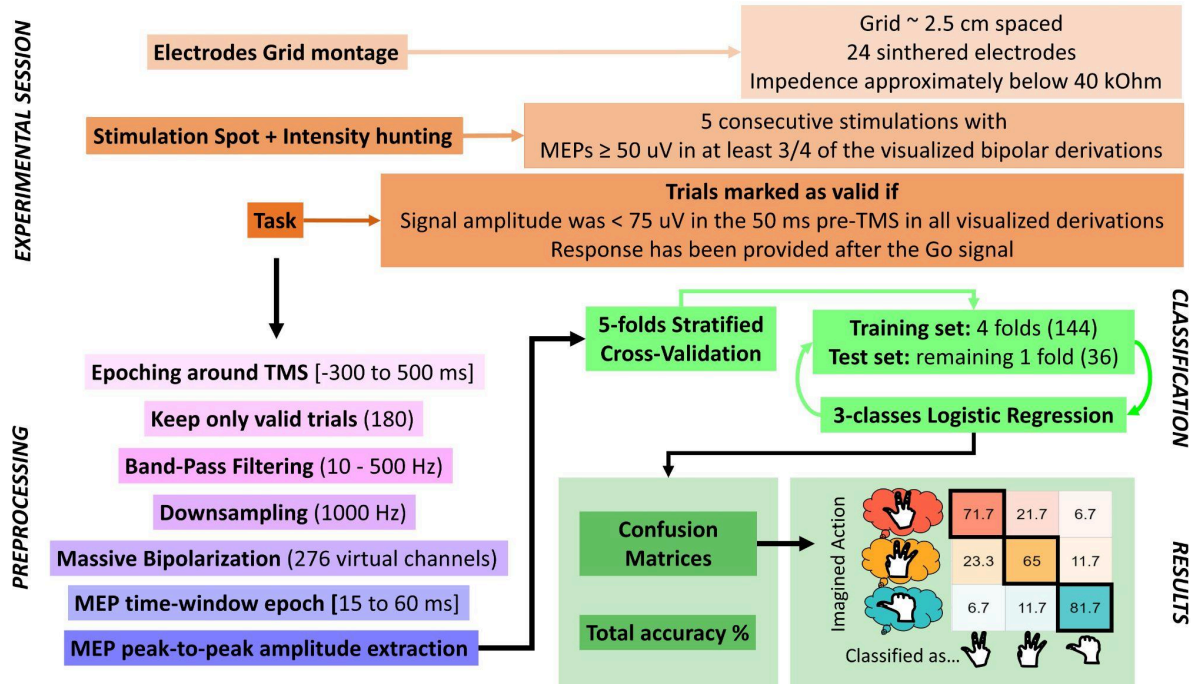

***Figure 4**. The main experimental steps: montage (shades of orange), preprocessing (shades of purple), classification and results (shades of green).*

3.1.2 Classification

The MultiMEP matrix was classified using 3-class logistic regression within a 5-fold stratified cross-validation algorithm: the dataset is randomly split into five equal parts (folds) ensuring that each fold maintained equal proportions of trials belonging to the three classes.

The logistic regression model is trained on four folds (training-set) and tested on the remaining one (test-set), repeating this process five times so that each fold serves as the test-set once (Figure 4 Classification panel). At the onset of each cross-validation cycle the training-set was z-scored, and the same transformation was applied to the test-set. The model performance is averaged across all five runs. No hyperparameter tuning was performed.

We then computed a confusion matrix for each participant: for each action imagined (rows) we calculated the classifier prediction percentages. Eventually a total accuracy metric has been computed indicating the percentage of correctly classified trials. (Figure 4 Results panel).

Three tests were performed:

Single subject total accuracy: We tested whether MultiMEP contained sufficient information to classify imagery content above chance at the single-subject level. To determine statistical significance, we followed the approach of Combrisson & Jerbi (Combrisson & Jerbi, 2015), who argue that the commonly assumed 33.3% chance level (for three classes) applies only with an infinite number of trials. Given that classification under the null hypothesis follows a cumulative binomial distribution, the probability of achieving 38.9%, 41.6%, or 44.4% accuracy in our experiment (3 classes, 60 trials per class) corresponded to p-values < 0.05, < 0.01, and < 0.001, respectively.

Group total accuracy: At the sample level, we tested if total classification accuracies were, on average, above the theoretical chance level by performing a non-parametric Wilcoxon test on total accuracies against 33.3%.

Classification Bias: To test if the total accuracy was inflated by a particularly easy-to-classify action, we extracted the accuracies corresponding to the on-diagonal elements of the confusion matrices for each participant. This resulted in 22 accuracy percentages for each imagined action, which underwent a Friedman test ($p < 0.05$).

3.2 Exploratory Analysis: Cross-classification between contraction and imagery patterns

In this exploratory analysis, we applied a MultiMEP decoding approach to test the claim, derived from classic 1m1a motor imagery works, that action production and motor imagery share common motor representations. To this aim, we took advantage of the contraction patterns (henceforth MultiEMG) collected when participants produced the observed actions. After checking that MultiEMG patterns were well discriminable, we performed two cross-classifications in two directions, by training the classifier on MultiMEP and testing it on MultiEMG (MultiMEP-to-MultiEMG), and vice versa (MultiEMG-to-MulitEMG).

3.2.1 MultiEMG Data Preprocessing

Preprocessing of MultiEMG patterns followed a procedure very similar to the one employed to obtain MultiMEP: the recorded EMG data were epoched between -300 and 1000 ms from the Go-signal. Valid trials were filtered, downsampled, and massively bipolarized. Differently from MultiMEP analysis, root-mean square value was calculated for each trial and each virtual channel, considering a time window ranging from the earliest contraction onset (see Supplementary Materials) until the following 200 ms. Eventually, a 180 trials x 276 matrix was computed along with the labels for each trial.

3.2.2 Sanity Check: MultiEMG Classification

To test if the executed actions were discriminable from MultiEMG patterns, we applied the MultiMEP pipeline analysis to the MultiEMG data.

3.2.3 Cross-classification between contraction and imagery patterns: Data Pre-processing

A crucial preprocessing step was aligning the multidimensional spaces of MultiEMG and MultiMEP, as their amplitudes are not directly comparable. TMS-induced MEPs exhibit muscle-specific biases (TMS-bias pattern) influenced by individual brain-skull anatomy, unlike voluntary contractions recorded in MultiEMG. Assuming a simple additive interaction, MultiMEPs during imagery reflect the combination of both the TMS-bias and imagery-specific patterns. Even if corticospinal activity during imagery mirrored execution, the TMS-bias would shift the position of the MultiMEP patterns within multidimensional space. To correct for this, we z-scored both datasets along the virtual channel dimension, reducing the TMS-bias while preserving trial-specific activity patterns and “along-electrode” relationships (Figure 5).

3.2.4 Cross-classification between contraction and imagery patterns: Classification

After the “along-electrode” standardization, we conducted two cross-classifications using a modified 5-fold cross-validation algorithm. In this approach, the training folds were drawn from one domain (MultiEMG or MultiMEP), while the test fold came from the other (MultiMEP or MultiEMG, respectively). This ensured that, in each fold, the

trials used to fit the classification model remained entirely independent from those in the test set.

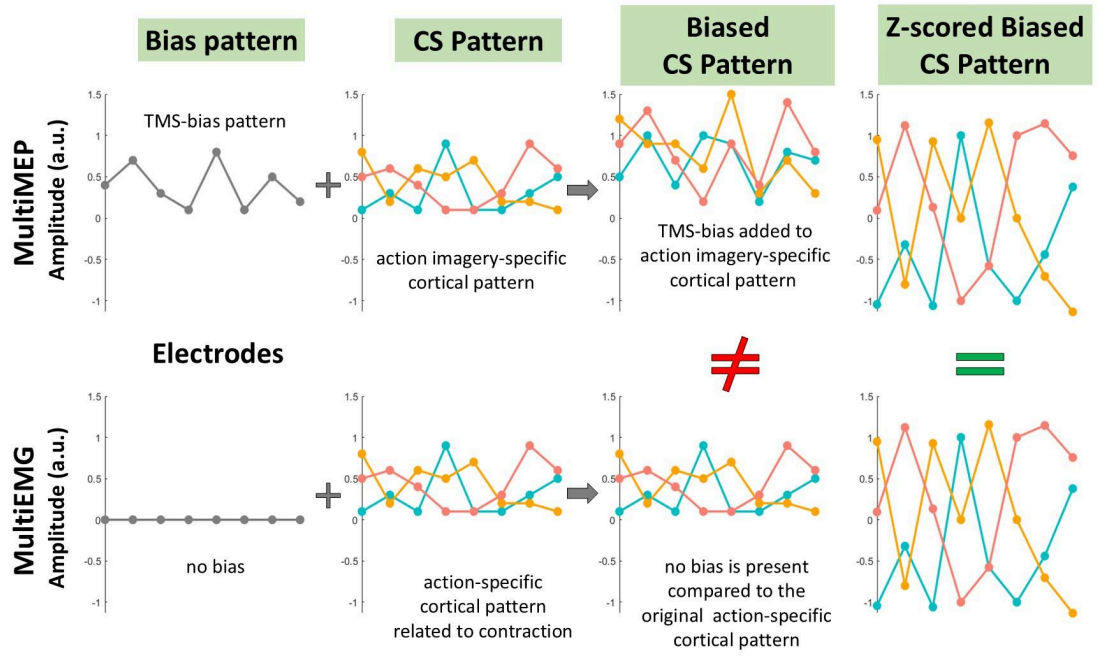

Figure 5. Schematic Representation of the TMS-bias Reduction by z-scoring along the Electrode dimension

Bias Pattern*. Two hypothetical datasets (MultiMEP and MultiEMG) each composed of three different motor representations (different colors) are displayed as rows. Each plot shows eight hypothetical electrodes on the x-axis and amplitude (arbitrary units) on the y-axis. The first plot of the two rows represents the influence of the TMS-bias on electrode amplitude: the bias affects MultiMEP only.*

CS Pattern. *Hypothetical corticospinal representation (CS) of the three motor representations, assumed to be identical between the two datasets by hypothesis.*

Biased CS Pattern. *Interaction between TMS-bias and corticospinal representation: the patterns of the three motor representations are no longer identical between the two datasets, as those in MultiMEP are influenced by the TMS-bias.*

Z-scored Biased CS Pattern*. z-scoring "along-electrode" brings back the similarity between the corresponding motor representations of the two datasets. Note that, compared to the CS Pattern, the relationship between conditions does not change within electrodes.*

3.3 Exploratory Analysis: Cross-Subjects MultiMEP classification

To assess the degree of similarity in decoding spaces across participants, we implemented the following pipeline: the MultiMEP dataset was first z-scored separately for each participant. Then, a logistic regression model was trained on

each participant's data (180 trials; 60 per action class) and tested on the data from each remaining participant. This procedure was repeated for each participant, resulting in 21 cross-subjects classification outcomes per individual. These were then averaged to produce a participant-level metric reflecting the similarity between their MultiMEP-derived decoding space and that of others. To determine whether this cross-subjects similarity exceeded chance level (33%), we compared each participant's average classification accuracy to chance using a Wilcoxon signed-rank test.

3.4 Exploratory Analysis: MultiMEP Classification Accuracy for Electrode Grid Subdivisions

In the present experiment, electromyographic activity was recorded from 24 surface electrodes arranged over the forearm. However, it is possible that not all 24 electrodes are necessary to achieve optimal classification accuracy. Furthermore, certain forearm regions may contribute more consistently to discriminating imagined actions than others.

To investigate these possibilities, we assessed classification performance using subsets of electrodes defined by spatial subdivisions: four Quadrants, four Half-Grid regions, and the Whole Grid.

Quadrants subdivisions: The grid was divided into four anatomical quadrants: dorsal anterior (electrodes 1, 2, 3, 7, 8, 9), dorsal posterior (4, 5, 6, 10, 11, 12), volar anterior (13, 14, 15, 19, 20, 21), and volar posterior (15, 16, 17, 22, 23, 24) aspects of the forearm.

Half-Grid subdivisions: Two orthogonal subdivisions were used: one separating the dorsal (electrodes 1–12) and volar (13–24) aspects, and another separating the anterior (1, 2, 3, 7, 8, 9, 13, 14, 15, 19, 20, 21) and posterior (4, 5, 6, 10, 11, 12, 15, 16, 17, 22, 23, 24) aspects of the forearm.

Whole Grid: All 24 electrodes (1–24) were included in the analysis.

For each spatial subdivision, we computed all possible bipolar derivations resulting in: 15 virtual channels for each Quadrant, 66 for each Half-Grid region, and 276 for the Whole Grid. This resulted in four 180 × 15 matrices (one per quadrant), four 180

× 66 matrices (one per Half-Grid), and one 180 × 276 matrix (Whole Grid). These matrices were entered into a classification procedure identical to that used for the MultiMEP-to-MultiMEP analysis.

To statistically assess differences in classification performance across the nine spatial configurations, we first applied a Friedman test. Pairwise comparisons between subdivisions were then conducted using paired Wilcoxon signed-rank tests, with Benjamini-Yekutieli correction applied to control for multiple comparisons.

4. RESULTS

4.1 MultiMEP classification

Single subject's total accuracy: We tested how well single participants' total accuracy classification departed from an empirical chance level: following (Combrisson & Jerbi, 2015), the likelihood that a 44.4% classification accuracy is obtained by chance corresponds to a p-value < 0.001, for a 41.6% accuracy to $p < 0.01$, and for a 38.9% accuracy to $p < 0.05$. Total accuracy was above 44.4% for every participant ($p<0.001$, Figure 6).

Group total accuracy: We calculated if the average total accuracy among participants was higher than the theoretical chance level of 33.3%. The average total accuracy was 74% +/- 10% sd, significantly higher than the group chance level of 33.3% ($W = 0$, $p < 0.001$, Figure 6).

Classification Bias: We tested whether actions were classified with different accuracies across participants: no biases have been detected (Friedman $Q = 0.17$, $p = 0.91$; average accuracies: "three" = 75.5%+/- 11% sd, "ok" = 74.1%+/- 10.5% sd, "hh" = 73.3%+/- 12.0% sd).

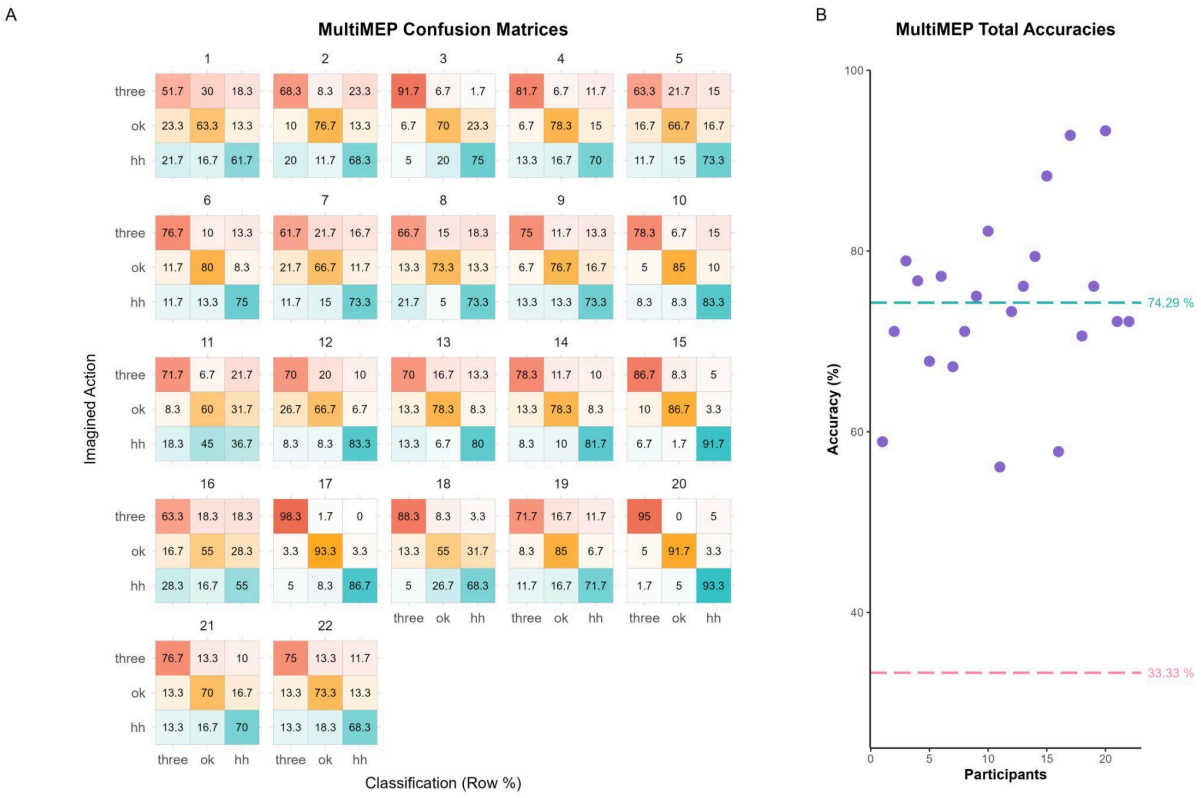

Figure 6. MultiMEP Confusion Matrices and Total Accuracies

A*. MultiMEP confusion matrices from each participant: within each imagined action category ("three" in red, "ok" in gold, "hitchhiking" - "hh" - in light blue), cells represent the % of trials predicted to belong to each of the three actions ("three", "ok" and "hh"). A perfect classification would highlight 100% on the main diagonal cells and 0% on the remaining cells of each row. Darker colors indicate better classification.*

B. *MultiMEP accuracies: each dot represents the classification total accuracy for each participant. The green dotted line represents the average total accuracy, and the pink one represents the theoretical chance level.*

4.2 Exploratory Analysis: Cross-classification between contraction and imagery patterns

4.2.1 Sanity Check: MultiEMG Classification

MultiEMG classification showed an almost flawless total classification accuracy (98.7% +/- 0.01% sd) with all participants above 44.4% (Supplementary Materials Figure S2).

4.2.2 MultiEMG-to-MultiMEP Classification

The MultiEMG-to-MultiMEP classification was produced by training a classifier model on the MultiEMG data and then predicting MultiMEP imagined actions based on the hyperplanes estimated from the MultiMEG contraction patterns.

Single subject's total accuracy: The total classification accuracy was above 44.4% ($p<0.001$, Figure 7) for 19 out of 22 participants, above 41.7% for 20 out of 22 participants, and above 38.9% ($p<0.05$) for 21 out of 22 participants.

Group total accuracy: The average total accuracy was 54.3% +/- 11.2% sd, significantly higher than the group chance level of 33.3% ($W = 0$, $p < 0.001$, Figure 7).

Classification Bias: Classification biases have been detected (Friedman $Q = 7.78$, $p = 0.02$), We employed three Wilcoxon tests to explore the differences in classification accuracy, comparing each combination of actions.

"three" vs. "ok": $W = 116.5$, $p = 0.75$, "three" (49.1% +/-16.9% sd), "ok" (51.5% +/- 17.5% sd).

"three" vs. "hh": $W = 35.0$, $p = 0.005$, with "hh" (62.2% +/- 14.4% sd) more easily classifiable than "three" (49.1% +/-16.9% sd).

"ok" vs. "hh": $W = 43.0$, $p = 0.02$, with "hh" (62.2% +/- 14.4% sd) more easily classifiable than "ok" (51.5% +/- 17.5% sd).

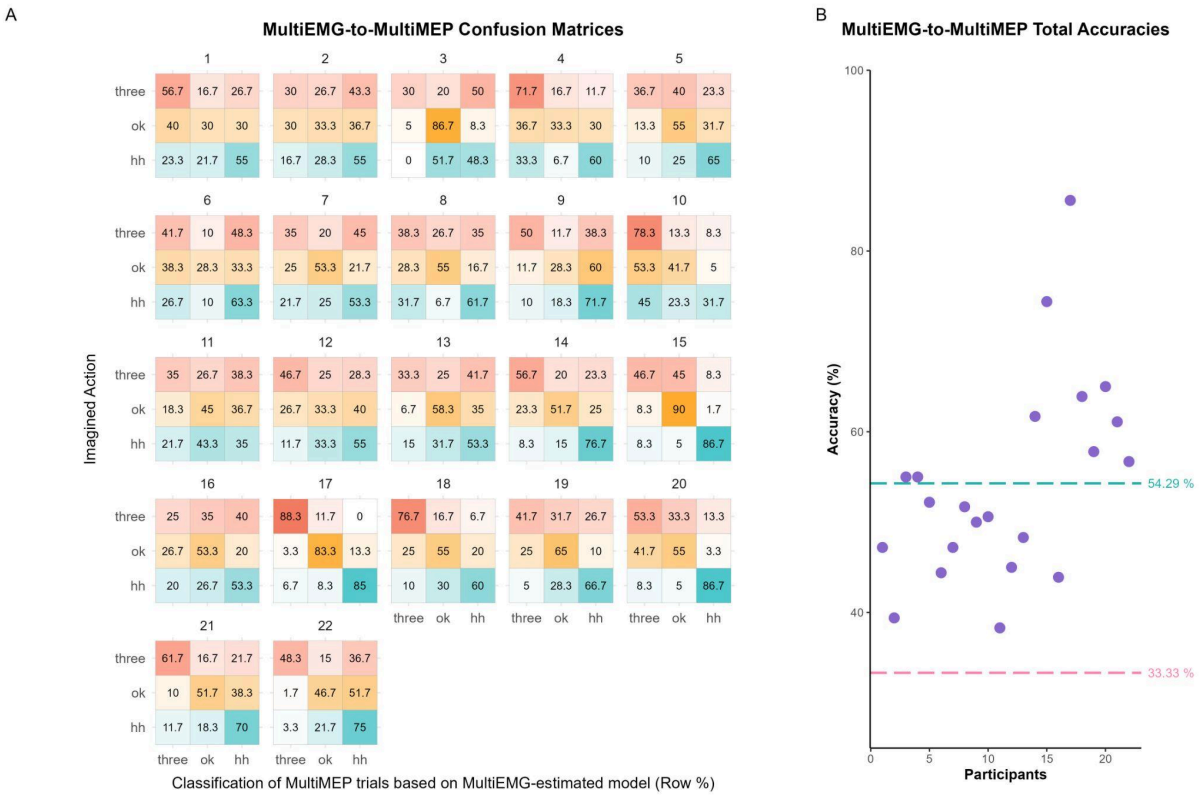

Figure 7*. *MultiEMG-to-MultiMEP Confusion Matrices and Total Accuracies

A. *MultiEMG-to-MultiMEP confusion matrices for each participant: rows represent the imagined action ("three" in red, "ok" in gold, "hitchhiking" - "hh" - in light blue). The cells constituting each row represent the % of trials predicted for the three action categories; differently from the MultiMEP analysis, the prediction model has been estimated using the MultiEMG patterns.* ***B.*** *MultiEMG-to-MultiMEP accuracies: each dot represents the classification total accuracy for each participant. The green dotted line represents the average total accuracy, and the pink one represents the theoretical chance level.*

4.2.3 MultiMEP-to-MultiEMG Classification

The MultiMEP-to-MultiEMG classification was produced by training a classifier model on the MultiMEP data and then predicting MultiEMG imagined actions based on the hyperplanes estimated from the MultiMEP patterns.

Single subject's total accuracy: The total classification accuracy was above 44.4% ($p<0.001$, Figure 8) for all participants

Group total accuracy: The average total accuracy was 71.7% +/- 13.1% sd, significantly higher than the group chance level of 33.3% ($W = 0$, $p < 0.001$, Figure 8).

Classification Bias: No classification bias has been detected (Friedman Q = 1.95, p-value = 0.37).

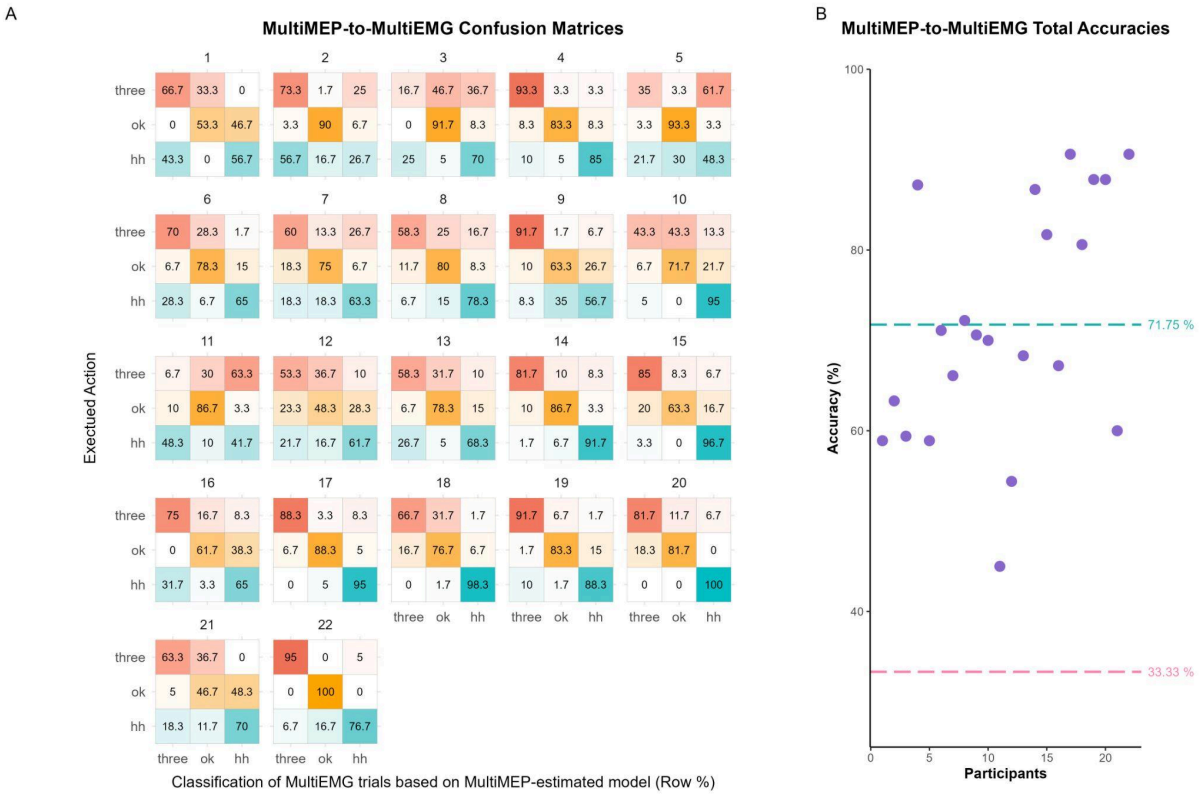

Figure 8. MultiMEP-to-MultiEMG Confusion Matrices and Total Accuracies

A. *MultiMEP-to-MultiEMG confusion matrices for each participant: rows represent the executed action ("three" in red, "ok" in gold, "hitchhiking" - "hh" - in light blue). The cells constituting each row represent the % of trials predicted for the three action categories; differently from the MultiMEP analysis, the prediction model has been tested on the MultiEMG patterns.* ***B.*** *MultiMEP-to-MultiEMG accuracies: each dot represents the classification total accuracy for each participant. The green dotted line represents the average total accuracy, and the pink one represents the theoretical chance level.*

4.3 Exploratory Analysis: Cross-Subjects MultiMEP classification

Results revealed an average cross-subject classification accuracy of 41.9% +/- 4.33% (Wilcoxon W = 253, $p < 0.001$) for the MultiMEP dataset. This indicates that, on average, a model trained on one participant's data was able to predict the imagined action from another participant's data nearly 42 times out of 100, suggesting a modest but reliable degree of similarity in decoding spaces across individuals (Figure 9).

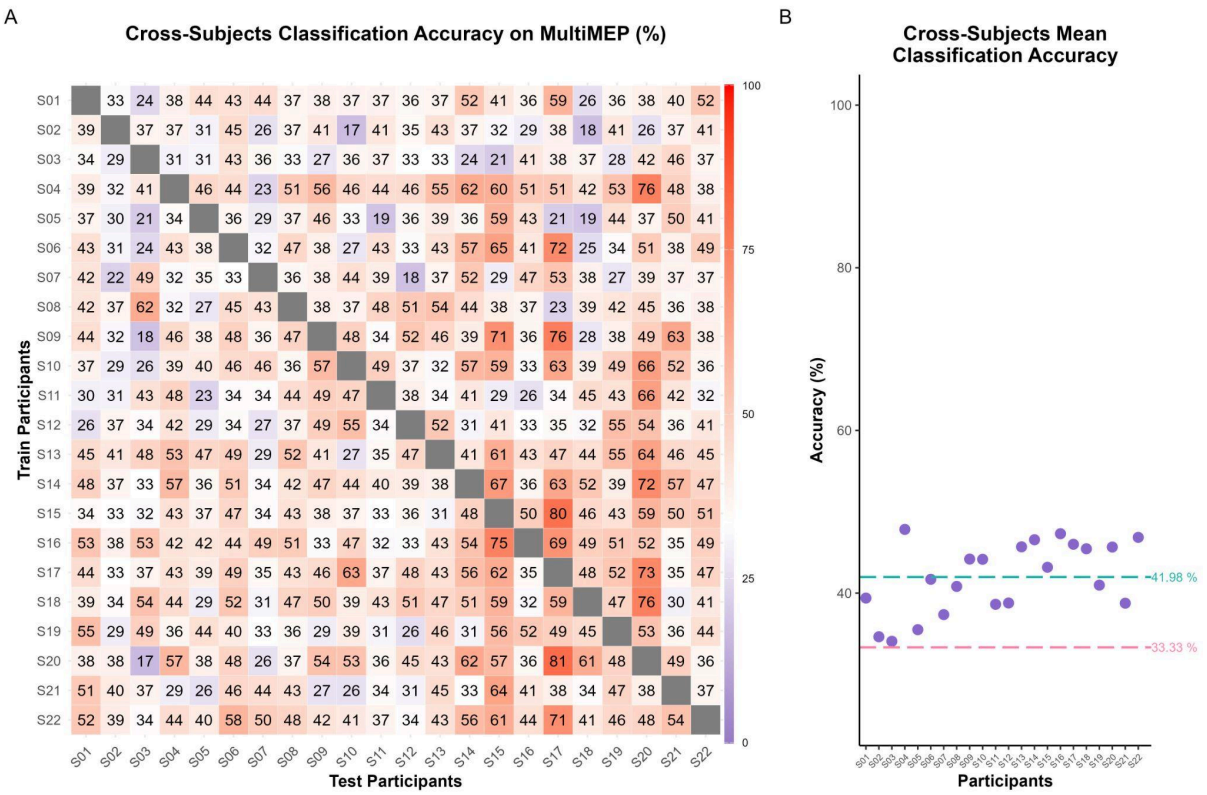

Figure 9. Cross-Subjects Classification on MultiMEP

A*. Cross-subjects confusion matrix: each cell indicates the classification accuracies obtained by training the classifier on the participant listed on each row and classifying on each remaining participant listed on the columns. A colorimetric scale is displayed with red indicating 100%, dark blue 0% and white 33% accuracy.*

B. *The plot displays the average classification accuracy of each participant.*

The same procedure applied to the MultiEMG dataset yielded a significantly higher cross-subject classification accuracy of 67% +/- 10% (W = 253, $p < 0.001$), suggesting that overt action execution involves more consistent patterns of muscular activation across participants (Figure 10).

These results suggest that participants partially share a common representational structure for imagining different actions. The relatively low classification accuracy likely reflects not only inter-subject variability in motor imagery, but also other factors that may have similarly impacted MultiEMG cross-subject classification. For instance, electrode spacing was fixed at 2.5 cm regardless of individual forearm anatomy, and slight variations in how participants executed the actions, cannot be entirely ruled out. Such sources of variability may have limited the consistency of the recorded patterns across individuals.

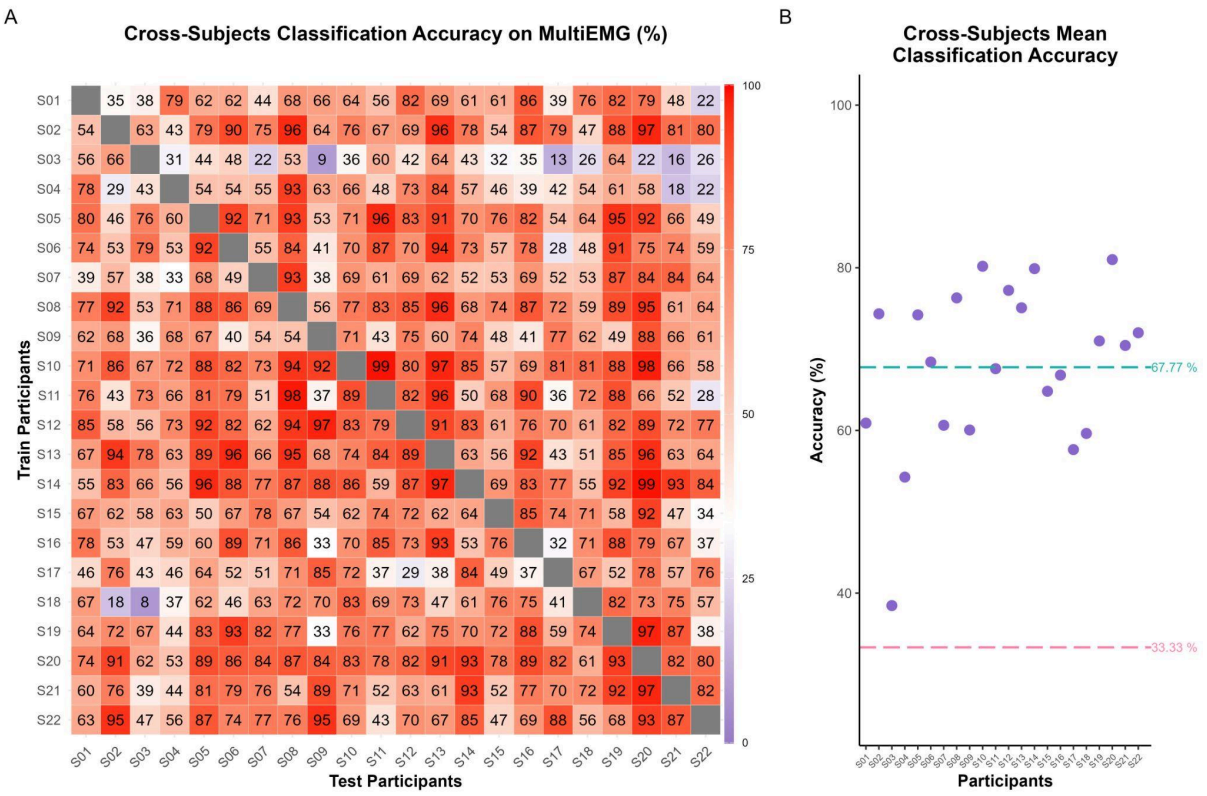

Figure 10. Cross-Subjects Classification on MultiEMG

A*. Cross-subjects confusion matrix: each cell indicates the classification accuracies obtained by training the classifier on the participant listed on each row and classifying on each remaining participant listed on the columns. A colorimetric scale is displayed with red indicating 100%, dark blue 0% and white 33% accuracy.*

B. *The plot displays the average classification accuracy of each participant.*

4.4 Exploratory Analysis: MultiMEP Classification Accuracies for Electrode Grid Subdivisions

The Friedman-test applied to the 9 subdivisions indicates that these differ with respect to classification accuracies (Friedman Q = 124.07, p < 0.001; average accuracies on Table 1). The exploration of these differences by means of pairwise Wilcoxon tests indicates that, within each size subdivision, no forearm portion detects information better than other, i.e., on average all Quadrants pick similar information, and all Half-grids subdivisions pick similar information as well. The only consistent differences have been detected between grid sizes, as the Whole Grid classifies consistently better than each Half-grid subdivision, and each Quadrant. Similarly

each Half-grid subdivision classifies better than each Quadrant (Figure 11; see Figures S3-S7 for more details).

Average MultiMEP Classification Accuracy for Electrode Grid Subdivisions			
	Subdivisions	**Means**	**SD**
Quadrants	Dorsal Anterior	59.5%	10.9%
	Volar Anterior	60.3%	10.3%
	Dorsal Posterior	60.5%	9.6%
	Volar Posterior	58.9%	10.7%
Half-Grids	Anterior	69.9%	10.4%
	Posterior	69.2%	9.7%
	Volar	67.6%	10.1%
	Dorsal	69.2%	11.1%
Whole Grid		74.3%	9.8%

Table1. Average MultiMEP classification accuracy for electrode grid subdivisions.

The table shows mean and standard deviation accuracy percentage for each subdivision.

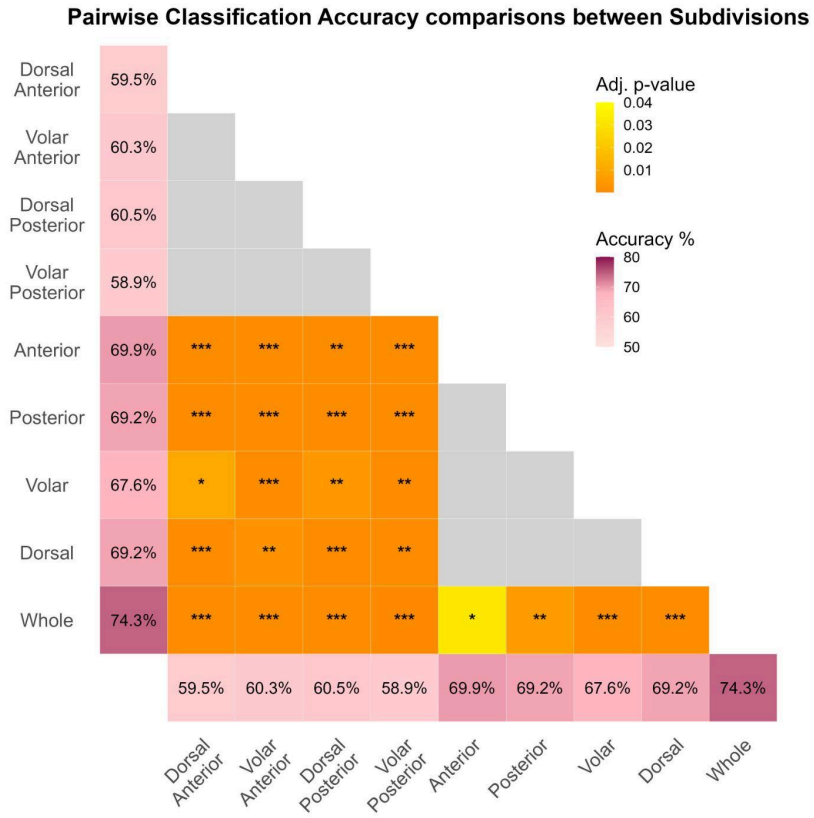

Figure 11. Pairwise classification accuracy comparisons between subdivisions.

The figure shows the pairwise classification accuracy comparisons between all subdivisions. The first column and the last row (pink-shaded vectors) show the mean classification accuracy for each subdivision. The orange-shaded cells report significant p-value for each comparison ($p < 0.05$, Benjamini-Yekutieli corrected), while grey cells represent non-significant comparisons.

5. DISCUSSION

Design constraints in traditional 1m1a studies on covert motor processes limit the information retrievable from MEPs. In this proof-of-principle study, we adopted a decoding approach to multidimensional MEPs, achieving over 70% classification accuracy for three complex imagined actions in all participants. This result suggests that single-trial MEP patterns carry more information than previously (implicitly) assumed.

Notably, we successfully classified actions involving different forearm muscles, broadening the potential of this technique to study covert motor processes for virtually any action. Expanding electrode coverage to the hand could further enhance accuracy. Once trained, the classifier enables single-trial classification of imagined actions, opening new research avenues. For example, MultiMEP can reveal the

geometry of motor representations by analyzing how different covert motor processes are positioned in MultiMEP-space, akin to primate studies on action preparation and production (Elsayed et al., 2016; Kaufman et al., 2013; Kaufman et al., 2015; Meirhaeghe et al., 2023).

Beyond distinguishing complex imagined actions, MultiMEP decoding allows inferences on the state of motor representations, much like 1m1a experiments. To test this, we computed contraction patterns for each action (MultiEMG) and compared them against MultiMEPs evoked during motor imagery. Our bidirectional cross-classification showed that hyperplanes estimated in one domain could discriminate classes in the other, in line with the claim that motor imagery and action production share the motor representations. However, despite being above chance in both directions, the analysis revealed an asymmetry: MultiEMG-to-MultiMEP achieved 54% accuracy (with a positive bias in classifying “hh” actions), while MultiMEP-to-MultiEMG reached an accuracy of 71%. This asymmetry likely reflects differences in the intrinsic discriminability of the two domains (classes within MultiEMG patterns being more separable than in MultiMEP), where classifiers trained on the more distinct dataset do not generalize well to the noisier one, despite a similar underlying structure. Following Van Den Hurk & Op De Beeck (2019[(Preprint)]), we interpret the MultiMEP-to-MultiEMG results as a better reflection of the actual overlap between domains.

To explore whether motor imagery of the three actions produce similar patterns across participants we conducted a cross-subject classification analysis by training the classifier on each participant and testing its predictions on each remaining subject. This yielded a mean accuracy of 42% for MultiMEP (motor imagery) and 67% for MultiEMG. Although the MultiMEP performance was lower compared to other results in the present work, it was consistent across individuals and may reflect both individual differences in motor imagery and broader variability also evident in the MultiEMG classification. Factors such as fixed electrode spacing and subtle differences in action execution likely contributed to this variability, highlighting the need for future work to better control for anatomical and behavioral heterogeneity.

Eventually, we also explored the contribution of different forearm regions and the effect of grid size on classification performance. Across subjects, no single forearm portion consistently outperformed others in decoding imagined actions, likely

because the muscles involved in each action spanned the entire grid. In contrast, grid size had a clear and consistent impact, with larger grids yielding higher classification accuracies. Notably, the improvement from Quadrants to the Whole Grid suggests a non-linear benefit, which suggests that our 24-electrode configuration may have approached a performance plateau.

In general, moving from a univariate to a multivariate datum allows to investigate problems from different perspectives, which might benefit the study of covert motor processes such as action observation or motor preparation. MultiMEP offers the possibility to study the processing of more ecological complex actions, adding generalizability to results, and putatively reducing the impact of sudden excitability shifts that decrease the signal-to-noise ratio, thus providing more stable outcomes. Most importantly though, it allows to explore different aspects of the datum that could not be captured through a 1m1a approach: for example, it has been shown that during motor preparation the corticospinal excitability inhibition precedes action execution (Duque et al., 2010). Speculating, this inhibition might be accompanied to a certain degree by changes in cortical excitability patterns as well, which might clarify the different ways preparation affects selected, non-selected and irrelevant effectors (Bundt & Huster, 2024). Hypothetically, different stages of preparations might be discovered thanks to similarity analyses of MultiMEP pattern. Theoretically it might be possible to classify different covert motor states, such as imagery vs. preparation, and characterize their underlying multidimensional structures. MultiMEP decoding offers the possibility to study single-subject specific patterns in order to highlight the differences and the commonalities between participants. Eventually, single-trial characteristics should now represent a richer source of information: for example, is the variability of the MultiMEP pattern shape different throughout the delay period before imperative signals?

Although the fine-grained informational content recoverable through a MultiMEP approach remains to be fully understood, it would be worth investigating whether covert motor activities involving different motor contents and effectors (e.g., ipsilateral hand, feet, or mouth) are reflected in subtle or subthreshold pattern-specific modulations within the hand motor cortex. For instance, previous research has shown that preparing foot movements increases beta-band amplitude in the hand motor cortex (Pfurtscheller et al., 1997); it is possible that different feet

actions preparation induce pattern-specific activities in the hand motor area. While such an outcome may seem counterintuitive, MultiMEP provides a unique opportunity to test similar possibilities and, importantly, to assess them in a subject-specific way. These are only a few open questions that we believe can be addressed when studying covert motor processes employing a MultiMEP setting.

Moreover, the MultiMEP setting can lead to different practical advantages. Notably, its putative increased sensitivity in detecting motor-related activity patterns may reduce the number of trials needed to detect condition differences, or on the other hand, extracting information-rich patterns that enhance effect sizes compared to traditional 1m1a approaches. The MultiMEP setting can also be used to inform new criteria for choosing stimulation hotspots. For instance, rather than relying on peak MEP amplitude from a single muscle, one could average MEP amplitudes across multiple electrodes to obtain a more stable estimate of general corticospinal excitability. Alternatively, hotspots could be defined based on non-standard criteria, such as maximal or minimal amplitude variability across electrodes, the presence of specific MultiMEP activation patterns, or any other metric computable on a distribution of data. To enable such applications, information must be extracted online via an automated MultiMEP-Viewer. Integrating MultiMEP with real-time, automated state monitoring provides immediate feedback about the recorded physiological signal. For example, real-time MultiMEP settings might assist motor imagery training: in this scenario participants may imagine performing an action, while MultiMEP are evoked to assess the similarity between the imagery-induced state and the pattern shape associated to actual motor execution of the action.

Given the richness of MultiMEP data, it could be integrated into other TMS protocols, including rTMS protocols (Hamada et al., 2013; Huang et al., 2005), paired-pulse (Coxon et al., 2006; Mrachacz-Kersting et al., 2021) and double-coil (Arai et al., 2012; Cattaneo et al., 2011; Davare et al., 2009; Davare et al., 2010) paradigms, shifting from a unidimensional excitation/inhibition framework to a multidimensional perspective. Similarly, combining MultiMEP with TMS-evoked EEG activity could provide insight into how motor measures relate across the two different domains (Bortoletto et al., 2021; Guidali et al., 2023; Zazio et al., 2022; Thong et al., 2025).

This paradigm shift may also benefit studies of pathological motor systems. Clinically, MEPs are used to assess corticospinal integrity (Arora et al., 2022; Chen

et al., 2008), but offer little insight into cortical motor representations. MultiMEP could characterize motor system alterations with high granularity, extending its application to neuropsychiatric conditions where unidimensional MEP-derived measures are commonly used as biomarkers (Kaskie & Ferrarelli, 2018; Vucic et al., 2023).

Compared to standard MVPA approaches in neuroimaging, MultiMEP offers key advantages. While fMRI and EEG-based Brain-Computer Interfaces achieve high classification accuracy (90–95%) in tasks with well-separated motor representations (e.g., left vs. right-hand movements), they rarely explore finer effector-specific actions (Khan et al., 2023; Lotte et al., 2018; Padfield et al., 2019; Sharma et al., 2022; Xygonakis et al., 2018). MultiMEP, by reflecting corticospinal states during covert motor tasks, extends motor system decoding from localized cortical patches to a physiologically meaningful space, mapping motor information onto hand and forearm muscles, where it can be more easily detected and discriminated. To evaluate MultiMEP potential as a Brain-Computer Interface, however, future studies must assess its test–retest reliability, with particular attention to how the latter is influenced by different tasks demands.

Finally, while the MultiMEP decoding framework differs methodologically from 1m1a experiments, the two forms do not represent a dichotomy. Intermediate strategies, such as decoding two-muscle setups, using unevenly spaced electrodes, or comparing averaged evoked patterns across conditions, can bridge the gap between traditional and multidimensional analyses, depending on the research question.

6. CONCLUSIONS

In summary, the MultiMEP decoding approach highlights the high discriminative power of MEP patterns at both single-subject and single-trial levels. Moreover, it extends to the inferential reach of 1m1a approaches while improving the generalizability of the results.

Supplementary Materials

Instructions

“In each trial, you are going to observe one of three actions for a brief period of time, after which a high-pitched sound will be played; when you hear the high-pitched sound (which represents your Go-signal), you have to execute that action with your right hand as fast as possible and as strongly as possible (you have less than half a second to respond). After a second, you will hear a low-pitched sound along with a “stop” label on the screen: at this point, you have to completely relax your hand and arm muscles, including shoulders and arms. Next, a feedback will be displayed: if the response was provided beyond the deadline or it was not provided at all, you will observe a “no response” label; if your muscles were contracted already before the Go-signal, you will observe an “already contracted” label; if you responded within the deadline, you will see you reaction time in milliseconds. Afterward, you will be presented with the same action drawing on top of which there will be a “imagine” blue label: once the picture appears, you have to produce sensorimotor imagery of the action you have just executed by recalling the muscular tension, the sensation on the skin and within the joints, that the production of that action evoked, but in doing so, do not contract the muscles.

After 2 s a TMS pulse will be delivered on the scalp, after which you can stop imaging. A blank screen will appear, and a new trial will begin next. If some of your muscles were already contracted before the TMS stimulation, an “already contracted” label will appear on the screen, followed by a “relax” label. In this case, try to focus on your muscle tension and relax completely from shoulders to fingers.”

Training session

At the beginning of the experimental session, participants underwent ~20–30 trials. They were required to only execute the action observed at the Go-signal, thus focusing exclusively on action performance without engaging in motor imagery. They were instructed to produce the action depicted on the monitor with maximal contraction, maintain the contraction until the stop signal, and then completely relax shoulders, arm, and hand muscles. The practice session continued until participants could quickly and consistently perform the action at the Go-signal, sustain the contraction, and achieve relaxation after the stop signal.

Once this was achieved, the monitor was turned off, and the motor imagery task was introduced. Participants were asked to close their eyes, execute one of the three actions, relax completely, and then imagine performing the same action by recalling the muscular tension and sensations from the skin and joints evoked by the action, all without contracting their muscles. The experimenter encouraged participants to enhance the vividness of their imagery until contractions appeared in the continuous EMG stream that was displayed in the meanwhile; once small contractions were observed by the experimenter, participants were instructed to stop and slightly scale down the intensity of their imagination during the experiment.

When participants felt confident with action imagery, they proceeded to ~20–30 complete experimental trials, during which they were required to both perform the action at the onset of each trial and subsequently imagine it motorically.

All the trials within training sessions were conducted without TMS stimulation.

After the training session, the experimental session started.

Muscular contraction and MEPs amplitude detection on the "MultiMEP Viewer"

The "MultiMEP Viewer" is a custom-made signal viewer displayed on the amplifier-PC (the computer acquiring streamed data from the amplifier). The Viewer is composed of two windows: the "continuous stream window," where the continuous native monopolar data streams were plotted, and the "multi-epochs window," where bipolar derivations of the recorded EMG signals were displayed between -150 to 400 ms from the TMS trigger. The time-wide window allowed to automatically calculate (online): a) the presence of pre-Go-signal or pre-TMS contractions defined as at least one of the bipolar derivations showing > 75 uV peak-to-peak amplitude in the 50 ms preceding the Go or the TMS trigger, b) MEPs peak-to-peak amplitude was automatically calculated between 15 ms to 65 ms from the TMS trigger, and c) the onset of an action production based on the contraction (reaction times).

Thanks to the Lab Streaming Layer functions a real-time communication was set up by sending triggers related to stimuli presentation from the presentation-PC to the amplifier-PC and, vice versa, sending events triggers from the amplifier-PC to the presentation-PC related to participants' behaviors, as pre-Go-Signal or pre-TMS trigger contractions, and reaction times, no-responses (or network malfunctioning). In case pre-TMS trigger contractions were presented, the trial was discarded and rescheduled randomly in order to obtain 60 analyzable trials for each action for each

participant (see Table S1); the same applied if a no-response was detected when participants had to produce the action observed at the onset of each trial, or if a network communication malfunctioning occurred.

Rescheduled trials				
Subjects	**Three**	**Ok Sign**	**Hitchhiking**	**Total**
1	5	10	18	33
2	6	9	4	19
3	9	7	8	24
4	8	21	14	43
5	8	10	6	24
6	58	22	3	83
7	30	15	17	62
8	10	12	6	28
9	20	9	7	36
10	12	25	11	48
11	12	3	0	15
12	9	14	10	33
13	6	3	7	16
14	13	20	11	44
15	31	26	13	70
16	16	12	10	38
17	23	14	19	56
18	8	17	10	35
19	3	24	3	30
20	5	14	5	24
21	11	8	10	29
22	8	11	12	31
Mean	**14.13**	**13.90**	**9.27**	**37.32**
SD	**12.42**	**6.72**	**5.00**	**17.48**

Table S1. *The table shows the number of rescheduled trials for each subject and action ("Three", "Ok Sign" and "Hitchhiking") in the first three columns. The last two rows of the table present the mean and standard deviation across all participants for each action.*

We did not set the Viewer to classify the type of action produced, so while one of the researchers was holding the coil on the participant's head, the other was visually checking both the continuous data stream to monitor the basal contraction of the recorded muscles, and that the action produced was the correct one.In case a contraction was spotted before the TMS pulse as described above or a response

could not be detected after the Go-signal, the trial was reinserted into the schedule of the remaining trials, and the schedule was reshuffled. This approach allowed exactly 60 clean trials i.e., without pre-TMS contractions in the 50 ms before the pulse) for each imagined action.

Rationale behind MultiMEP intensity-hunting procedure

One challenge we faced was determining the most appropriate criterion for setting stimulation intensity in an experimental context where many bipolar derivations were expected to elicit MEPs. To our knowledge, no previous study has addressed this issue as nobody employed a MultiMEP approach as we did. Consequently, our decision-making was guided entirely by pilot sessions conducted beforehand.

Had we adopted the classic Rossini-Rothwell method, where the threshold is determined using a single muscle, we would have had to perform threshold hunting on just one bipolar derivation. However, this approach presents two major limitations: (1) while it ensures reliable MEPs for that specific derivation, it does not guarantee similar responses for other forearm sites; and (2) it requires an a priori selection of a single derivation.

An alternative strategy would have been to apply the Rossini-Rothwell thresholding criterion to all 24 bipolar derivations as if they constituted a single electrode, requiring at least 5 out of 10 trials across all derivations to exceed 50 μV. Although this would avoid the issues of the previous approach, it imposes a stringent and perhaps unrealistic constraint: requiring every derivation to exceed 50 μV could lead to excessively high stimulation intensities. This is particularly problematic because some cortical neurons responsible for activating specific muscle fibers may not be recruitable in all participants due to factors such as scalp-cortex distance or cortical organization. Pushing the stimulation intensity to meet this criterion may result in (a) unreasonably high intensities that exceed tolerable limits and (b) ceiling effects in easily excitable neurons, potentially obscuring modulation effects induced by motor imagery and thereby reducing classification accuracy.

Instead, based on our pilot data, we adopted a more flexible and pragmatic criterion: requiring that at least 18 out of the 24 bipolar derivations consistently elicit MEPs larger than 50 μV in 5 out of 5 consecutive trials. This approach strikes a balance between reliability and practicality: it does not demand uniform MEP responses

across all derivations, allowing for inter-muscle variability, yet still enforces a meaningful and consistent level of excitability across the majority of recording sites.
A potential limitation of this solution is its dependence on task-specific features. While it proved suitable for our motor imagery paradigm, where excitability levels are relatively stable between the thresholding and experimental phases, it may be less appropriate in paradigms such as passive action observation. In such contexts, general excitability may decrease from the thresholding to the experimental session, potentially reducing the signal-to-noise ratio and compromising classification performance.
The average stimulation intensity for the current intensity-hunting procedure is 56.8%+/- 8% (paragraph 2.2.2 TMS) with a minimum of 40% and a maximum of 72%. While the intensity is typically measured in MSO%, and thus a direct comparison is typically difficult between stimulators, the average intensity applied in the present experiment well within ranges of intensities employed in studies using the same stimulator (Emandi et al., 2015; Rossi et al., 2021).

Reaction Times on-line calculation (MultiMEP Viewer)

For each channel, within the 1 ms to 400 ms window after the Go-signal trigger, we calculated the difference between the maximum and the minimum of the EMG between -5 ms to +5 ms centered around each sampling point of the epoch (i.e., using a sliding window of 10 ms centered on each sample by using the result of movmax - movmin, orange line in Figure S1). Next, the algorithm sought the first window around a sample of the “smoothed max-min EMG” signal (window radius +/-25 ms) in which the “smoothed max-min EMG” signal was always > 75 uV (magenta vertical line, while the cyan horizontal represents 75 uV).
To account for the centering of the 50 ms sliding window, 25 ms were removed from the detected sample, identifying the sample around which the first -/+5ms sliding window showed a max-min difference > than 75 uV (green vertical line). Next, as the last sample of this window contains the first >75 uV peak-to-peak on the original EMG trace, 5 ms have been added to produce the RT for the electrode (black vertical line). Finally, one RT should be detected in at least 12 out of 24 bipolar derivations; the shortest has been considered the final RT to be sent to the presentation PC as feedback.

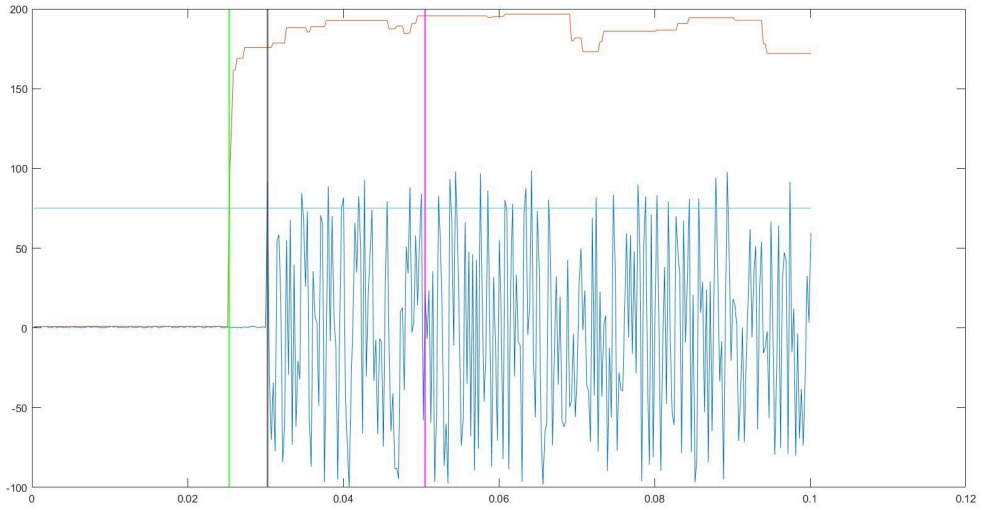

Figure S1 *Example of synthetic EMG time series with hypothetical contraction onset (reaction time) at 30 ms. The blue time series represents an EMG trace. The orange horizontal time series is the result of the sliding movmax - movmin centered at each sample (radius -/+5 ms around each sample, "smoothed max-min EMG"). The magenta vertical line is the sample at which each sample of the orange line around the +/- 25 ms has a movmax - movmin greater than 75 uV (cyan horizontal line). The green vertical line represents the magenta sample - 25 ms, accounting for the movsum function centering used to estimate the magenta sample point. That is the first of the 10 ms sliding window of the 50 consecutive ones where the signal was greater than 75 uV. However, this window is shifted backward by 5 ms as its last sample contains the first sampling point greater than > 75. Thus, 5 ms were added, resulting in the sample identified by the black line, which represents the actual reaction time sample (correctly estimated as 30.8 ms).*

Root-mean square calculation
(MultiEMG and Cross-classification analyses)

Within the MultiEMG analysis we calculated the root-mean square metric (RMS) of the EMG trace for each trial and each virtual channel as a measure of contraction magnitude. We chose to calculate the RMS in the time interval starting from the earliest detected reaction time onset among all the bipolar derivations, until the following 200 ms. To detect the reaction time onset we implemented a similar algorithm as the one employed for the reaction times calculation during the experimental session (Figure S1). As it was theoretically possible that valid trials showed pre-Go-signal contraction but no reaction time onset in the post-Go-signal window, we decided to calculate the reaction time starting from -300 ms from the

Go-signal, and, in case the criteria for detecting a reaction time were not satisfied in the considered time window, we calculated the RMS from 200 to 400 ms from the Go-Signal.

Sanity Check: MultiEMG Classification

Single subject's total accuracy: The total classification accuracy was above 44.4% ($p<0.001$, Figure S2) for each participant.

Group total accuracy: The average total accuracy was 98.6% +/- 0.01% sd, significantly higher than the group chance level of 33.3% ($W = 0$, $p < 0.001$, Figure S2).

Classification Bias: Each of the three actions was classified with similar average accuracies across participants (Friedman $Q = 0.66$, $p = 0.71$; average accuracies: "three" = 98.6% +/-1.6% sd, "ok" = 98.8% +/- 1.3% sd, "hh" = 98.4% +/- 1.4% sd).

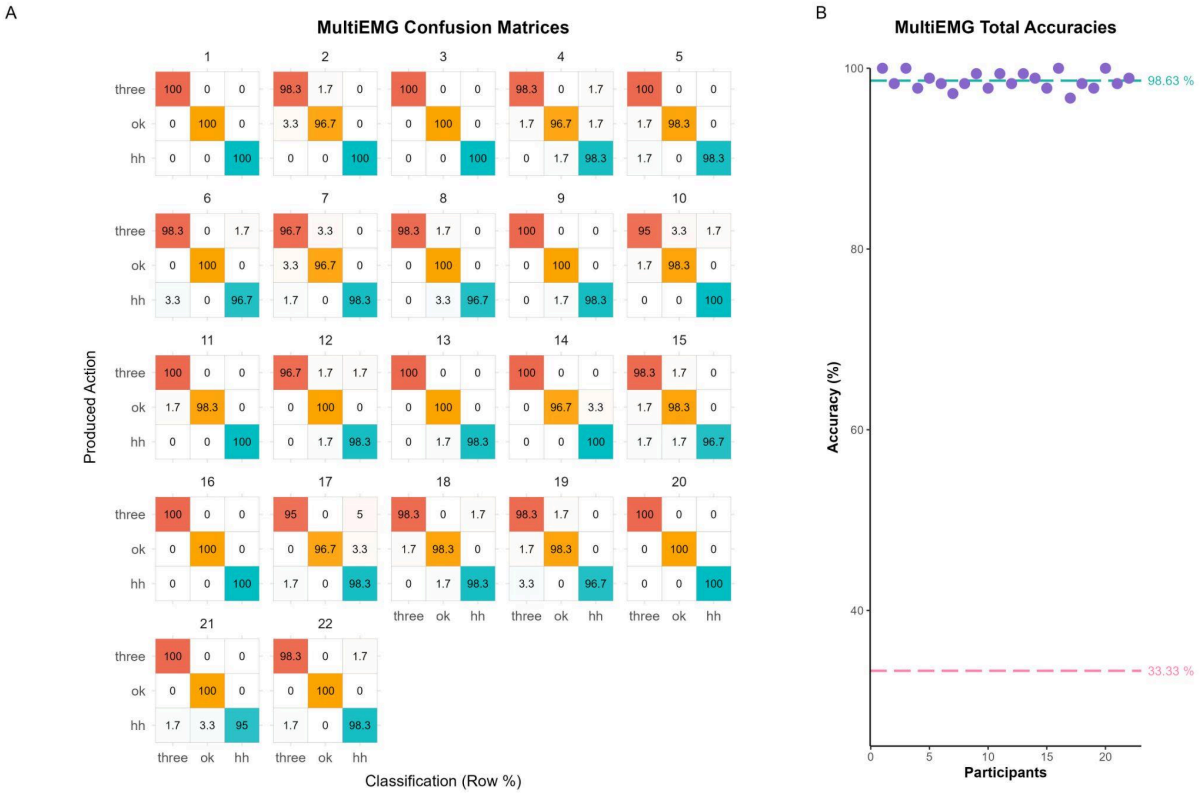

Figure S2. MultiEMG Confusion Matrices and Total Accuracies

***A**. MultiEMG confusion matrices for each participant: within each produced action category (row), cells represent the % of trials predicted to belong to each of the three actions. A perfect classification would highlight 100% on the main diagonal cells and 0% on the remaining cells of each row.*

B. *MultiEMG accuracies: each dot represents the total accuracy of the classification for each participant. The green dotted line represents the average total accuracy, and the pink one the chance level.*

MultiEMG classification showed an almost flawless accuracy in distinguishing between produced actions based on the MultiEMG contraction pattern. Each of the three actions were classified with similar average accuracies across participants.

MultiMEP Classification Accuracies for Electrode Grid Subdivisions

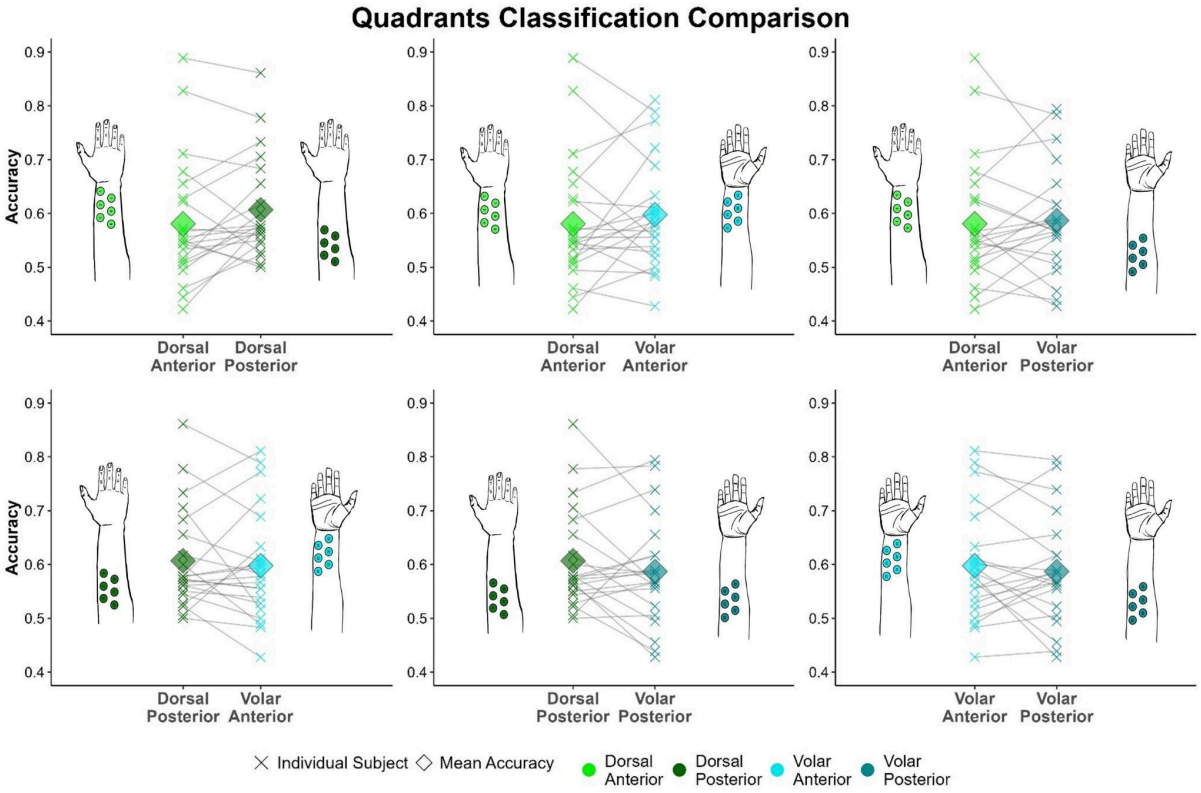

Figure S3. *Pairwise comparisons between Quadrant subdivisions. Crosses represent single participants classifications and diamonds represent mean accuracy across subjects.*

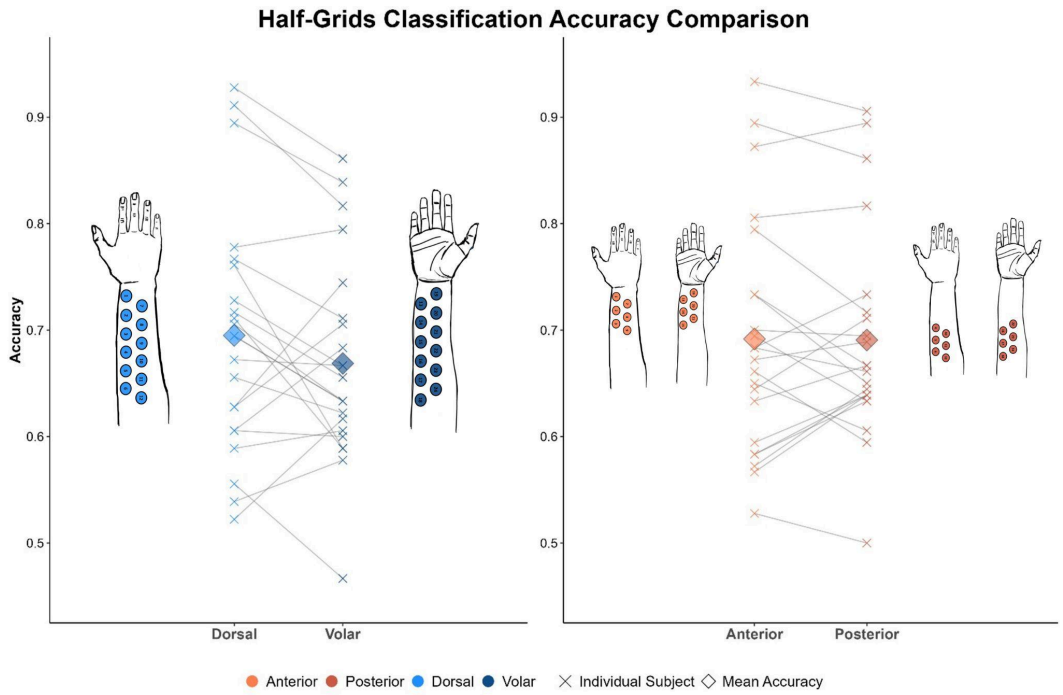

Figure S4. *The left panel shows Half-Grid classification accuracy comparisons between Dorsal and Volar subdivisions; the right panel shows the comparison between Anterior and Posterior subdivisions. Crosses represent single participants classifications and diamonds represent mean accuracy across subjects.*

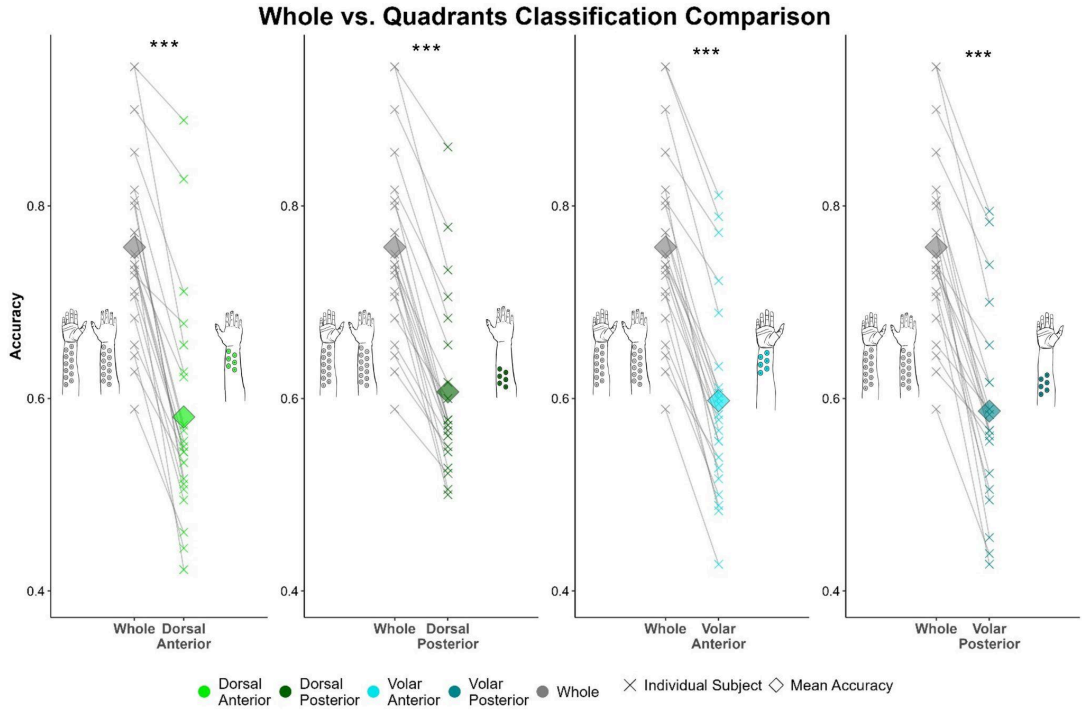

Figure S5. *The figure shows Whole-Grid classification accuracies compared to Quadrant classifications. Crosses represent single participants classifications and diamonds represent mean accuracy across subjects.*

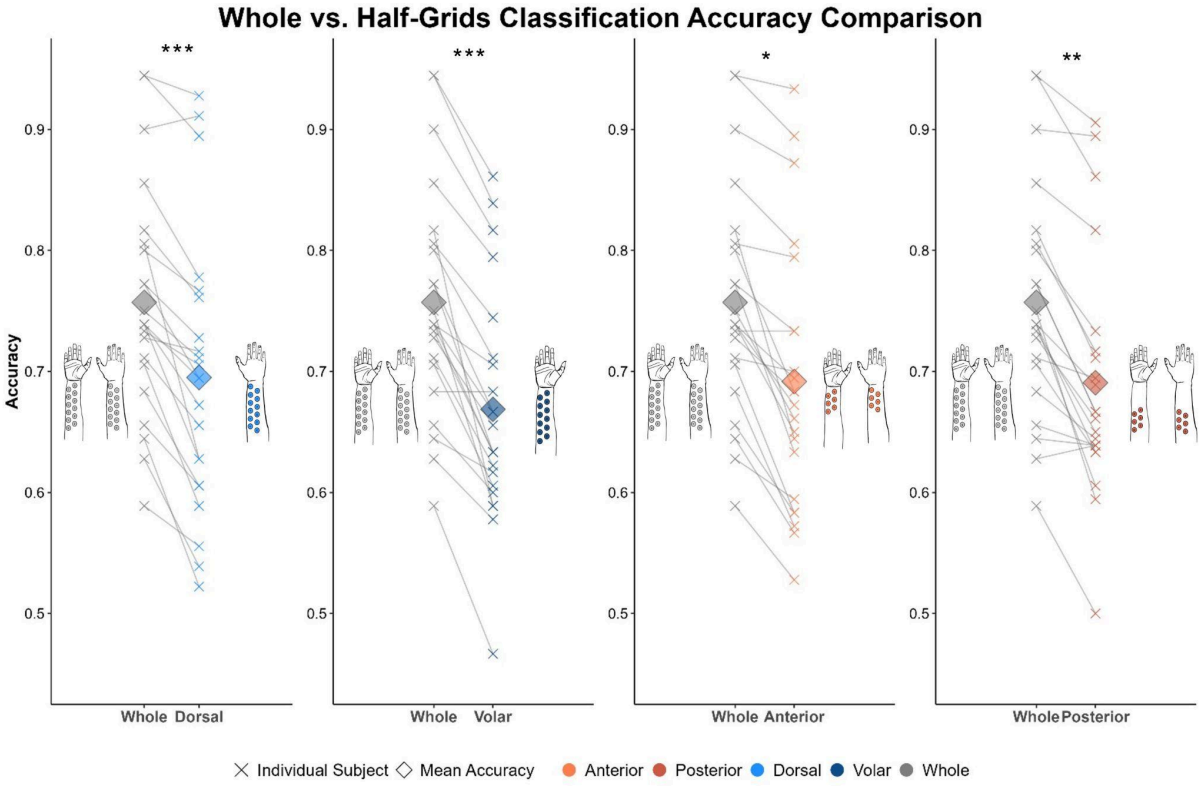

Figure S6. *The figure shows Whole-Grid classification accuracies compared to Half-Grid classifications. Crosses represent single participants classifications and diamonds represent mean accuracy across subjects.*

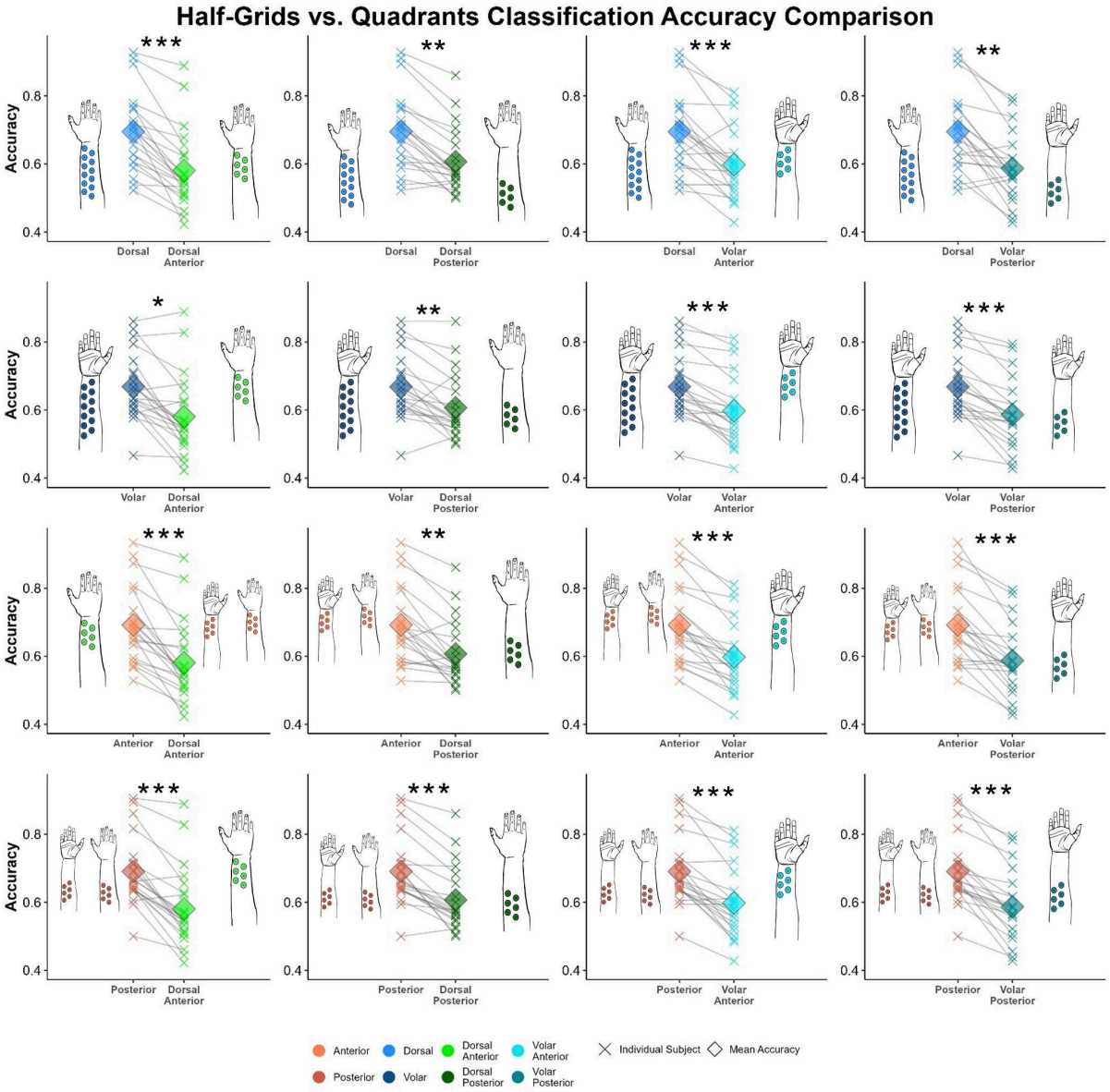

Figure S7. *The figure shows Half-Grid classification accuracies compared to Quadrants classifications. Crosses represent single participants classifications and diamonds represent mean accuracy across subjects.*

REFERENCES

Arai, N., Lu, M.-K., Ugawa, Y., & Ziemann, U. (2012). Effective connectivity between human supplementary motor area and primary motor cortex: a paired-coil TMS study. Experimental Brain Research, 220, 79–87.

Aron, A. R., & Verbruggen, F. (2008). Stop the presses: Dissociating a selective from a global mechanism for stopping. Psychological Science, 19, 1146–1153.

Arora, T., Desai, N., Kirshblum, S., & Chen, R. (2022). Utility of transcranial magnetic stimulation in the assessment of spinal cord injury: Current status and future directions. Frontiers in Rehabilitation Science, 3.

Barchiesi, G., & Cattaneo, L. (2013). Early and late motor responses to action observation. Social Cognitive and Affective Neuroscience, 8, 711–719.

Barchiesi, G., et al. (2020). Head magnetomyography (hMMG): A novel approach to monitor face and whole head muscular activity. Psychophysiology, 57.

Barchiesi, G., et al. (2022). Sharing motor plans while acting jointly: A TMS study. Cortex, 151, 224–239.

Bestmann, S., & Krakauer, J. W. (2015). The uses and interpretations of the motor-evoked potential for understanding behaviour. Experimental Brain Research, 233, 679–689.

Bortoletto, M., et al. (2021). Asymmetric transcallosal conduction delay leads to finer bimanual coordination. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 14, 379–388.

Bundt, C., & Huster, R. J. (2024). Corticospinal excitability reductions during action preparation and action stopping in humans: Different sides of the same inhibitory coin? Neuropsychologia, 195, 108799.

Catmur, C., Walsh, V., & Heyes, C. (2007). Sensorimotor learning configures the human mirror system. Current Biology, 17, 1527–1531.

Cattaneo, L., & Barchiesi, G. (2011). Transcranial magnetic mapping of the short-latency modulations of corticospinal activity from the ipsilateral hemisphere during rest. Frontiers in Neural Circuits, 5.

Cattaneo, L., Caruana, F., Jezzini, A., & Rizzolatti, G. (2009). Representation of goal and movements without overt motor behavior in the human motor cortex: a transcranial magnetic stimulation study. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 29, 11134–8.

Cattaneo, L., Maule, F., Barchiesi, G., & Rizzolatti, G. (2013). The motor system resonates to the distal goal of observed actions: Testing the inverse pliers paradigm in an ecological setting. Experimental Brain Research, 231, 37–49.

Cattaneo, L., et al. (2015). Online repetitive transcranial magnetic stimulation (TMS) to the parietal operculum disrupts haptic memory for grasping: Haptic working memory in the parietal operculum. Human Brain Mapping, 36, 4262–4271.

Cavallo, A., Heyes, C., Becchio, C., Bird, G., & Catmur, C. (2014). Timecourse of mirror and counter-mirror effects measured with transcranial magnetic stimulation. Social Cognitive and Affective Neuroscience, 9, 1082–1088.

Chen, R., et al. (2008). The clinical diagnostic utility of transcranial magnetic stimulation: Report of an IFCN committee. Clinical Neurophysiology, 119, 504–532.

Claffey, M. P., Sheldon, S., Stinear, C. M., Verbruggen, F., & Aron, A. R. (2010). Having a goal to stop action is associated with advance control of specific motor representations. Neuropsychologia, 48, 541–548.

Classen, J., Liepert, J., Wise, S. P., Hallett, M., & Cohen, L. G. (1998). Rapid plasticity of human cortical movement representation induced by practice. Journal of Neurophysiology, 79, 1117–1123.

Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. Journal of Neuroscience Methods, 250, 126–136.

Coxon, J. P., Stinear, C. M., & Byblow, W. D. (2006). Intracortical inhibition during volitional inhibition of prepared action. Journal of Neurophysiology, 95, 3371–3383.

Craighero, L., Zorzi, V., Canto, R., & Franca, M. (2014). Same kinematics but different objects during action observation: Detection times and motor evoked potentials. Visual Cognition, 22, 653–671.

D'Ausilio, A., et al. (2014). Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging. Philosophical Transactions of the Royal Society B: Biological Sciences, 369, 20130418.

Davare, M., Montague, K., Olivier, E., Rothwell, J. C., & Lemon, R. N. (2009). Ventral premotor to primary motor cortical interactions during object-driven grasp in

humans. Cortex, 45, 1050–1057.

Davare, M., Rothwell, J. C., & Lemon, R. N. (2010). Causal connectivity between the human anterior intraparietal area and premotor cortex during grasp. Current Biology, 20, 176–181.

Derosiere, G., Vassiliadis, P., & Duque, J. (2020). Advanced TMS approaches to probe corticospinal excitability during action preparation. NeuroImage, 213, 116746.

Duque, J., & Ivry, R. B. (2009). Role of corticospinal suppression during motor preparation. Cerebral Cortex, 19, 2013–2024.

Duque, J., Labruna, L., Cazares, C., & Ivry, R. B. (2014). Dissociating the influence of response selection and task anticipation on corticospinal suppression during response preparation. Neuropsychologia, 65, 287–296.

Duque, J., Lew, D., Mazzocchio, R., Olivier, E., & Ivry, R. B. (2010). Evidence for two concurrent inhibitory mechanisms during response preparation. Journal of Neuroscience, 30, 3793–3802.

Elsayed, G. F., Lara, A. H., Kaufman, M. T., Churchland, M. M., & Cunningham, J. P. (2016). Reorganization between preparatory and movement population responses in motor cortex. Nature Communications, 7, 13239.

Emadi Andani, M., Tinazzi, M., Corsi, N., & Fiorio, M. (2015). Modulation of inhibitory corticospinal circuits induced by a nocebo procedure in motor performance. PLOS ONE, 10, e0125223.

Fadiga, L., et al. (1998). Corticospinal excitability is specifically modulated by motor imagery a magnetic stimulation study. 37, 147–158.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. Journal of Neurophysiology, 73, 2608–2611.

Farina, D., Mesin, L., Martina, S., & Merletti, R. (2004). Comparison of spatial filter selectivity in surface myoelectric signal detection: Influence of the volume conductor model. Medical and Biological Engineering and Computing, 42, 114–120.

Ferracuti, F., Iarlori, S., Mansour, Z., Monteriù, A., & Porcaro, C. (2022). Comparing between different sets of preprocessing, classifiers, and channels selection techniques to optimise motor imagery pattern classification system from EEG pattern recognition. Brain Sciences, 12, 57.

Ficarella, S. C., & Battelli, L. (2019). Motor preparation for action inhibition: A review of single pulse TMS studies using the Go/NoGo paradigm. Frontiers in Psychology, 10, 340.

Fourkas, A. D., Avenanti, A., Urgesi, C., & Aglioti, S. M. (2006). Corticospinal facilitation during first and third person imagery. Experimental Brain Research, 168, 143–151.

Fukumoto, Y., et al. (2022). Differences in motor imagery strategy change behavioral outcome. Scientific Reports, 12, 13868.

Grosprêtre, S., Ruffino, C., & Lebon, F. (2016). Motor imagery and cortico-spinal excitability: A review. European Journal of Sport Science, 16, 317–324.

Guerrero, F. N., García, P. A., & Spinelli, E. M. (2017). Signal modes for design-oriented analysis of active sEMG spatial filter electrodes. In I. Torres, J. Bustamante, & D. A. Sierra (Eds.), VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th -28th, 2016 (Vol. 60, pp. 504–507). Springer Singapore.

Guidali, G., et al. (2023). Effects of transcranial magnetic stimulation (TMS) current direction and pulse waveform on cortico-cortical connectivity: A registered report TMS-EEG study. European Journal of Neuroscience, 58, 3785–3809.

Hamada, M., Murase, N., Hasan, A., Balaratnam, M., & Rothwell, J. C. (2013). The role of interneuron networks in driving human motor cortical plasticity. Cerebral Cortex, 23, 1593–1605.

Hannah, R., Cavanagh, S. E., Tremblay, S., Simeoni, S., & Rothwell, J. C. (2018). Selective suppression of local interneuron circuits in human motor cortex contributes to movement preparation. The Journal of Neuroscience, 38, 1264–1276.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. Annual Review of Neuroscience, 37, 435–456.

Hebart, M. N., & Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function. NeuroImage, 180, 4–18.

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. Neuron, 45, 201–206.

Jimura, K., & Poldrack, R. A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia,

50, 544–552.

Kaskie, R. E., & Ferrarelli, F. (2018). Investigating the neurobiology of schizophrenia and other major psychiatric disorders with transcranial magnetic stimulation. Schizophrenia Research, 192, 30–38.

Kaufman, M. T., Churchland, M. M., & Shenoy, K. V. (2013). The roles of monkey M1 neuron classes in movement preparation and execution. Journal of Neurophysiology, 110, 817–825.

Kaufman, M. T., Churchland, M. M., Ryu, S. I., & Shenoy, K. V. (2015). Vacillation, indecision and hesitation in moment-by-moment decoding of monkey motor cortex. eLife, 4, e04677.

Keogh, R., Bergmann, J., & Pearson, J. (2020). Cortical excitability controls the strength of mental imagery. eLife, 9, e50232.

Khan, R. A., et al. (2023). A novel framework for classification of two-class motor imagery EEG signals using logistic regression classification algorithm. PLOS ONE, 18, e0276133.

Kothe, C., et al. (2025). The Lab streaming layer for synchronized multimodal recording. Preprint. https://doi.org/10.1101/2024.02.13.580071

Lotte, F., et al. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. Journal of Neural Engineering, 15, 031005.

Majid, D. S. A., Cai, W., George, J. S., Verbruggen, F., & Aron, A. R. (2012). Transcranial magnetic stimulation reveals dissociable mechanisms for global versus selective corticomotor suppression underlying the stopping of action. Cerebral Cortex, 22, 363–371.

Maule, F., Barchiesi, G., Brochier, T., & Cattaneo, L. (2015). Haptic working memory for grasping: the role of the parietal operculum.

McMillan, S., Ivry, R. B., & Byblow, W. D. (2006). Corticomotor excitability during a choice-hand reaction time task. Experimental Brain Research, 172, 230–245.

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.

Meirhaeghe, N., Riehle, A., & Brochier, T. (2023). Parallel movement planning is achieved via an optimal preparatory state in motor cortex. Cell Reports, 42, 112136.

Merletti, R., & Cerone, G. L. (2020). Tutorial. Surface EMG detection, conditioning and pre-processing: Best practices. Journal of Electromyography and Kinesiology, 54, 102440.

Merletti, R., & Muceli, S. (2019). Tutorial. Surface EMG detection in space and time: Best practices. Journal of Electromyography and Kinesiology, 49, 102363.

Messina, I., et al. (2016). Sex-specific automatic responses to infant cries: TMS reveals greater excitability in females than males in motor evoked potentials. Frontiers in Psychology, 6.

Mrachacz-Kersting, N., Stevenson, A. J. T., & Ziemann, U. (2021). Short-interval intracortical inhibition and facilitation targeting upper and lower limb muscles. Scientific Reports, 11, 21993.

Naish, K. R., & Obhi, S. S. (2015). Timing and specificity of early changes in motor excitability during movement observation. Experimental Brain Research, 233, 1867–1874.

Naish, K. R., Houston-Price, C., Bremner, A. J., & Holmes, N. P. (2014). Effects of action observation on corticospinal excitability: Muscle specificity, direction, and timing of the mirror response. Neuropsychologia, 64, 331–348.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience, 2011, 1568–1569.

Padfield, N., Zabalza, J., Zhao, H., Masero, V., & Ren, J. (2019). EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges. Sensors, 19, 1423.

Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Pfurtscheller, G., Neuper, C., Andrew, C., & Edlinger, G. (1997). Foot and hand area mu rhythms. International Journal of Psychophysiology, 26, 121–135.

Rossi, S., Hallett, M., Rossini, P. M., & Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. Clinical Neurophysiology, 120, 2008–2039.

Rossi, S., Pasqualetti, P., Tecchio, F., Pauri, F., & Rossini, P. M. (1998). Corticospinal excitability modulation during mental simulation of wrist movements in human

subjects. Neuroscience Letters, 243, 147–151.

Rossi, S., et al. (2021). Emerging of new bioartificial corticospinal motor synergies using a robotic additional thumb. Scientific Reports, 11, 18487.

Rossi, S., et al. (2021). Safety and recommendations for TMS use in healthy subjects and patient populations, with updates on training, ethical and regulatory issues: Expert guidelines. Clinical Neurophysiology, 132, 269–306.

Ruffino, C., Gaveau, J., Papaxanthis, C., & Lebon, F. (2019). An acute session of motor imagery training induces use-dependent plasticity. Scientific Reports, 9, 20002.

Sharma, R., Kim, M., & Gupta, A. (2022). Motor imagery classification in brain-machine interface with machine learning algorithms: Classical approach to multi-layer perceptron model. Biomedical Signal Processing and Control, 171, 103101.

Stinear, C. M., Byblow, W. D., Steyvers, M., Levin, O., & Swinnen, S. P. (2006). Kinesthetic, but not visual, motor imagery modulates corticomotor excitability. Experimental Brain Research, 168, 157–164.

Tandonnet, C., Garry, M. I., & Summers, J. J. (2011). Selective suppression of the incorrect response implementation in choice behavior assessed by transcranial magnetic stimulation: Suppression of incorrect response implementation. Psychophysiology, 48, 462–469.

Thong, S., et al. (2025). Disinhibition across secondary motor cortical regions during motor sequence learning: A TMS-EEG study. The Journal of Neuroscience, 45.

Ubaldi, S., Barchiesi, G., & Cattaneo, L. (2015). Bottom-up and top-down visuomotor responses to action observation. Cerebral Cortex, 25, 1032–1041.

Urgesi, C., Candidi, M., Fabbro, F., Romani, M., & Aglioti, S. M. (2006). Motor facilitation during action observation: topographic mapping of the target muscle and influence of the onlooker’s posture. European Journal of Neuroscience, 23, 2522–2530.

Urgesi, C., et al. (2010). Simulating the future of actions in the human corticospinal system. Cerebral Cortex, 20, 2511–2521.

Van Den Hurk, J., & Op De Beeck, H. P. (2019). Generalization asymmetry in multivariate cross-classification: When representation A generalizes better to representation B than B to A. Preprint. https://doi.org/10.1101/592410

Vucic, S., et al. (2023). Clinical diagnostic utility of transcranial magnetic stimulation in neurological disorders. Updated report of an IFCN committee. Clinical Neurophysiology, 150, 131–175.

Weaverdyck, M. E., Lieberman, M. D., & Parkinson, C. (2020). Tools of the trade multivoxel pattern analysis in fMRI: a practical introduction for social and affective neuroscientists. Social Cognitive and Affective Neuroscience, 15, 487–509.

World Medical Association. (2013). World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA, 310, 2191–2194.

Xygonakis, I., Athanasiou, A., Pandria, N., Kugiumtzis, D., & Bamidis, P. D. (2018). Decoding motor imagery through common spatial pattern filters at the EEG source space. Computational Intelligence and Neuroscience, 2018, 1–10.

Zazio, A., Barchiesi, G., Ferrari, C., & Bortoletto, M. (2022). M1-P15 as a cortical marker for transcallosal inhibition: A preregistered TMS-EEG study. Frontiers in Human Neuroscience,

Chapter 2

Unlocking Motor Preparation With Multidimensional Motor Evoked Potentials[3]

1.INTRODUCTION

A multitude of overt and covert motor processes characterise human actions. In the previous chapter, I explored a modified motor imagery task to test a new method for the study of covert motor processing with TMS and MEPs. Here, I would like to introduce a new work centered on another covert process, motor preparation. Motor preparation is defined as the change in the state of an organism that improves its ability to adapt to potential environmental changes, thereby ensuring that an appropriate response can be executed effectively (Bestmann and Duque, 2016; Neige et al., 2018; Wong et al., 2015; Bundt and Huster 2024). To reach the optimal state for action production, changes occur within the sensorimotor system that are typically not detectable from peripheral recordings, i.e., EMG or kinematics recording might not provide any information related to the state optimization processes occurring within the motor system. To this aim a fruitful approach might involve the employment of neuroimaging instrumentation such as magneto/electroencephalography (Di Russo et al., 2017) or functional magnetic resonance (Cunnington et al., 2002) which can detect modification in brain activity while actions are planned. Another popular way of studying action preparation involves the use of TMS on M1 delivered between a cue that specifies parameters of the action to be produced and an imperative or “Go” signal, commanding participants to execute the action they were preparing (van Elswijk et al., 2008).

Different paradigms have been employed to study different characteristics of motor preparation: motor preparation can be coupled with action selection in choice reaction times paradigms (Soto et al., 2010), can be implemented in single-choice paradigms where actions to be executed are pre-specified at each trial before the Go-signal, or in Go-NoGo paradigms where the imperative signal specifies whether

[3] Manuscript in preparation

the action has to be executed or not (Leocani et al., 2000). TMS pulses can be delivered right before the Imperative signal, at baseline time (before the cue) or in between the cue and the imperative signal to test the temporal evolution of the cortical excitability between the cue and the Imperative signal (Duque et al., 2010; 2014; 2017). The Imperative signal can be predictable in time, or rather unpredictable, usually employing long (at least 2 seconds).

Despite these differences between experimental settings and designs, the limitations characterizing TMS-MEP motor preparation experiments are the same as those illustrated in the previous chapter as these always employ a 1m1a approach: MEP amplitude from single muscles are treated as proxies of instantiated motor representations, tested actions are typically single joint movements, averages are needed to increase the signal-to-noise ratio and if two muscles are employed to test the specificity of the representations instantiated, muscles are treated as levels of an independent variable.

In light of our results obtained with the MultiMEP method in the motor imagery task, which proved capable of classifying three different ecological actions, we explored whether the technique could be extended to other covert motor processes such as motor preparation and by increasing the number of actions and their similarity in order to explore the limits of the technique.

To this aim we presented participants with drawings representing hands in different counting positions, from one to five. Participants had to prepare to execute the displayed action when a sound was presented, trying to provide a response as fast as possible. A single TMS pulse was delivered when the sound was played, thus testing the corticospinal state during preparation but before action execution.

2. METHODS

2.1 Participants

We tested twenty-five right handed participants (12 females, mean age: 24.8). All of them were previously screened for neurological and psychiatric conditions with TMS exclusion criteria questionnaire before starting the procedure. Twenty-four participants were included in the final analysis: one has been excluded due to difficulty relaxing the target muscles. The ethical committee of the University of Milan

approved the study (“MultiMEP: Potenziali Evocati Motori Multidimensionali”) and it was conducted in accordance with the revised Helsinki Declaration (World medical association, 2013) with written informed consent obtained from all participants. Since the MultiMEP method is new and has been used only in our previous work, we decided to recruit a sample of participants corresponding to the sample size usedin the preceding experiment (25 participants, Chapter 1, paragraph 2.1 Participants).

2.2 Procedure

The procedure consisted of the following phases: electrode grid montage, hotspot-intensity hunting procedure, instruction, training session, experimental blocks. It lasted approximately 2.5-3 hours.

2.2.1 Electrodes grid montage

After we ensured that all exclusion criteria were met, the procedure started with the scrubbing phase. The skin of the right hand and forearm was scrubbed using a scrubbing paste (Merletti & Cerone, 2020). The electromyographic activity was recorded using 32 passive sintered Ag/AgCl ring electrodes (model: B10, length: 150 cm, EasyCap GmbH, Germany) arranged in four strips of six electrodes each, spaced 2.5 cm apart. Differently from the first MultiMEP study, the four strips were all placed on the dorsal forearm; the first strip was aligned along the radial side (electrodes from 1 to 6), the second strip (electrodes from 7 to 12)was placed in the space between the first and the second electrodes of the first strip moved 2.5 cm towards the ulnar side. The third strip (electrodes from 13 to 18) was placed aligned with the first and the fourth strip (electrodes from 19 to 24) with the second, each 2.5 cm apart covering the dorsal part of the forearm. On the hand we recorded electromyography activity by means of Ag/AgCl disposable electrodes to ensure that they fit the shape of the hand better. A total of eight electrodes were used: electrodes from 25 to 27 were placed on the palm of the hand aligned with the thumb; electrodes from 28 to 30 on the dorsum, aligned with the thumb and eventually, 31 and 32 were aligned with the little finger on the side on the hand (Figure 1). Disposable ground and reference electrodes were placed on the dorsal side of the forearm with adhesive tape. Conductive gel was then applied to the inside of each ring and under the disposable electrodes of the hand using a syringe to ensure conductance levels were below 40 kΩ.

Monopolar EMG recordings were achieved by connecting the electrodes to 32 pins on a BePlusPro Advanced amplifier (EBNeuro, Italy), with data sampled at 4096 Hz. A 5 Hz high-pass filter was applied solely for visualisation purposes and did not affect the recorded data.

As in the first MultiMEP experiment (Chapter 1; paragraph Electrodes Grid Montage) for visualization purposes have been computed bipolar derivations of the electrodes by computing the difference between pairs of monopolar recordings (Farina et al., 2004, Guerrero et al., 2007) along the longitudinal dimension of the forearm for each of the four strips (1-2 to 5-6, 7-8 to 11-12, 13-14 to 17-18, 19-20 to 23-24, 25-26 to 31-32) and eventually between 6-32 and 24-25. We employed our “MultiMEP Viewer” (Chapter 1; Supplementary materials: Muscular contraction and MEPs amplitude detection on the “MultiMEP Viewer”) and the “multi-epochs window”, where MEPs calculated on the 27 bipolar derivations, were displayed between -150 to 600 ms from the TMS trigger, and whose peak-to-peak amplitude was automatically calculated between 15 ms to 65 ms from the TMS trigger.

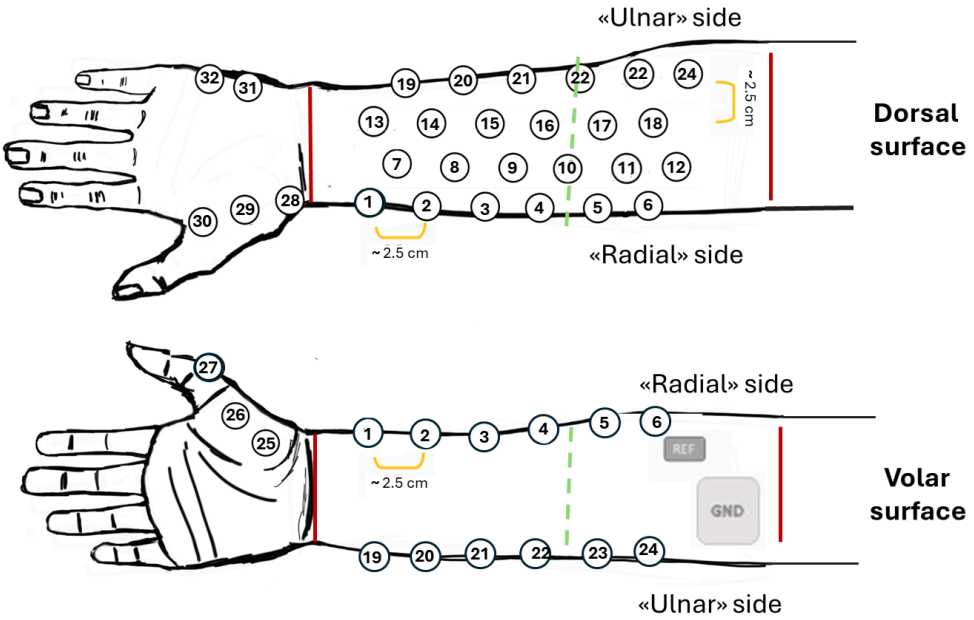

Figure 1. Electrodes grid montage. *Four strips, each containing six passive sintered ring electrodes (24 electrodes; typically used for electroencephalographic recordings) placed on the dorsal forearm spaced 2.5 cm apart and eight Ag/AgCl disposable electrodes placed on the hand (32 electrodes total) were used for electromyographic recordings. Ground and reference electrodes were disposable*

electrodes and were placed on the proximal area of the forearm.

2.2.2 TMS

TMS pulses were administered via a 70-mm butterfly coil, connected to STM9000 stimulator (inducing anterior-to-posterior currents) and guided by the NETBRAIN9000 neuronavigation system (EBNeuro, Italy). The participants' cranial positions and the coil were coregistered to a standardized three-dimensional Montreal Neurological Institute (MNI) brain mesh. Participants sat on a chair with no armrests, with their head placed on a chin rest and their right arm relaxed along their body with the hand pointing downward. This position was the position requested during the experiment in order to minimize the likelihood of pre-TMS contractions. Participants were also allowed to sit with their back on the back of the chair, if this position produced less muscular contractions.

As in the first MultiMEP experiment, we found the stimulation spot (the MultiMEP hotspot) on the scalp that allowed, at rest, to produce MEPs in the recorded bipolar derivations of at least 50 uV in amplitude on at least 21 out of 27 bipolar derivations (3/4 of the visualized bipolar derivations) in 5 consecutive trials. The intensity estimated with this procedure was then used throughout the experiment. In order to detect the amplitude of the MEPs detected on the 32 bipolar derivations, the Viewer was programmed to calculate the peak-to-peak amplitude of a time window comprised between 15 and 60 ms from the TMS trigger onset, which was displayed near the bipolar channel name; if the peak-to-peak amplitude detected was greater than 50μV the amplitude was displayed in green, otherwise in gray. After the threshold hunting procedure was concluded, the experimental session started.

2.3 Stimuli, task and trial structure

The stimuli consisted of drawing of a right hand representing the most used 1-5 European finger counting configuration: number one, (hitchhiking), executed with the thumb extended and the remaining fingers flexed; number two, executed with index and middle finger extended and the remaining flexed; number three, with thumb, index and middle finger extended and the remaining flexed; number four, with all fingers extended except for the thumb and number five with all the fingers extended (Figure 2a).

The experimental session consisted of 4 experimental blocks for a total of 240 trials

(48) for each action.

Participants were instructed to prepare to execute the action presented on the screen and execute as fast as possible as soon as they saw the imperative signal “Perform” (in italian “Esegui”), and to maintain it until the stop signal. After that they were asked to completely relax. In order to ensure that participants actually prepared the action to be performed, they underwent a training session before the experimental session, to minimize their reaction times (see “Training session” paragraph).

Each trial of the experimental block started with the preparatory cue which was the image representing one of the five actions.The image was presented on the screen for a variable time between 2s and 3s. To ensure consistency across actions to be prepared, the delay between preparatory cue and imperative signal has been set at eight counterbalanced delays between actions. The imperative signal “Esegui” with a high pitched sound represented the Go signal; participants executed the action until the Stop signal, represented by a low-pitched auditory sound lasting 500ms (with a pitch of 440 Hz). The Go signal corresponded also to the TMS delivery pulse on the scalp.

Eventually, participants were presented with a visual feedback on the screen: if the action was executed correctly, under 250 ms from the Go-signal, and without contraction in the 50 ms before TMS pulse, they would see their reaction time. In any other cases, they could see the following phrases in red: reaction time (over their RT threshold), no response, already contracted before sound. Trial counting did not decrease because it would be rescheduled. Feedback screen appeared on the screen for 1s. It was followed by a 1-second blank screen, during which participants were required to relax (Figure 2b).

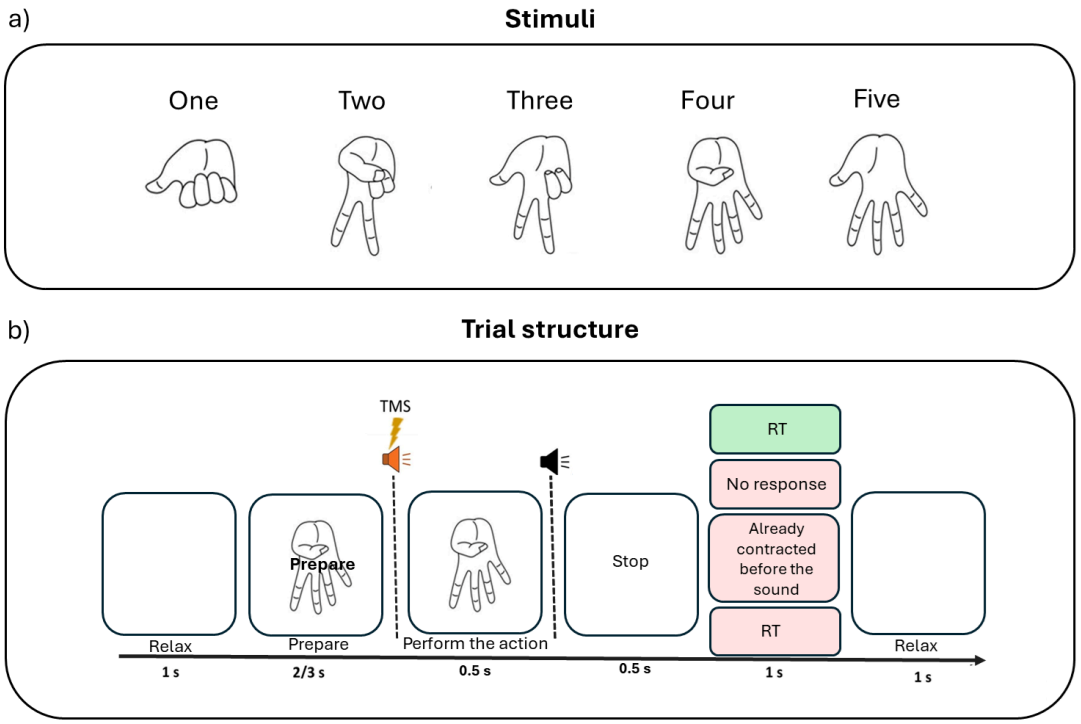

Figure 2. Stimuli and Trial Structure Stimuli. *a) Schematic representation of the five stimuli employed in the task (European gestures for numbers): one, two, three, four, five. The downward orientation of the drawings matched the participants' hand position. b) Schematic representation on the trial structure with the duration of each step: preparatory phase where participants were required to stay relaxed lasted 2/3 s, Go signal with TMS pulse were participants were asked to execute the action seen on the screen for 0.5 s until the Stop signal (0.5 s) followed by the feedback which remained on the screen for 1s.*

2.4 Training session

In order to induce participants to actually prepare the actions between the display of the drawings and the Go-signal, they underwent a training session, similar to the experimental session that followed, but without TMS pulses delivered and with a changing RT threshold. In the training session participants completed a practice block consisting of 40 trials; each action has been shown eight times. The initial RT threshold was set at 250 ms. After the fifth trial, this threshold was adjusted to correspond to the trimmed mean of the previous five reaction times. If RTs were above threshold the RT was displayed in red, otherwise, if faster, in green.

While the optimal strategy would have been to set the RTs threshold to a value close to individual RTs, in pilots experiments a silent period-like EMG after the MEP was

detected when delivering TMS during the experimental session, which would make it almost impossible to react within the individual RT estimated during the training session. We then decided to keep the RT threshold for the experimental session to 250 ms; the training session lasted approximately 5 minutes.

3. ANALYSES

3.1 MultiMEP Data processing

To analyse data we employed MATLAB with Fieldtrip toolbox (Oostenveld et al., 2011) and custom Python scripts, including Scikit-learn libraries for classification algorithms (Pedregosa et al., 2011). Figures have been created with R, Version 4.3.2 (R Core Team, 2022), and the package ggplot2, Version 3.4.4 (Wickham, 2016).

For each participant's dataset has been epoched 300ms before and 200ms after TMS trigger. Trials with pre-TMS contractions were eliminated to ensure that exactly 48 trials for each action would be analysed. Electromyography signal was preprocessed using 4th order Butterworth 10 Hz high-pass and 500 Hz low-pass filters. Downsampling (1000 Hz) has been applied to the filtered epochs to reduce the computational load.

The "massive bipolarization" (Chapter 1; paragraph 3.1.1 Data Preprocessing) has been applied computing all possible bipolar derivations, subtracting the signal of one electrode from another every unique pair of 32 channels, resulting in 496 bipolar derivations (virtual channels). For each of them we calculated the peak-to-peak amplitude of the MEPs within 15 and 60 ms after the TMS trigger. Eventually, we computed the MultiMEP matrix with 240 trials (row) × 496 MEPs virtual channels (columns) (Figure 3).

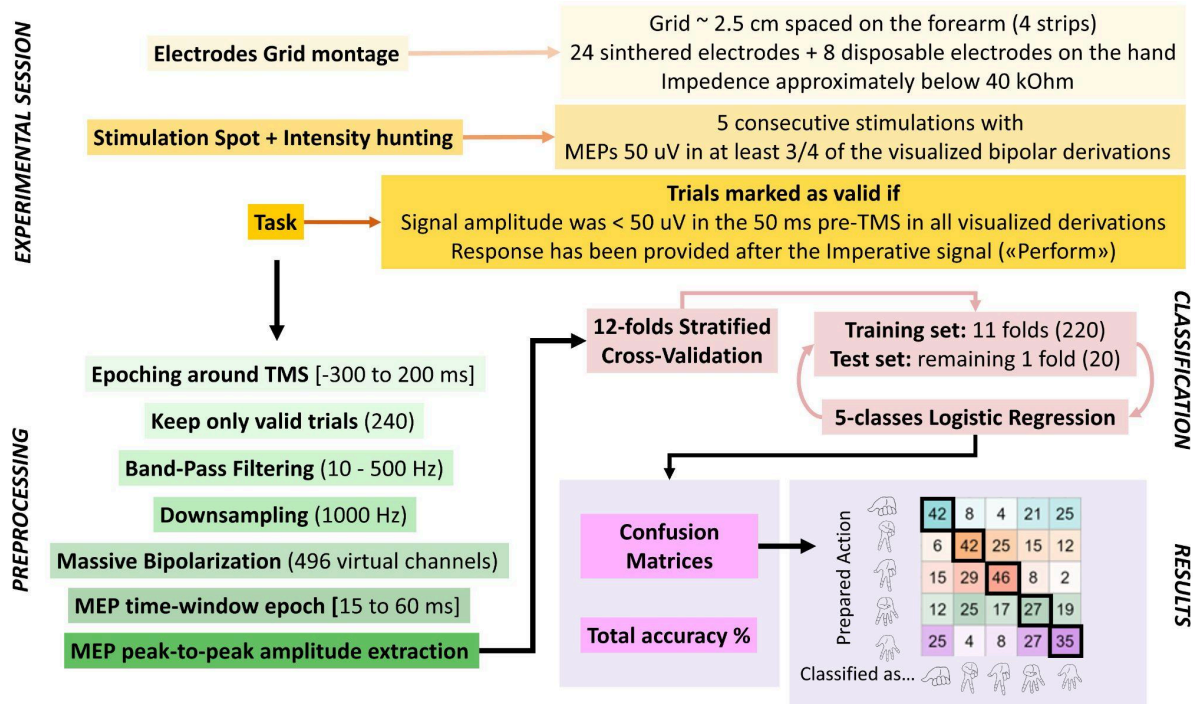

Figure 3. Main experimental steps: montage (shades of yellow), preprocessing (shades of green), classification (shades of pink) and results (shades of purple).

3.2 MultiIMEP classification

The classification followed the same pipeline applied to our first MultiMEP experiment (Chapter 1; paragraph 3.1.2 Classification): a 5-class logistic regression within 12-fold stratified cross-validation has been employed: we have five classes of action so the 12folds in which the dataset is randomly split have the equal proportions of trials belonging to each of them. The logistic regression model is trained on 11 folds and tested on the 12th, the same procedure is repeated for each fold. The dataset used as training-set has been z-scored at the beginning of each cross-validation cycle as well as the test-set. At the end of the process we computed a confusion matrix for each participant: for each prepared action the classifier prediction percentage has been calculated

Single subject total accuracy: We tested if MultiMEP pattern contained sufficient information to classify prepared actions above chance at the single-subject level. As in our previous work we followed the approach of Combrisson & Jerbi (2015), our chance level, with five classes is 20%; the probability of achieving 28.3%, 26.2% or 24.2% accuracy in our experiment (5 classes, 48 trials per class) corresponded to p-value < 0.001, < 0.01, and < 0.05 respectively.

Group total accuracy: At the sample level, we tested if total classification accuracies were, on average, above the theoretical chance level by performing a non-parametric Wilcoxon test on total accuracies against 20%.

3.3 Exploratory analysis: MultiEMG data processing

We computed the analysis on the contraction pattern, obtained from the execution of the actions in each trial. We searched for contractions from 60ms (after the end of the MEP) and analysed 200ms after the start of it. Eventually, the root mean square has been computed on these data (see Chapter 1; paragraph 3.2.1 MultiEMG Data Preprocessing) and a 240 trials x 496 matrix was computed along with the labels for each trial.

3.3.1 MultiEMG classification

If we had not been able to classify voluntary contraction pattern (MultiEMG), we would not have been able to classify the MultiMEP pattern. For this reason we conduct a sanity check and we tested if the executed actions were discriminable from MultiEMG patterns, applying the MultiMEP pipeline analysis to the MultiEMG data (Chapter 1; paragraph 3.2.2 Sanity Check: MultiEMG Classification).

3.3.2 Cross-Classification between contraction and preparatory patterns

To correctly compare MultiMEP and MultiEMG data avoiding TMS bias, we computed data standardization and z-scored both datasets along the virtual channel dimension.

4. RESULTS

4.1 MultiMEP classification

Single subject's total accuracy: We tested how well single participants' total accuracy classification departed from an empirical chance level: following Combrisson & Jerbi (2015), the likelihood that a 28.3% classification accuracy is obtained by chance corresponds to a p-value < 0.001, for a 26.2% accuracy to p< 0.01, and for a 24.2% accuracy to p< 0.05. Total accuracy was above 24.2% for 23 out of 24 participants

(p<0.05), above 26.2% for 23 out of 24 (p<0.01) and above 28.3% for for 23 participants out of 24(p<0.001).

Group total accuracy: We calculated if the average total accuracy among participants was higher than the theoretical chance level of 20%. The average total accuracy was 41.26% +/- 9% sd, significantly higher than the group chance level of 20% (W = 0, p < 0.001, Figure 4).

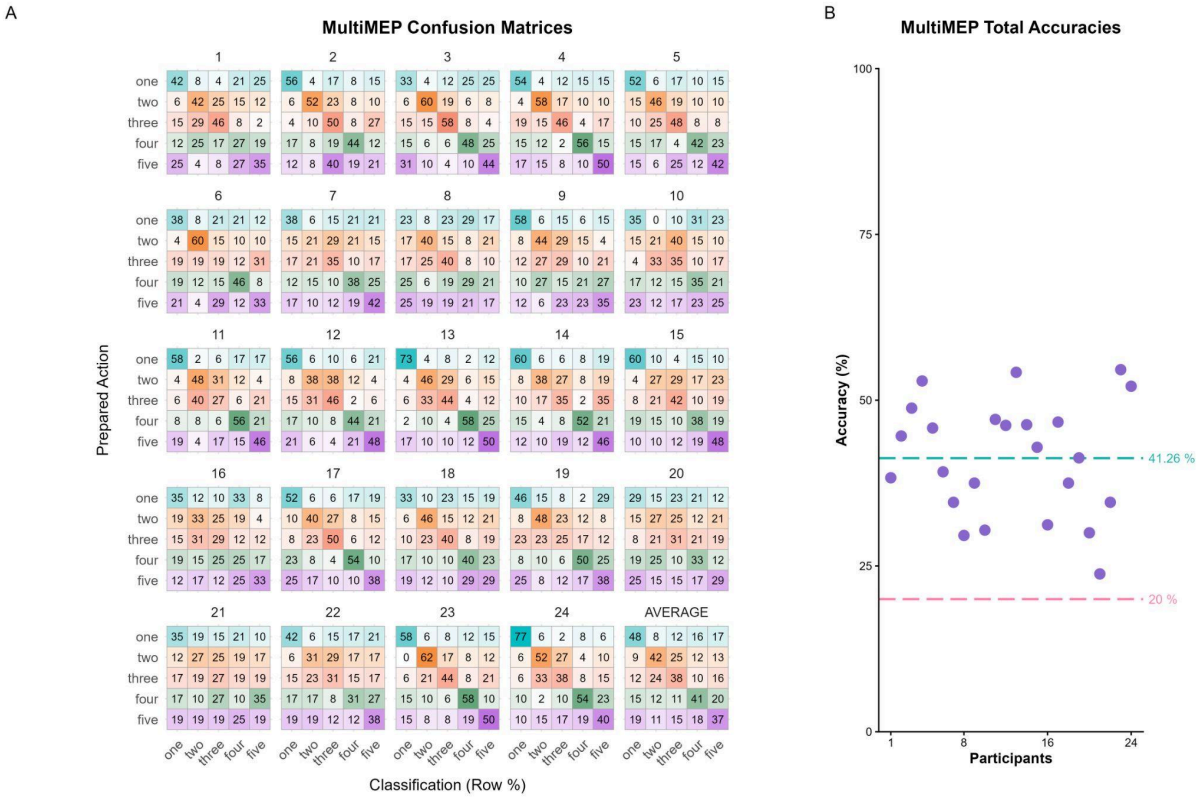

Figure 4. MultiMEP Confusion Matrices and Total Accuracies

A*. MultiMEP confusion matrices from each participant: within each prepared action category ("one" in turquoise, "two" in orange, "three" in coral, "four" in seagreen and "five" in violet), cells represent the % of trials predicted to belong to each of the five actions ("one", "two", "three", "four" and "five"). A perfect classification would highlight 100% on the main diagonal cells and 0% on the remaining cells of each row. Darker colors indicate better classification.****B.*** *MultiMEP accuracies: each dot represents the classification total accuracy for each participant. The green dotted line represents the average total accuracy, and the pink one represents the theoretical chance level.*

4.2 Exploratory analysis: MultiEMG classification

Single subject's total accuracy: We repeated single subjects' test for MultiEMG, calculating total classification accuracy for each of them (Combrisson & Jerbi 2015). Total accuracy was above 24.2% for all 24 participants (p<0.05), above 26.2% for 24 participants (p<0.01) and above 28.3% for 24 participants (p<0.001).

Group total accuracy: The average total classification accuracy for MultiEMG is 94.90% +/-5%, significantly higher than the group chance level of 20% (W = 0, p < 0.001, Figure 5).

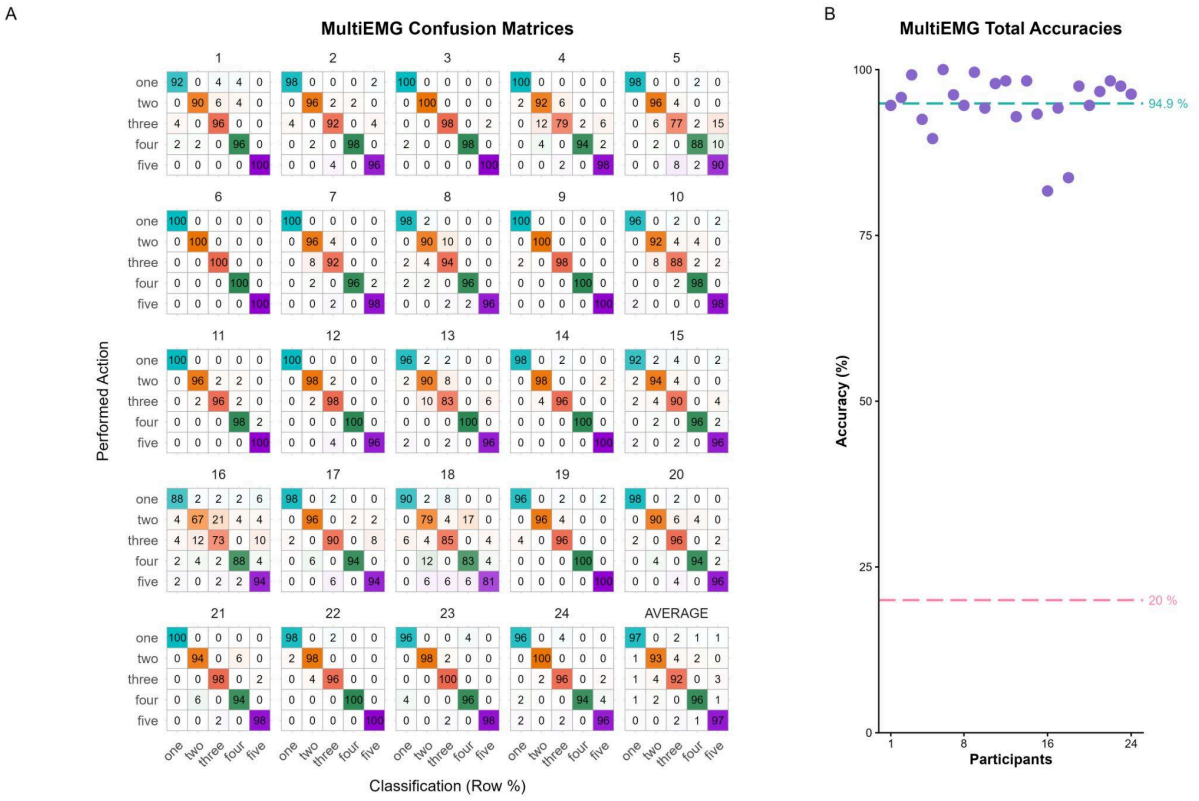

Figure 5. MultiEMG Confusion Matrices and Total Accuracies
A*. MultiEMG confusion matrices from each participant: within each executed action category ("one" in turquoise, "two" in orange, "three" in coral, "four" in seagreen and "five" in violet), cells represent the % of trials predicted to belong to each of the five actions ("one", "two", "three", "four" and "five"). A perfect classification would highlight 100% on the main diagonal cells and 0% on the remaining cells of each row. Darker colors indicate better classification.* ***B.*** *MultiEMG accuracies: each dot represents the classification total accuracy for each participant. The green dotted line represents the average total accuracy, and the pink one represents the theoretical chance level.*

4.2.1 MultiMEP-to- MultiEMG

The MultiMEP-to-MultiEMG classification was produced by training a classifier model on the MultiMEP data and then predicting MultiEMG prepared actions based on the hyperplanes estimated from the MultiMEP patterns.

Single subject's total accuracy: The total classification accuracy was above 28.3% (p<0.001) for 11 out of 24 participants, above 26.2% (p<0.01) for 13 out of 24 participants and above 24.2% (p<0.05) for 15 participants out of 24.

Group total accuracy: The average total accuracy was 26.77% +/- 11% sd, higher than group chance level of 20% (W = 59.5, p = 0.009).

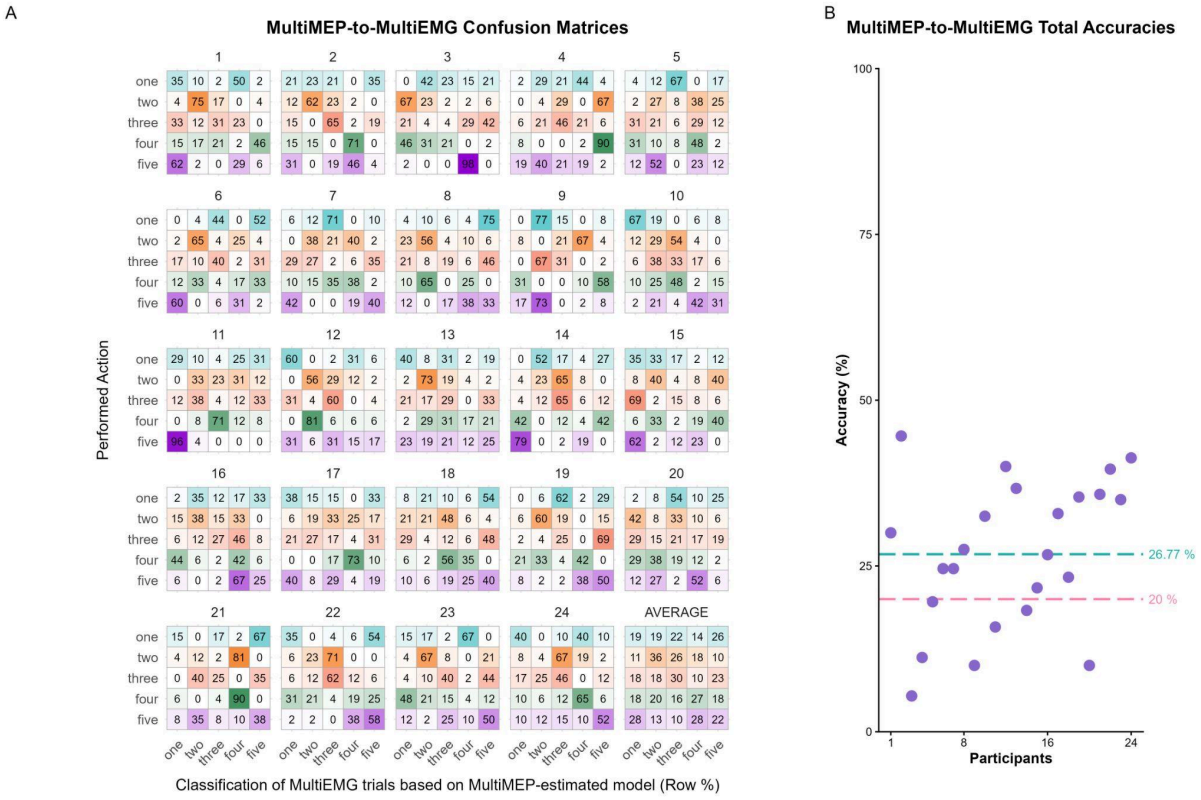

Figure 6. MultiMEP-to-MultiEMG Confusion Matrices and Total Accuracies

A*. MultiMEP-to-MultiEMG confusion matrices from each participant: rows represent executed action ("one" in turquoise, "two" in orange, "three" in coral, "four" in seagreen and "five" in violet). The cells represent the % of trials predicted for the five action categories but, differently from previous analysis, the prediction model has been estimated using MultiMEP patterns and tested on MultiEMG contraction patterns.* ***B.*** *MultiMEP-to-MultiEMG accuracies: each dot represents the classification total accuracy for each participant. The green dotted line represents the average total accuracy, and the pink one represents the theoretical chance level.*

4.2.2 MultiEMG-to-MultiMEP

The MultiEMG-to-MultiMEP classification was produced by training a classifier model on the MultiEMG data and then predicting MultiMEP prepared actions based on the hyperplanes estimated from the MultiEMG contraction patterns.

Single subject's total accuracy: The total classification accuracy was above 28.3% ($p<0.001$) for 5 out of 24 participants, above 26.2% ($p<0.01$) for 5 out of 24 participants and above 24.2% ($p<0.05$) for 9 participants out of 24.

Group total accuracy: The average total accuracy was 24.20% +/- 4% sd, higher than group chance level of 20% ($W = 11$, $p< 0.001$).

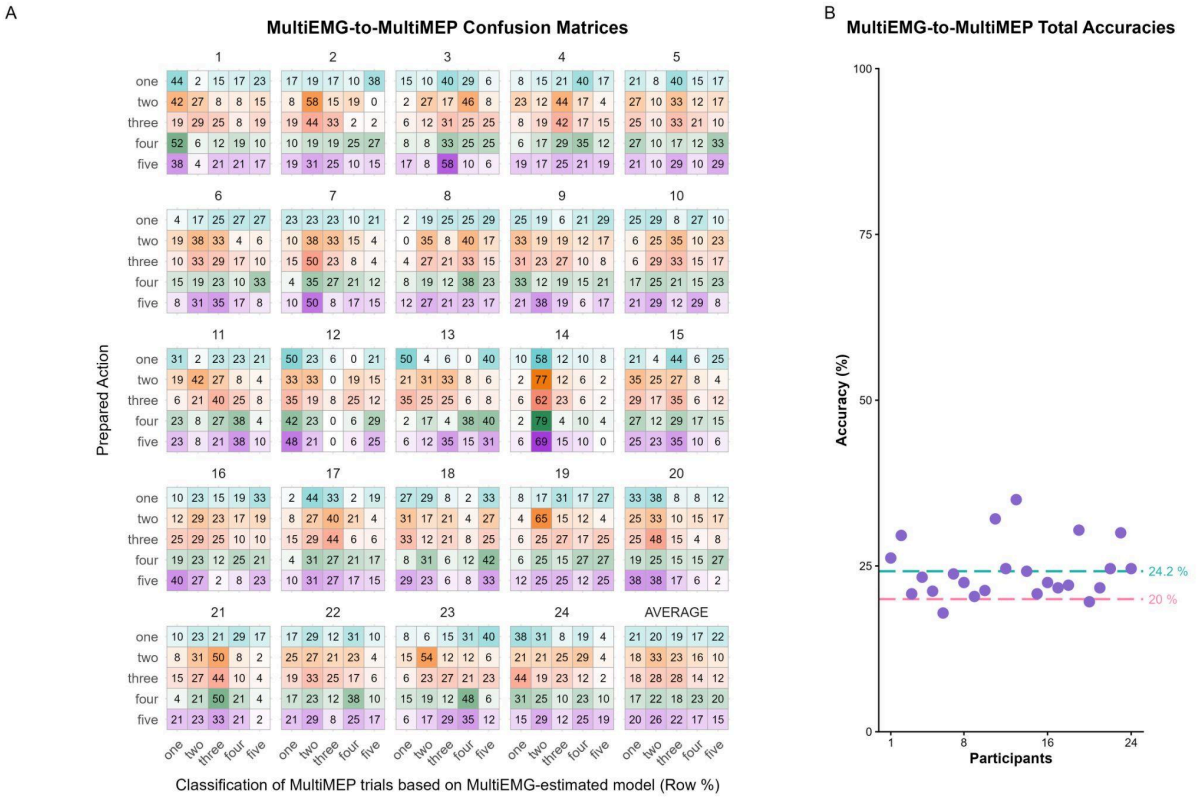

Figure 7. MultiEMG-to-MultiMEP Confusion Matrices and Total Accuracies
A. MultiEMG-to-MultiMEP confusion matrices from each participant: rows represent executed action ("one" in turquoise, "two" in orange, "three" in coral, "four" in seagreen and "five" in violet). The cells represent the % of trials predicted for the five action categories but, differently from previous analysis, the prediction model has been estimated using MultiEMG patterns and tested on MultiMEP patterns. **B.** MultiEMG-to-MultiMEP accuracies: each dot represents the classification total accuracy for each participant. The green dotted line represents the average total accuracy, and the pink one represents the theoretical chance level.

5. DISCUSSION

The classification analysis employed in our previous study allows us to distinguish three different ecological actions from the MultiMEP pattern during an imagery task. Here we tested the decoding approach to the MultiMEP method on a different covert

motor process, action preparation and increasing the number of actions and the similarities between them.

As for the motor imagery experiment, we conducted an exploratory cross-classification analysis trying to predict prepared action labels based on the contraction patterns (MultiMEP to MultiEMG) and vice versa.

Prepared actions have been classified on average with a 41.25% (chance level = 20%). Single subjects analysis revealed that 23 participants' accuracy was above 28.3% (the probability that classification accuracy is obtained by chance corresponds to $p < 0.001$).

We analysed the MultiEMG classification pattern as a necessary step to proceed with MultiMEP pattern classification and cross-classification between preparation and contraction domains. As expected, MultiEMG classification had very high mean classification accuracy, 94.90%.

The exploratory cross-classification between MultiMEP and MultiEMG patterns revealed a lower classification accuracy in both directions: 26.77% for MultiMEP-to-MultiEMG and 24.20% for MultiEMG-to-MultiMEP. Although at group level cross-classification in both directions is statistically significant, at the single subject level, MultiMEP-to-MultiEMG classification revealed that only for 11 participants out of 24 the total classification accuracy was above 28.3% while in the opposite direction, only for 5 out of 24 participants. The noticeable difference between these cross-classification accuracies and those obtained in the motor imagery task might have been caused by different non-mutually exclusive factors: while unlikely, we cannot exclude that a TMS pulse delivered several milliseconds before voluntary execution affected the pattern of contraction activity in the present experiment. Indeed the trial structure of the imagery task was different with respect to the present one as the participants executed the action several seconds before a TMS pulse. The signal to noise ratio might be different between preparatory and imagery tasks, leading to a lower cross-classification in the present study, due to the fact that we tested five ecological actions (instead of three) very similar to each other, thereby, in the MultiMEP space they are less distinguishable. In addition, in the preparatory task we employed a different electrode grid (four strips placed on the volar side of the forearm and eight electrodes on the hand). Another hypothesis to explain this different result relates to the TMS bias: in order to get rid of the TMS-bias, and to make the the imagery and execution domains comparable, in the

imagery experiment we z-scored both datasets (imagery and contraction) along the virtual-channel dimension and found that the accuracies resulting from the cross-classifications were well above chance (54% and 71%, chance = 33%); the application of the z-score transformation was performed to remove the TMS bias under the assumptions that 1) this was the only difference between the two domains and 2) that TMS-bias would simply add to the action-specific pattern of activity.

Speculating, the fact that in the present experiment a similar transformation produced much lower accuracies and that the MultiMEP-to-MultiEMG cross-classification direction produced lower accuracies than the opposite direction, might indicate that the relationship among actions within the MultiMEP preparation space might be scaled non-linearly with respect to the relationship between actions in a contracted state. This would imply that the state of the corticospinal system is both different across covert motor processes considered to be quite similar, and, in the case of preparation, different from the contraction pattern.

A widely accepted hypothesis is that preparatory activity consists in a form of reduced activity of movement (Cisek, 2006), but recent neural population studies show different evidence, demonstrating that preparatory activity is a structured and non accidental activity orthogonal to contraction. Therefore, preparatory activity might not constitute a sub-threshold version of movement-related activity, but a different pattern that is functional for the following execution phase. Clearly, a direct comparison between these studies and ours should be approached with caution due to several differences, as we did not directly record neural activity in primates where continuous neural data were available (Churchland et al., 2010, Elsayed et al., 2016, Kaufman et al., 2014).

Thus, considering the fairly different outcomes between imagery and preparation, a dedicated study should address, within-subjects, the differences between imagery and preparation processing, specifically investigating whether a) classification outcomes are comparable between the two covert processes, and if so 2) whether a diminished cross-classification accuracy is obtained for preparation compared to motor imagery.

6. CONCLUSIONS

The present results confirm that the MultiMEP decoding approach is a viable strategy for testing covert motor processing related to more than two actions. The

cross-classification results might indicate that preparation does not imply similar motor states compared to imagery or action execution; however, future ad-hoc studies are needed to carefully study these relationships.

REFERENCES

Bestmann, S., & Duque, J. (2016). Transcranial Magnetic Stimulation: Decomposing the Processes Underlying Action Preparation. The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry, 22(4), 392–405. https://doi.org/10.1177/1073858415592594

Bundt, C., & Huster, R. J. (2024). Corticospinal excitability reductions during action preparation and action stopping in humans: Different sides of the same inhibitory coin?. Neuropsychologia, 195, 108799. https://doi.org/10.1016/j.neuropsychologia.2024.108799

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I., & Shenoy, K. V. (2010). Cortical preparatory activity: representation of movement or first cog in a dynamical machine?. Neuron, 68(3), 387–400. https://doi.org/10.1016/j.neuron.2010.09.015

Cisek P. (2006). Integrated neural processes for defining potential actions and deciding between them: a computational model. The Journal of neuroscience : the official journal of the Society for Neuroscience, 26(38), 9761–9770. https://doi.org/10.1523/JNEUROSCI.5605-05.2006

Combrisson, E. & Jerbi, K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J. Neurosci. Methods 250, 126–136 (2015).

Cunnington, R., Windischberger, C., Deecke, L., & Moser, E. (2002). The preparation and execution of self-initiated and externally-triggered movement: a study of

event-related fMRI. NeuroImage, 15(2), 373–385. https://doi.org/10.1006/nimg.2001.0976

Di Russo, F., Berchicci, M., Bozzacchi, C., Perri, R. L., Pitzalis, S., & Spinelli, D. (2017). Beyond the "Bereitschaftspotential": Action preparation behind cognitive functions. Neuroscience and biobehavioral reviews, 78, 57–81. https://doi.org/10.1016/j.neubiorev.2017.04.019

Duque, J., Greenhouse, I., Labruna, L., & Ivry, R. B. (2017). Physiological Markers of Motor Inhibition during Human Behavior. Trends in neurosciences, 40(4), 219–236. https://doi.org/10.1016/j.tins.2017.02.006

Duque, J., Labruna, L., Cazares, C., & Ivry, R. B. (2014). Dissociating the influence of response selection and task anticipation on corticospinal suppression during response preparation. Neuropsychologia, 65, 287-296.

Duque, J., Lew, D., Mazzocchio, R., Olivier, E., & Ivry, R. B. (2010). Evidence for two concurrent inhibitory mechanisms during response preparation. Journal of Neuroscience, 30(10), 3793-3802.

Elsayed, G. F., Lara, A. H., Kaufman, M. T., Churchland, M. M., & Cunningham, J. P. (2016). Reorganization between preparatory and movement population responses in motor cortex. Nature communications, 7, 13239. https://doi.org/10.1038/ncomms13239

Farina, D., Mesin, L., Martina, S. & Merletti, R. Comparison of spatial filter selectivity in surface myoelectric signal detection: Influence of the volume conductor model. Med. Biol. Eng. Comput. 42, 114–120 (2004).

Ficarella, S. C. & Battelli, L. Motor Preparation for Action Inhibition: A Review of Single Pulse TMS Studies Using the Go/NoGo Paradigm. Front. Psychol. 10, 340 (2019).

Guerrero, F. N., García, P. A. & Spinelli, E. M. Signal modes for design-oriented analysis of active sEMG spatial filter electrodes. in VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th -28th, 2016 (eds Torres, I., Bustamante, J. & Sierra,

D. A.) vol. 60 504–507 (Springer Singapore, Singapore, 2017).

Kaufman, M. T., Churchland, M. M., Ryu, S. I., & Shenoy, K. V. (2014). Cortical activity in the null space: permitting preparation without movement. Nature neuroscience, 17(3), 440–448. https://doi.org/10.1038/nn.3643

Leocani, L., Cohen, L. G., Wassermann, E. M., Ikoma, K., & Hallett, M. (2000). Human corticospinal excitability evaluated with transcranial magnetic stimulation during different reaction time paradigms. *Brain : a journal of neurology*, *123 (Pt 6)*, 1161–1173. https://doi.org/10.1093/brain/123.6.1161

Merletti, R. & Cerone, G. L. Tutorial. Surface EMG detection, conditioning and pre-processing: Best practices. J. Electromyogr. Kinesiol. 54, 102440 (2020).

Neige, C., Mavromatis, N., Gagné, M., Bouyer, L. J., & Mercier, C. (2018). Effect of movement-related pain on behaviour and corticospinal excitability changes associated with arm movement preparation. The Journal of physiology, 596(14), 2917–2929. https://doi.org/10.1113/JP276011

Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 1568–1569 (2011).

Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011).

R Core Team. (2022). *R: The R Project for Statistical Computing*. https://www.r-project.org/. https://www.r-project.org/

Soto, O., Valls-Solé, J., & Kumru, H. (2010). Paired-pulse transcranial magnetic stimulation during preparation for simple and choice reaction time tasks. Journal of neurophysiology, 104(3), 1392–1400. https://doi.org/10.1152/jn.00620.2009

van Elswijk, G., Schot, W. D., Stegeman, D. F., & Overeem, S. (2008). Changes in corticospinal excitability and the direction of evoked movements during motor preparation: a TMS study. BMC neuroscience, 9, 51.

https://doi.org/10.1186/1471-2202-9-51

Wickham, H. (2016). Ggplot2. Springer Science+Business Media, LLC

Wong, A. L., Haith, A. M., & Krakauer, J. W. (2015). Motor Planning. The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry, 21(4), 385–398. https://doi.org/10.1177/1073858414541484

World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. JAMA 310, 2191–2194 (2013).

Chapter 3

Does Experience Modulate Automatic Imitation? A New Look[4]

1. INTRODUCTION

Automatic imitation is a stimulus-response compatibility effect, wherein observing another's action heavily affects the observer's motor performance. In a typical automatic imitation task, participants are visually presented with one of two actions (task-irrelevant stimuli) and a cue (colors or numbers) indicating the action they have to perform (task-relevant stimuli). Participants usually produce the cue commanded action faster if the latter is similar to the observed action. Such an effect has been replicated several times since its discovery, employing different experimental variants (Brass et al., 2000, 2001; Stürmer et al., 2000; for a review, see Cracco et al., 2018; Heyes, 2011).

Automatic imitation has been explained through various mechanisms. The Ideomotor theory posits a tight relationship between perception and action (Greenwald, 1970; Prinz, 1990, 2005). It claims that representing an action's perceptual consequences is crucial for controlling its execution, as it enables the agent to anticipate the expected sensory feedback when the action is performed correctly. The same mechanism also accounts for automatic imitation: the perception of an event similar to the events one has experienced following one's own actions facilitates executing that action.

Similarly, the Motor Resonance theory suggests that when observing another's action, the sensory information concerning that action is automatically transformed into the motor representation involved in planning and executing that action (Rizzolatti & Sinigaglia, 2010, 2016). This explains the automatic imitation

[4] This chapter is based on the manuscript in preprint: Genovese, F., Fanghella, M., Sinigaglia, C., Barchiesi, G. Does Experience Modulate Automatic Imitation? A New Look bioRxiv 2024.04.17.589868; doi: https://doi.org/10.1101/2024.04.17.58986

effect, with the participants performing the observed action faster even if the sensory information concerning that action is task-irrelevant.

The Associative Sequence Learning (ASL) theory differs from the Ideomotor Theory and Motor Resonance by claiming instead that automatic imitation relies on bidirectional sensorimotor contingencies (Catmur et al., 2009; Cook et al., 2014; Heyes, 2010). According to ASL, the links between perceptual and action events develop through bidirectional Pavlovian-like associations between sensory and motor experiences. These associations are purely contingent and arise whenever actions happen to be performed and observed together without prioritizing action execution over action observation (Cook et al., 2014). The contingent co-occurrence of similar perceptual and action experiences would explain why observing an action might enhance its execution. However, this enhancement would be reduced or disappear if the contingencies differed (Ray & Heyes, 2011).

In a demonstration that has since become classic, Heyes et al. (2005) provided the first evidence that automatic imitation can be affected by contingent sensorimotor experiences. Two groups of participants underwent two distinct visuomotor training sessions. In the "imitative" training, participants watched a hand extend or flex its fingers and mimicked the observed action. In the "counter-imitative" training, participants were instructed to perform the opposite action to what they saw, flexing their fingers when the hand extended them, and vice versa. Following the training, both groups participated in a single-choice version of an automatic imitation task, extending or flexing their fingers. The training sessions had differing effects on the magnitude of automatic imitation, consistently with ASL predictions. While the imitative group exhibited an automatic imitation effect, the counter-imitative group abolished this effect.

ASL predictions have also been tested at the neurophysiological level through a passive observation paradigm. Catmur et al. (2007) employed transcranial magnetic stimulation (TMS) to stimulate the primary motor cortex and record motor-evoked potentials (MEP) from the first dorsal interosseus (FDI) and abductor digiti minimi (ADM) muscles while participants passively observed index and little finger movements involving those muscles. Before and after imitative training, MEPs in the ADM and FDI muscles were greater when observing little-finger and index-finger movements, respectively. Following counter-imitative training, the MEP amplitude pattern reversed.

Although these findings demonstrated that contingent sensorimotor experiences could modulate automatic imitation, it is still unclear how this modulation occurs. Indeed, two TMS studies on passive action observation produced contrasting results. Barchiesi & Cattaneo (2013) delivered single TMS pulses on the primary motor cortex at four intervals from the onset of a visually presented action, showing that counter-imitative training affected visuomotor facilitation only at late time points, leaving earlier ones unaffected. This result was interpreted as the interaction between two mechanisms: a fixed visuomotor transformation responsible for the early motor facilitation effects and a flexible arbitrary visuomotor association mechanism for the later effects. Conversely, Cavallo et al. (2014) found that the counter-imitative training impacted visuomotor facilitation at all time points, where it was found during the pre-training session, thus suggesting that the facilitation effect would be due to only one mechanism, which is contingent association.

The present study aims to shed new light on the automatic imitation effect and its potential modulation by describing, on the behavioral front, the high-resolution time course of this effect before and after imitative and counter-imitative training. We leveraged the time course design logic implemented in Barchiesi & Cattaneo (2013) and Cavallo et al. (2014), employing a double-choice version of the automatic imitation paradigm in which the delay between the task-irrelevant event (the observation of a hand flexing or extending its fingers) and the task-relevant event (a colored cue indicating whether the action to be performed was a flexion or an extension of the fingers) was parametrically modulated.

Our main aim was to assess, for the first time, whether counter-imitative training consistently affects the time course of the automatic imitation effect. In a pre-training session we tested the presence of automatic imitation across nine delays between the presentation of the colored cue and the observed action. After a session of counter-imitative training, we conducted a post-training session similar to the pre-training one. Then, we compared the magnitude of the automatic imitation effect between pre and post training sessions, focusing on delays where automatic imitation emerged during the pre-training session.

Suppose automatic imitation entirely depends on sensorimotor contingencies, as assumed by ASL and suggested by Heyes et al. (2005)'s findings. In that case, one should expect that the counter-imitative training would cause a suppression or (at least) a reduction of the automatic imitation effect in all time points where it emerged

in the pre-training session. If this is not the case, and the automatic imitation effect remains unaffected by counter-imitative training at any point during its progression, it would involve that the automatic imitation effect is not, or not only, a matter of contingent association, at least not in the way it has been understood thus far.

2. MATERIALS AND METHODS

2.1 Sensitivity Analysis

Three previous behavioral experiments implementing a "training-test" rationale employed 10 (Heyes et al., 2005), 8 (Press et al., 2007), and 12 (Cook et al., 2010) participants for each training group. To detect even smaller effects resulting from our experimental design, we decided to roughly quadruplicate the number of participants for each training group to 40.

The result of a sensitivity analysis performed using MorePower (Campbell & Thompson, 2012) (alpha = 0.05, power = 0.8, 2 trainings x 2 sessions x 9 delays) indicates that employing a total of 80 participants for our experimental design is sufficient for detecting effects as small as Cohen's f = 0.156 (partial eta-squared = 0.238), which is typically considered between "small" and "medium" (Perugini et al., 2018).

Data, analysis code, research materials employed, are available at https://osf.io/a3zns/. Data were analyzed using Jamovi (Jamovi project, 2024), JASP (JASP Team 2024), R, Version 4.3.2 (R Core Team, 2022), and the package ggplot2, Version 3.4.4 (Wickham, 2016).

2.2 Participants

We collected data from 84 participants (right-handed, with no neurological or psychiatric disorders, vision either corrected or normal); four of them did not participate in all the experimental sessions and were excluded from the analyses. Most participants were university students enrolled in courses held at the Department of Philosophy of the University of Milan. The participation provided them with university credits. Two groups of 40 participants were enrolled in the "Counter" experiment (17 women, age: 25.1 +/- 7.60) or the "Imitative" experiment (17 women age: 25.2 +/- 5.97). The study complied with the revised Helsinki Declaration (World

Medical Association General Assembly, 2008) and was approved by the local Ethical Committee. All participants gave written informed consent for their participation.

2.3 Stimuli

Four right-hand pictures on a black background (two male and two female hands) were presented in the experimental sessions. Each hand could be presented with the fingers in three positions: flexed, intermediate (neutrally positioned), or extended. Building on previous research (Barchiesi & Cattaneo, 2015; Heyes et al., 2005; Ubaldi et al., 2015), we leveraged the illusion of movement produced by the rapid transition from images of a neutrally positioned hand to those showing flexed or extended fingers. This solution allows an easy control of event timings. Two small, round objects were shown on the images: one placed between the fingers and the other on the dorsal side of the presented hand. These served as the targets of the displayed goal-directed actions. When finger flexion was shown, the action was a goal-directed grasp of the object between the fingers. When finger extension was shown, the action was a goal-directed touch of the external round object with the dorsal side of the fingertips.

2.4 Response Detection System

2.4.1 Flex-sensor

As in recent studies (Barchiesi et al., 2022; Barchiesi & Cattaneo, 2015; Cattaneo et al., 2015; Guidali et al., 2023; Zazio et al., 2022), we took advantage of variable resistors to capture participants' behavior. One key advantage of this method is its capacity to unambiguously identify and record opposing responses along a single continuous dimension such as the magnitude of an applied force or an effector's bending amount.

In the present experiment, participants had to produce a flexion or an extension of their right hand fingers following the appearance of a colored cue on the screen (see next paragraph). To record finger movements, we used bending variable resistors (“flex sensors”), which can precisely capture opposing responses along a single continuous dimension, such as the degree of bending. In our task, participants flexed or extended the fingers of their right hand in response to a colored cue on the

screen. A 95 mm flex-sensor was mounted on the dorsal side of the right-hand index finger. Two soft sheaths (heat-shrink sleeves, ~0.5 cm diameter × ~2.5 cm length) were attached to the proximal and middle phalanges and secured with skin adhesive tape (without flattening them). The flex sensor was then inserted into the sheaths and fixed to the index metacarpus. This arrangement allowed the sensor to bend and slide within the sheaths during finger movements, and to return to its original position when the fingers relaxed (Figures 1–3). The flex-sensor was connected in series with a fixed resistor in a voltage-divider circuit, powered and read out by an Arduino® board. Voltage variations caused by finger bending were converted into conductivity values. For each session, maximal flexion and extension were measured and mapped to 0 and 1000 arbitrary units, respectively. Response thresholds were set to ±100 units relative to the initial conductivity at cue onset. For example, an initial reading of 430 units yielded extension and flexion thresholds of 530 and 330 units, respectively. During each session, the right hand was concealed under an opaque box to prevent visual feedback.

2.4.2 Foam Rubber Manipulandum

Between the participant and the PC monitor, a foam rubber column (diagonal ~4.5 cm, height ~14 cm) and a semi-rigid base were fixed to the table. The fifth metacarpal of the right hand rested on the base, with the fingers placed around the foam column. This setup positioned the fingers midway between extension and flexion at trial onset and allowed comfortable flexion due to the column’s softness (Figures 1–3).

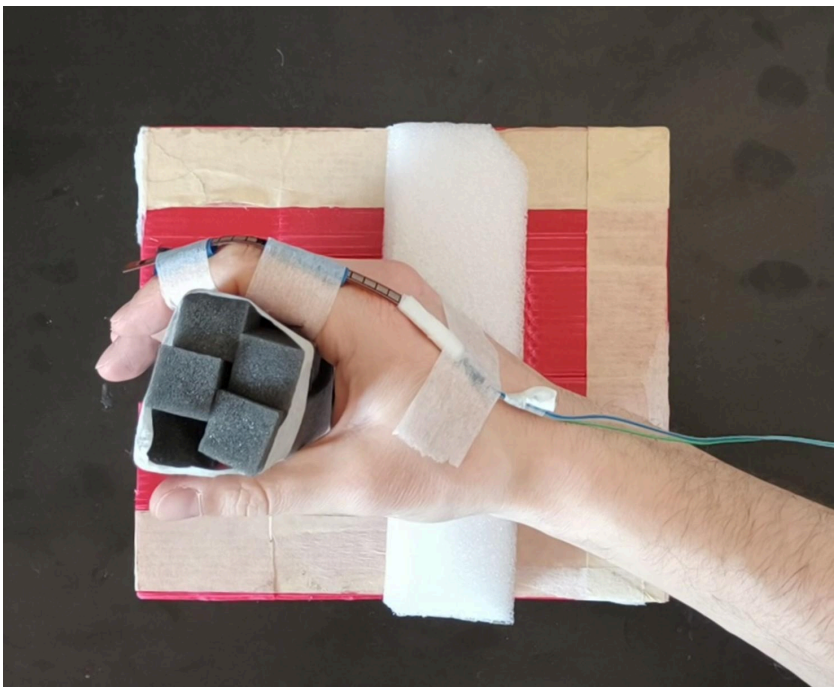

Figure 1*: Example of a participant's hand starting position at the beginning of each trial. The foam rubber structure (dark grey) keeps the fingers in an intermediate position while waiting for the stimuli to appear on the screen. The hand lays on a semi-rigid base (white), which allows the fingers to move freely during extension by creating space between the table and the fifth metacarpal side of the hand. Additionally, if necessary, a soft foam rubber support has been placed beneath the forearm to maintain alignment of the hand-forearm complex.*

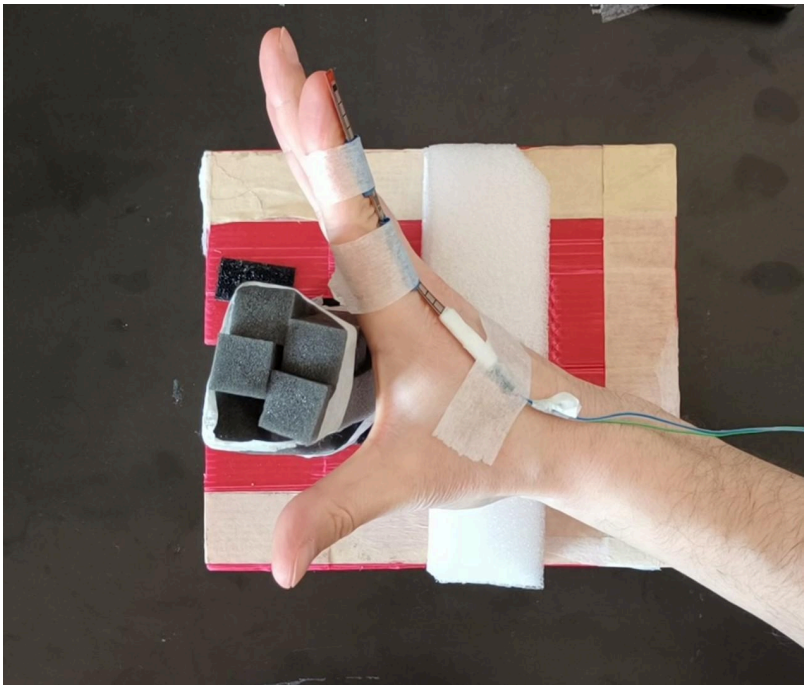

Figure 2*. Example of fingers extension movement.*

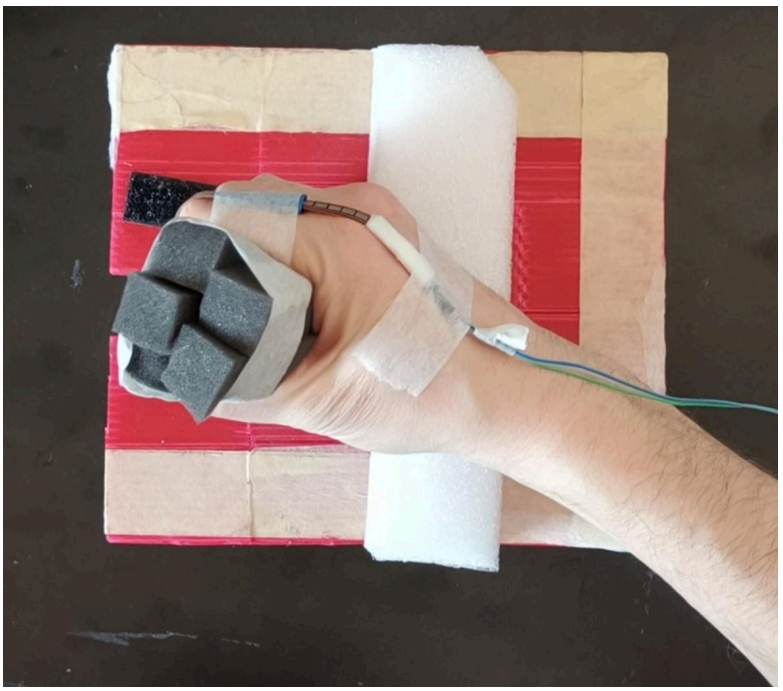

Figure 3. *Example of fingers flexion movement.*

2.5 General Experimental Structure

Each participant underwent three sessions: a “Pre-training” session (common to both groups), in which a forced double-choice reaction time version of the automatic imitation paradigm was tested. A training session followed the Pre-training: in this session, one group performed a counter-imitative visuomotor training (“Counter” group) while the other performed an imitative visuomotor training (“Imitative” group). Both groups eventually performed a “Post-training” session, similar to the Pre-training session, the following day. Each session lasted approximately 40 minutes.

2.5.1 Automatic Imitation Time Course (Pre-training and Post-training Sessions)

2.5.1.1 Task and Trial Structure

The experimental design was implemented and run by custom Matlab R2021b (Mathworks®) scripts employing Psychtoolbox functions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Stimuli were presented on a 60 Hz refresh rate monitor (DELL P2419H, resolution 1920 x 1080 pixels, screen size: 23.8 inches). Each trial presented a randomly oriented (0-360°) neutrally positioned hand on the screen,

which lasted randomly between 1000 and 1500 ms. In the remainder of the trial, a colored cue (either light green or purple, lasting 67 ms) and an action (either a flexion or an extension of the fingers, hereafter referred to here as “observed action”) were presented on the screen; the critical feature of each trial was the temporal delay between the presentation of the colored cue and the presentation of the observed action. Here we conceptually grouped them into three categories: in the “negative delays” trials, the neutrally positioned hand turned colored, and after 700, 300, 200, or 100 ms, an action was presented on the screen; in the “simultaneous delay” trials, the colored cue was presented simultaneously with the observed action (0 ms delay); eventually in the “positive delays” trials, the colored cue could be presented 100, 200, 300, 400 ms after the observed action (Figure 4).

The order of the nine delays was randomized across sessions and participants. The participants’ task was to produce a flexion or an extension of their right hand fingers according to the presented colored cue as fast as possible and as accurately as possible; the association rule between the color and the movement to be produced was counterbalanced across participants. After either a detected response or the end of the response time window (until 700 ms from the colored cue presentation), participants were provided with a feedback screen for 500 ms: if the response was correct and provided between 150 and 700 ms, their reaction time in milliseconds was displayed; if the detected reaction time was faster than 150 ms, a “Do not guess!” alert was displayed, if it exceeded the deadline time, a “Too Slow” alert was presented; if the provided response was wrong, a “Wrong!” label was shown (feedback labels have been displayed in Italian). Eventually, a black screen was displayed for 1000 ms before the subsequent trial started.

To enhance participants' motivation and encourage them to perform at their best, we informed them at the beginning of each session that they would receive a prize at the end of the study if they ranked first based on average reaction times and accuracy. Responses beyond the deadline or before the color appeared incurred penalties.

Each Pre-training and Post-training session comprised 720 trials, with 80 trials for each of the nine delays. Within each delay, there were 40 congruent trials, where the observed action matched the movement commanded by the colored cue, and 40 incongruent trials, where the observed actions were the opposite of those

commanded by the colored cue. Participants could briefly rest within each session once every 240 trials (2 breaks).

On each trial, at the onset of the colored cue, a TTL trigger was delivered through the parallel port of the presentation PC into a digital input pin of an Arduino board. A custom-made routine, loaded on the board, measured the value of the flex-sensor output and compared it with the flexion and extension thresholds for response detection. Once the response and the reaction time were calculated, they were sent back, via serial communication, to Matlab for visual feedback presentation.

Before the Pre-training session, participants underwent a brief familiarization session (~ 5 minutes, 80 trials).

2.5.1.2 Familiarization Session

Participants underwent a familiarization session to learn the association between the movements to be produced and the colored cues. The task employed was identical to the one employed in the Pre-training and Post-training sessions, with the difference that the hands on the screen did not move (they only changed color) and that the time window for response production was increased to 3000 ms. Participants underwent the familiarization session only before the Pre-training session.

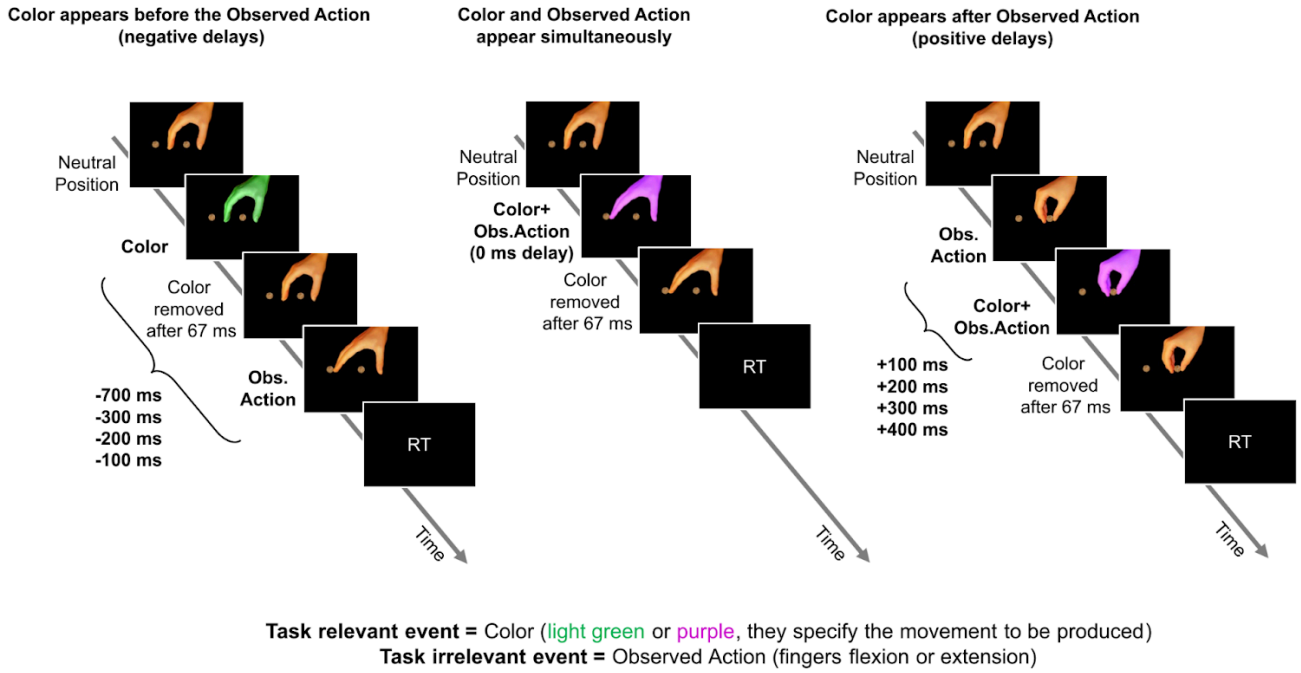

***Figure 4**: Structure of the automatic imitation task. In each trial of the initial (Pre-training) and the last session (Post-training), the task-relevant event, the colored cue (either light green or purple), appears at one of nine different delays with respect to the task-irrelevant event, the observed action (either fingers flexion or extension). The order of the delays has been randomized across trials. The leftmost timeline depicts a trial where the colored cue flashes on the neutrally positioned hand -700, -300, -200, or -100 ms relative to the subsequent presentation of the observed action ("negative delays trials"). The middle timeline represents a trial where the colored cue and the observed action appear simultaneously (delay = 0 ms). The rightmost timeline represents a trial in which the colored cue appears 100, 200, 300, or 400 ms after the observed action has been presented on the screen ("positive delays trials"). In each trial type, the colored cue lasts 67 ms. The orientation of the displayed hand was randomized across trials between 0 and 360 degrees.*

2.5.2 Counter-imitative Training and Imitative Training Sessions

The Counter and Imitative groups training sessions comprised 720 trials divided into 10 consecutive blocks. The trial structure was similar to the Pre-training and Post-training sessions but with a few differences: No colored cue was displayed in this session, as the participants' task was to produce, as quickly and accurately as possible, the opposite action compared to the one they observed in the Couter-Imitative training, while it was to produce the same action observed in the Imitative training session. Each participant underwent only one of the two trainings, depending on the group they were assigned to. The feedback screens were the same as in the Pre- and Post-training sessions. A further difference compared to the Pre- and Post-training sessions was that the displayed hands were always oriented in an egocentric perspective, in the attempt of mimicking a natural visuo-motor contingency acquisition situation (such as self-observation, Ray & Heyes, 2011) where the spatial relationship between the effector and the movement sensory outcomes remain stable across time.

2.6 Data Analysis

2.6.1 Analysis: Pre and Post-training Sessions

Only reaction times (RTs) between 150 ms and 700 ms from correct responses were included in the following analysis. To understand whether the automatic imitation effect was differently affected by the two trainings, a mixed three-way 2x9x2 ANOVA has been conducted with SESSION (Pre-training and Post-training) and DELAY (-700, -300, -200, -100, 0, 100, 200, 300, and 400 ms) as within-subject factors and TRAINING (Counter and Imitative) as a between-subject factor. To facilitate the interpretation of the results, an index of automatic imitation effect was computed as the dependent variable: we calculated the median RT of the congruent trials, and we subtracted it from the median RT of the incongruent ones separately for each cell of the three-way ANOVA and for each participant. Thus, if participants were slower in the incongruent trials than congruent ones, this would have been reflected as a positive automatic imitation index, and vice versa if incongruent trials were faster than congruent ones.

Given that a significant three-way interaction was detected, we separated the three-way design into two two-way 2x9 repeated-measures ANOVAs (one for each training group), with SESSION and DELAY as within-subject factors, to understand whether the training changed the automatic imitation effect in the Post-training compared to the Pre-training session.

A significant two-way interaction was only obtained for the Counter group. Crucially, since the experiment aimed to describe how the time course of the automatic imitation effect was modulated between the Pre- and Post-training sessions, we first identified, within the Pre-training session, at which delays an automatic imitation effect was detected; secondly, we tested if, at these delays, the automatic imitation effect was modulated in the Post-training session.

Thus, we performed a one-way repeated-measures ANOVA on the Counter group's Pre-training session with DELAY as the only factor, followed by nine Bonferroni-corrected t-tests (0.05/9) at each delay.

Finally, using five Bonferroni-corrected (0.05/5 = 0.01) paired sample t-tests, we tested the modulation of the automatic imitation effect between the Post- and Pre-training sessions at those delays where an automatic imitation effect was detected in the Pre-training session.

As a post-hoc analysis, we tested a residual automatic imitation effect within the Post-training session using nine Bonferroni-corrected post-hoc tests across delays.

Given the absence of a two-way interaction within the Imitative group, we did not further analyze the modulation of the automatic imitation effect between the Pre- and Post-training sessions. However, for the sake of Figure 5 clarity and interpretability, we performed nine Bonferroni-corrected paired-sample t-tests for each session across all delays also for the Imitative group (see Figure 5).

2.6.2 Analysis: Counter and Imitative Training Sessions

To test participants' performance improvement during the training sessions we conducted, separately for each group, a one-way repeated-measures ANOVA with TRAINING BLOCKS (from the first to the tenth) as the independent variable and the median RT for each block as the dependent variable. RTs between 150 ms and 700 ms from correct responses were included in the analysis. We fitted a linear-mixed regression model to test a linear trend across the blocks. The model predicted the

RTs based on training blocks as a fixed effect, accounting for random intercepts and slopes for training blocks across subjects.

3. RESULTS

3.1 Results: Pre and Post-training Sessions

The percentage of trials included in the analysis is 89.8% for the Counter group and 88.4% for the Imitative group. The mixed three-way ANOVA (TRAINING x SESSION x DELAY) produced a significant main effect of SESSION ($F_{1, 78}$ = 10.05, p = 0.002, partial-η2 = 0.114) with the automatic imitation effect greater in the Pre-training session compared to the Post-training session (23.1 ms vs. 19.1 ms) and a significant main effect of DELAY ($F_{6,78}$ = 94.84, p < .001, partial-η2 = 0.549; -700 = 0.12 ms, -300 = -0.35 ms, -200 = 2.26 ms, -100 = 8.78 ms, 0 = 27.03 ms, 100 = 41.9 ms, 200 = 40.72 ms, 300 = 37.75 ms, 400 = 31.85 ms), with later delays showing greater automatic imitation effects. A main effect of TRAINING was found ($F_{1, 78}$ = 8.03, p < 0.006, Counter = 17.97 ms, Imitative = 24.26 ms) with greater automatic imitation effect in the Imitative compared to the Counter training. Moreover, three significant two-way interactions were found: a SESSION x TRAINING ($F_{1, 78}$ = 13.08, p < 0.001, partial-η2 = 0.144), showing that the difference in automatic imitation effect between Pre- and Post-training in the Counter group was greater compared to the Imitative group (counter Pre-training = 22.3 ms, counter Post-training = 13.6 ms, Imitative Pre-training = 24 ms, Imitative post-training = 24.5 ms). A DELAY x TRAINING ($F_{6, 467.63}$ =3.09, p = 0.006, partial-η2 = 0.038) interaction was found, indicating greater automatic imitation effect in the Imitative group for later delays (Counter group: -700 = 0.56 ms, -300 = 1.69 ms, -200 = 2.04 ms, -100 = 5.59 ms, 0 = 22.12 ms, 100 = 38.98 ms, 200 = 35.05 ms, 300 = 30.38 ms, 400 = 25.41 ms; Imitative group: -700 = -0.32 ms, -300 = -2.38 ms, -200 = 2.48 ms, -100 = 11.98 ms, 0 = 31.95 ms, 100 = 44.82 ms, 200 = 46.4 ms, 300 = 45.11 ms, 400 = 38.28 ms). A SESSION x DELAY interaction ($F_{6. 99}$ = 4.02, p < 0.001, partial-η2 = 0.049) indicated that a greater automatic imitation effect was detected in the Pre-training session for later delays compared to Post-training automatic imitation effect (Pre-training: -700 = -0.43 ms, -300 = 3.60 ms, -200 = -0.07 ms, -100 = 8.37 ms, 0 = 26.14 ms, 100 = 44.47 ms, 200 = 46.82 ms, 300 = 42.37 ms, 400 = 36.97 ms; post: -700 = 0.67 ms, -300 = -4.3 ms, -200 = 4.58 ms, -100 = 9.2 ms, 0 = 27.92 ms, 100 = 39.32 ms, 200 =

34.63 ms, 300 = 33.12 ms, 400 = 26.72 ms). Eventually, a mixed three-way ANOVA (TRAINING x SESSION x DELAY) produced a significant three-way interaction ($F_{6.99, 545.01}$ = 3.132; p = 0.003; partial-η2 = 0.039, indicating that the automatic imitation effect in the Post-training session was affected differently by the two types of training across delays, compared to the Pre-training session. Greenhouse-Geisser correction was applied for departure from sphericity when needed (Table 1). Following the detection of a three-way interaction in the mixed three-way ANOVA, we explored each group separately to test the effect of each training through two two-way ANOVAs. The Counter group two-way ANOVA produced a main effect of SESSION ($F_{1, 39}$ = 22.29 p = < 0.001, partial-η2 = 0.364) with greater Pre-training automatic imitation compared to the Post-training and a main effect of DELAY ($F_{5.71, 222.57}$ =36.02, p < .001, partial-η2 = 0.48) with greater automatic imitation for later delays. A SESSION x DELAY interaction for the Counter group was found ($F_{6.17, 240.76}$ = 6.55; p < 0.001; partial-η2 = 0.14), indicating that the automatic imitation effect in the Post-training session was differently modulated across delays compared to the Pre-training session. The analysis of the two two-way repeated-measures ANOVAs for the Imitative group produced a significant main effect of DELAY ($F_{5.26, 205.12}$ = 60.936, p < .001, partial-η2 = 0.61) with greater automatic imitation for later delays. However, we did not find such a two-way interaction for the DELAY x SESSION effect in the Imitative group ($F_{6.59, 256.99}$ = 0.921; p = 0.487; partial-η2 = 0.023), indicating that the automatic imitation effect in the Post-training session was not modulated across delays with respect to the Pre-training session (Table 2).

The one-way repeated-measures ANOVA conducted on the delays of the Pre-training session of the Counter group produced a significant main effect of DELAY (Table 3) ($F_{5.88, 229.31}$ = 32.8, p < 0.001; partial-η2 = 0.457 Greenhouse-Geisser corrected), indicating that the automatic imitation effect was increased with greater delays. Nine Bonferroni-corrected t-tests against zero (0.05/9 = 0.0056, Table 4) detected a positive automatic imitation effect from 0 to 400 ms delays.

We then tested whether the automatic imitation effect was reduced between the Post-training and Pre-training sessions within the Counter group. We limited the testing to those delays in which the automatic imitation was found in the Pre-training session (0, +100, +200, +300, +400 ms). The paired sample t-tests (Bonferroni

corrected for five comparisons) showed that, compared to the Pre-training session, Post-training automatic imitation effects were reduced from +100 ms to 400 ms delays (all ps <0.01; see Table 5). In contrast, they were not modulated at the 0 ms delay, i.e., at the first delay in which automatic imitation emerged in the Pre-training session (t_{39}= 0.52 *p* = 0.604, Cohen's d = 0.083).

Eventually, as a post-hoc analysis, we tested the (residual) presence of the automatic imitation effect in the Counter group's Post-training session. As in the Pre-training session, the automatic imitation effect was detected from the 0 ms delay onward (Bonferroni corrected for nine comparisons) (Table 4).

Table 1

Three-way Mixed Anova

	Sum of Squares	df	Mean Square	F	*p*	η^2p
Session	5887	1	5887	10.05	**0.002**	0.114
Session ✻ training	7666	1	7666	13.08	**< .001**	0.144
Residual	45702	78	586			
Delay	424390	6	70788	94.84	**< .001**	0.549
Delay ✻ training	13833	6	2307	3.09	**0.006**	0.038
Residual	349023	467.63	746			
Session ✻ Delay	12319	6.99	1763	4.02	**< .001**	0.049
Session ✻ Delay ✻ training	9593	6.99	1373	3.13	**0.003**	0.039
Residual	238901	545.01	438			
Training	14191	1	14191	8.03	**0.006**	0.093
Residuals	137903	78	1768			

Table 2

Two-way ANOVAs													
Counter Group							**Imitative Group**						
	Sum of squares	**df**	**Mean square**	**F**	***p***	**η^2p**		**Sum of squares**	**df**	**Mean square**	**F**	***p***	**η^2p**
Session	13494	1	13494	22.29	**< .001**	0.364	**Session**	58.7	1	58.7	0.104	0.749	0.003
Residual	23611	39	605				Residual	22090.7	39	566.4			
Delay	154813	5.71	19352	36.02	**< .001**	0.48	**Delay**	283410	5.26	53885	60.936	**< .001**	0.61
Residual	167637	222.57	537				Residual	181386.3	205.12	884.3			
Session ✻ Delay	18931	6.17	2366	6.55	**< .001**	0.144	**Session ✻ Delay**	2980.6	6.59	452.3	0.921	0.487	0.023
Residual	112697	240.76	361				Residual	126204	256.99	491.1			

Table 3

One-way ANOVA						
	Sum of Squares	**df**	**Mean Square**	**F**	**p**	**η^2p**
Delay	130427	5.88	22183	32.8	**< .001**	0.457
Residual	155240	229.31	677			

Table 4

Counter Group															
Pre-training								**Post-training**							
Delay (ms)	**Incongruent (ms)**	**Congruent (ms)**	**Automatic Imitation (ms)**	**T-stat**	**df**	**Cohen's d**	**P-value *(p<0.0056)***	**Delay (ms)**	**Incongruent (ms)**	**Congruent (ms)**	**Automatic Imitation (ms)**	**T-stat**	**df**	**Cohen's d**	**P-value *(p<0.0056)***
-700	384	387	-3	-0.86	39	-0.14	0.393	-700	370	366	4	1.50	39	0.24	0.141
-300	391	384	7	2.12	39	0.34	0.040	-300	366	370	-4	-1.39	39	-0.22	0.173
-200	383	383	0	0.11	39	0.02	0.916	-200	370	366	4	2.19	39	0.35	0.034
-100	389	383	6	1.61	39	0.25	0.116	-100	370	365	5	1.73	39	0.27	0.091
0	396	372	24	6.37	39	1.01	***<.001***	0	373	352	21	7.14	39	1.13	***<.001***
100	394	347	47	12.15	39	1.92	***<.001***	100	366	335	31	7.91	39	1.25	***<.001***
200	383	336	47	11.18	39	1.77	***<.001***	200	352	329	23	6.01	39	0.95	***<.001***
300	375	336	39	8.40	39	1.33	***<.001***	300	341	320	21	5.71	39	0.90	***<.001***
400	367	332	35	8.65	39	1.37	***<.001***	400	339	322	17	3.52	39	0.56	***0.001***
Imitative Group															
Pre-training								**Post-training**							
Delay (ms)	**Incongruent (ms)**	**Congruent (ms)**	**Automatic Imitation (ms)**	**T-stat**	**df**	**Cohen's d**	**P-value *(p<0.0056)***	**Delay (ms)**	**Incongruent (ms)**	**Congruent (ms)**	**Automatic Imitation (ms)**	**T-stat**	**df**	**Cohen's d**	**P-value *(p<0.0056)***
-700	385	383	2	0.66	39	0.10	0.513	-700	368	371	-3	-0.85	39	-0.14	0.398
-300	385	385	0	0.02	39	0.00	0.985	-300	368	373	-5	-1.79	39	-0.28	0.082
-200	382	382	0	-0.16	39	-0.03	0.871	-200	372	366	6	1.89	39	0.30	0.066
-100	389	378	11	2.95	39	0.47	***0.005***	-100	375	362	13	3.66	39	0.58	***<.001***
0	398	369	29	6.90	39	1.09	***<.001***	0	387	352	35	8.24	39	1.30	***<.001***
100	386	343	43	9.11	39	1.44	***<.001***	100	374	326	48	12.34	39	1.95	***<.001***
200	380	333	47	9.13	39	1.44	***<.001***	200	364	318	46	13.31	39	2.10	***<.001***
300	372	326	46	9.50	39	1.50	***<.001***	300	357	312	45	13.25	39	2.10	***<.001***
400	366	326	40	7.60	39	1.20	***<.001***	400	349	312	37	8.20	39	1.30	***<.001***

Table 4. The *Delay* column indicates the time delays between the presentation of the observed action and the colored cue. The *Congruent* column presents the mean RTs, averaged across all participants, for trials where the observed action matched the action to be performed. The *Incongruent* column displays the mean RTs for trials where the observed and performed actions differed. The *t-test statistic*, *degrees of freedom*, *effect size*, and *p-value* columns report the statistical outcomes related to the automatic imitation effect, with significant effects highlighted in bold. A Bonferroni-corrected significance threshold of $p < 0.0056$ was applied to account for the nine comparisons.

Table 5

Pre vs. Post - Automatic Imitation Effect (Counter Group)						
Delay (ms)	**Pre (ms)**	**Post (ms)**	**t-statistic**	**df**	**Cohen's d**	**P-value *(p<0.01)***
0	23.238	20.987	0.5236	39	0.08279	0.604
100	46.775	31.188	4.4742	39	0.70744	***<.001***
200	46.938	23.163	4.877	39	0.77112	***<.001***
300	39.312	21.45	3.0773	39	0.48656	***0.004***
400	34.237	16.587	3.7388	39	0.59115	***<.001***

We did not analyze the Imitative group's interactions further, given the absence of a two-way interaction within the two-way ANOVA. As stated in the Analysis paragraph, we performed nine Bonferroni-corrected t-tests separately for each session across all delays also for the Imitative group. Results are displayed in Figure 5 and Table 2.

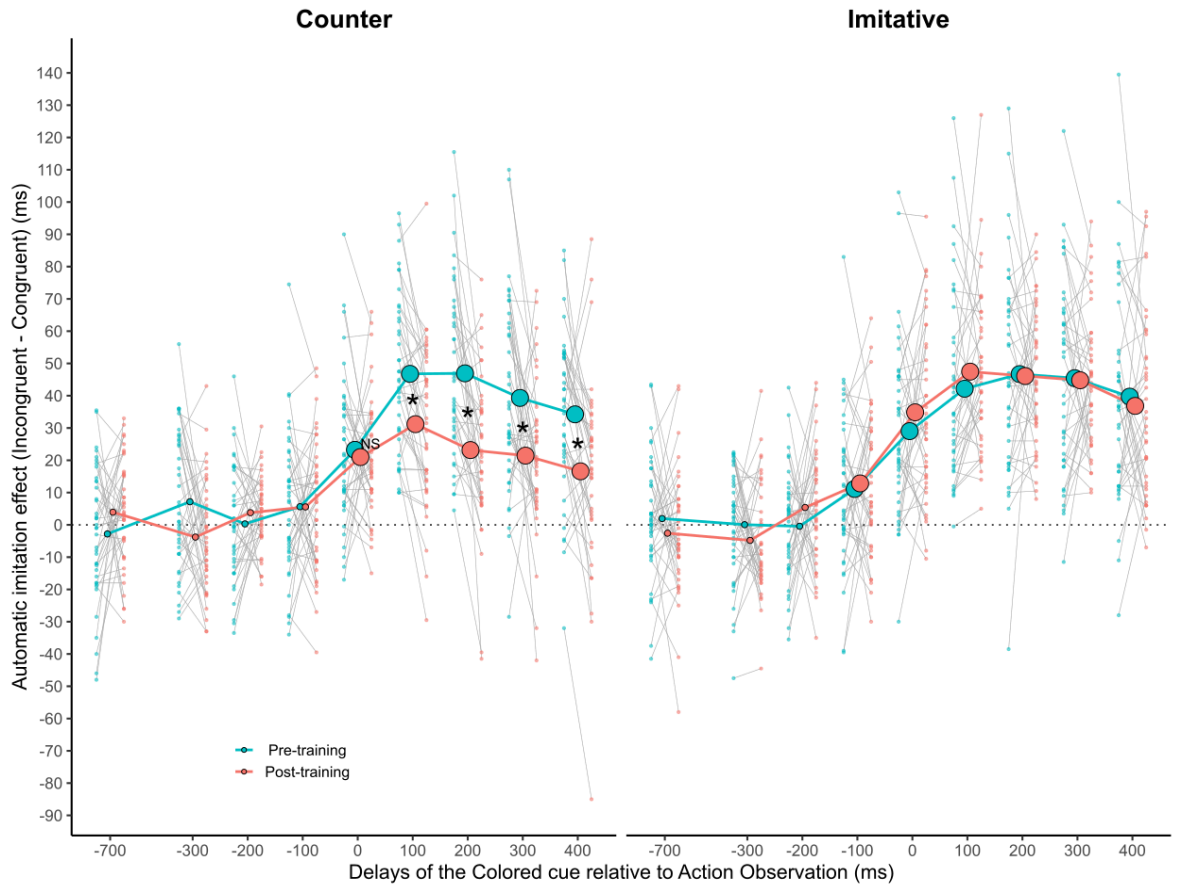

Figure 5*: The plot shows the average time course of the automatic imitation effect (RTs from incongruent trials minus RTs from congruent trials) within the Counter group (left) and the Imitative group (right), both in the Pre-training (thick turquoise line) and Post-training session (thick salmon line). Dots on thick lines represent delays at which the automatic imitation effect is significantly greater than zero (Bonferroni corrected for nine comparisons separately for each session and each group). For each delay where a significant automatic imitation was found in the Pre-training session, we compared the magnitude of the automatic imitation effect between Pre-training and Post-training sessions. The NS label signals a non-significant difference between Pre- and Post-training sessions. In contrast, the "*" symbol signals a significant difference in automatic imitation magnitude between the two sessions (Bonferroni corrected for five comparisons). No labels are present in the Imitative group plot since no two-way interaction has been detected; indeed, no statistical testing has been conducted relating the differences between automatic imitation effects between the Pre-training and*

the Post-training sessions. Thin grey lines at each delay show single-subject data. Within each delay, the left side of each thin line (small turquoise dots) identifies Pre-training data, and the right (small salmon dots) identifies Post-training data.

3.2 Results: Counter and Imitative Training Sessions

The percentage of trials included in the analysis is 89.5% for the Counter group and 94.2% for the Imitative group. The one-way repeated-measures ANOVA conducted on the training session of the Counter group produced a main effect of the TRAINING BLOCKS variable ($F_{3.89,\ 151.82}$ = 19.26, p<0.001, partial-η^2 = 0.33, Greenhouse-Geisser corrected). A significant linear trend has been detected across the counter-imitative training blocks, showing an average RT decrease of 5 ms per training block (estimated -5.0037, SE = 0.72, t_{39} = -6.977 p < 0.001). Also, for the Imitative group, the one-way repeated-measures ANOVA produced a main effect of the TRAINING BLOCKS variable ($F_{5.84,\ 227.85}$ = 3.50, p < 0.003, partial-η^2 = 0.08, Greenhouse-Geisser corrected). A significant linear trend has been detected across imitative training blocks, showing an average RT decrease of 0.97 ms per training block (estimated -0.9729, SE = 0.3137, $t_{38.99}$ = -3.102, p < 0.004).

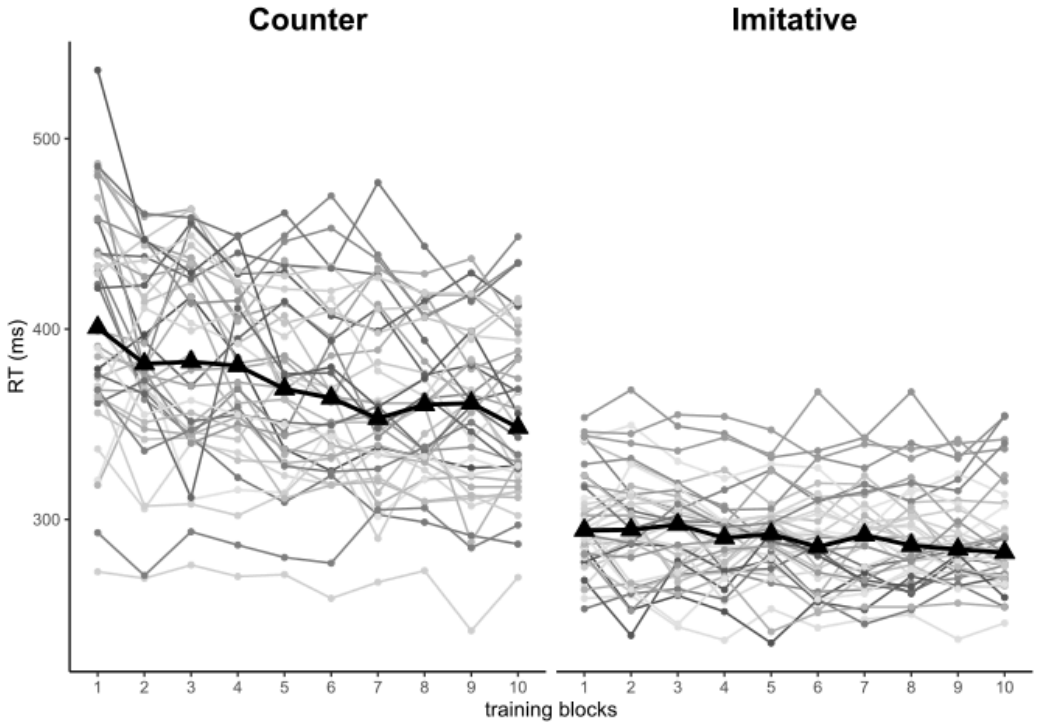

Figure 6*: The plot represents the modulation of the RTs across the training block for the Counter group (left) and the Imitative group(right). The black triangles connected by the thick black line show*

the average effect of the training across the ten blocks. The grey dots connected by the thin lines of different shades of grey represent the median RT value of each participant at each training block.

3.3 Exploratory analysis: The 0 ms Delay

To deeply understand our result at 0 ms delay we conducted different exploratory analysis. To quantify the likelihood that the absence of difference between Pre and Post-training automatic imitation effects does not result from an underpowered comparison between a greater Pre-training compared to the Post-training automatic effect at 0 ms, we took advantage of the Bayesian Factor (BF) being a metric that allows to quantify the probability that H0 (or H1) is true (Keysers et al., 2020). We conducted a Bayesian paired sample t-test using JASP software (JASP Team 2024) on the automatic imitation effect detected at 0 ms delay. We tested the hypothesis that the Pre-training automatic imitation is greater than the Post-training automatic imitation (one-tailed). Cauchy factor has been set as default (0.707).

The test resulted in a BF+0 = 0.269 (error ~0.00006), providing “moderate” evidence in favor of H0 (Jeffreys, 1961; Keysers et al., 2020), that is, the hypothesis that Pre-training automatic imitation is not greater than Post-training automatic imitation. We tested whether the automatic imitation varies with response quickness, particularly if, at longer RTs, there were hints that automatic imitation was greater in the Pre-training session than in the Post-training one at 0 ms delay within the Counter group. We performed, within the 0 ms delay, the same analysis performed in the main text for the Counter group, with percentile variable in the place of delay: the percentile variable is constituted by the 10th, 30th, 50th, 70th and 90th RTs percentiles for each design session x congruency combination.

Similar to the main analysis, we subtracted RTs at each percentile of the Congruent trials from the Incongruent trials to obtain an automatic imitation index effect separately for each session (Figure 7).

An ANOVA on automatic imitation with SESSION (Pre-training, Post-training) and PERCENTILE (10, 30, 50, 70, 90) as within factors has been conducted, showing a main effect of the PERCENTILE and an interaction between SESSION and PERCENTILE (Table 6).

Table 6

Effect	df	MSE	F	η^2p	p-value

session	1, 39	1614.21	0.75	.019	.393
percentiles	1.86, 72.72	819.35	15.50 ***	.284	**<.001**
session:percentiles	2.34, 91.40	552.09	4.34 *	.100	**.012**

Table 6: ANOVA results testing Session x Percentile effects on automatic imitation

Following the strategy employed for the paper's main analyses, we tested whether an automatic imitation effect was present in the Pre-training percentiles (Bonferroni corrected 0.05/5 = 0.01) using a one-sample t-test against zero, showing that automatic imitation was present at each tested percentile within the pre-training session (Table 7).

Table 7

Percentile	df	automatic imitation (ms)	t-statistic	P-value (p<0.01)
10th	1,39	13.68	3.3003	**0.002**
30th	1,39	17.14	4.8193	**< 0.001**
50th	1,39	23.2375	6.3652	**< 0.001**
70th	1,39	30.615	7.4458	**< 0.001**
90th	1,39	45.8075	6.9298	**< 0.001**

Next, we tested whether the training significantly affected the automatic imitation effect at each tested percentile by comparing the automatic imitation effect recorded during Post- vs. Pre-training sessions. Paired-sample t-test comparisons performed on automatic imitation effect difference at each percentile between Post- and Pre-training sessions are shown in table 8:

Table 8

Percentile	df	automatic imitation (post vs pre ms)	t-statistic	P-value (p<0.01)
10th	1,39	4.68	0.88428	0.382
30th	1,39	3.32	0.84204	0.4049
50th	1,39	-2.25	-0.5235	0.6035
70th	1,39	-7.34	-1.6952	0.09801
90th	1,39	-15.78	-1.9636	0.05674

None of the tested percentiles showed a significant decrease.

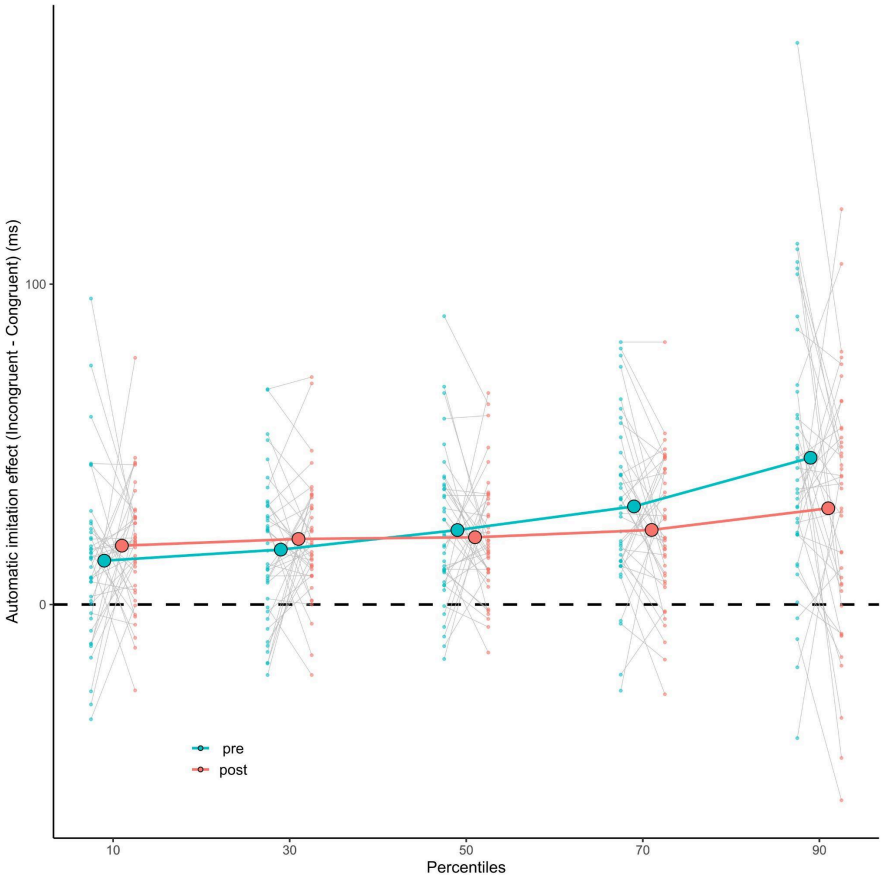

Figure 7: Automatic imitation relationship across Session X Percentiles at 0 ms delay. Automatic imitation is present in all percentiles during Pre-training sessions. No significant differences in automatic imitation effects between Pre- and Post-training have been detected at any of the percentiles.

Even though no effect of the session has been detected when directly comparing Pre- and Post-training automatic imitation at each percentile, the ANOVA shows an interaction between SESSION and PERCENTILE, in which the automatic imitation effect difference between Pre- and Post-training grows over the percentiles (Figure S5). These results suggest that assuming a linear trend, responses longer than the deadline would have increased the difference between Pre- and Post-training automatic imitation.

Eventually, as did for Pre- versus Post-training comparison at 0 ms delay, we computed a Bayesian Factor for each percentile to estimate the likelihood that null

results hid a higher automatic imitation for the Pre- compared to Post-training session (Table 9).

Table 9

Percentile	BF+0	error
10th	0.097	~ 0.002
30th	0.100	~ 0.002
50th	0.269	~ 6.259×10-6
70th	1.193	~ 2.120×10-5
90th	1.871	~ 7.553×10-5

Results from the Bayesian Factor show that the evidence for the null hypothesis (Pre- is not greater than Post-training automatic imitation) is strong (BF+0 <= 0.1) until the 30th percentile and moderate at the 50th (between 0.1 and 0.33). At the 70th and the 90th percentiles, the evidence instead changes the likelihood towards the alternative hypothesis; however the evidence is only anecdotal (BF+0 between 1 and 3).

4. DISCUSSION

In the present study, we tested the time course of the automatic imitation effect by employing a forced-choice reaction time paradigm in which we parametrically varied the delay between the task-relevant event (a colored cue) and the task-irrelevant event (the observed action) before and after an imitative or a counter-imitative training. This allowed us to investigate whether counter-imitative training abolished or at least reduced the automatic imitation effect. In particular, we aimed to assess whether the putative abolishment or reduction was detected at every delay between the observed action and the colored cue, where automatic imitation was present during the Pre-training session.

Our results showed that the automatic imitation effect is already present when the delay between the colored cue and the action observed is 0 ms and lasts until 400 ms. Strikingly, our results highlighted that the automatic imitation effect was modulated but not abolished after counter-imitative training. We detected a consistent reduction of the automatic imitation effect in the Counter group during the Post-training session. However, this reduction occurred only when the colored cue

was presented between 100 and 400 ms after the action observed. No reduction was detected at 0 ms. For the Imitative group, we did not find any modulatory effect of the imitative training between the Pre- and Post-training sessions. This suggested that the modulation caused by counter-imitative training was training-specific.

These results contrast with those of Heyes et al. (2005), who found a non-significant difference between incompatible and compatible actions after counter-imitative training. To fully evaluate this contrast, it is essential to remember that our and Heyes et al.'s studies differ in some critical features. We employed a test-retest strategy within each training group, while they tested only the equivalent of our Post-training session in the two groups, thus being unable to remove the between-subject variance in their design. Another important difference is that Heyes and colleagues employed a single-choice reaction time paradigm whereby the observed action was used as a go-signal, thus representing a relevant event, while in our procedure, it represented the irrelevant event; this potential difference in attention allocation might have magnified the motor bias determined by the action observation (Chong et al., 2009). However, we did find an automatic imitation effect in both sessions, which was reduced at later delays of the Post-training one of the Counter group, thus suggesting that the attentional allocation was sufficient to produce both the automatic imitation effect and the reduction of the latter due to the counter-imitative training. Finally, our sample was four times greater than the one they employed, potentially making it more sensitive to slight differences between conditions.

Automatic imitation is a behavioral effect thought to arise due to visuomotor mappings that facilitate the corresponding motor representation connected to the action observed (Brass et al., 2000, 2001; Stürmer et al., 2000). ASL states that these mappings arise from contingent visual and motor activity associations. When contingencies vary, the same associative mechanism producing automatic imitation in the Pre-training session abolishes or reduces this effect in the Post-training session (Catmur et al., 2009; Cook et al., 2014; Heyes, 2010).

If this were the case, we would have expected to find that, for each delay at which the automatic imitation effect was found in the Pre-training session of the Counter group, a corresponding abolishment or reduction would have also been detected in the Post-training session. However, our results indicate a reduction in the

automatic imitation effect is present after a counter-imitative training session, but only at some delays where automatic imitation was found during the Pre-training session.

A dual-route model might explain the differential impact of counter-imitative training on the automatic imitation effect. Dual-route models have been applied to many different domains, ranging from attention (Schneider & Shiffrin, 1977) to decision-making (Kahneman & Frederick, 2002) and moral cognition (Greene, 2007). Although these models are often characterized in various (and sometimes incompatible) ways (Evans & Stanovich, 2013), they usually share the minimal assumption that the condition that influences whether one process or mechanism occurs diverges from the conditions that influence whether another occurs.

According to the dual-route model, the time course of automatic imitation could result from the interaction of two different visuomotor mechanisms: a rigid and faster visuomotor mapping, similar to that postulated by the Ideomotor and Motor resonance theories, that would cause the automatic imitation effect, and a more flexible and slower “arbitrary association” mapping (Barchiesi & Cattaneo, 2013), that the counter-imitative training session would have induced. According to this model, action observation after the counter-imitative training would trigger the automatic imitation route, which facilitates motor responses similar to those observed, and the arbitrary association route, which facilitates the opposed motor responses. While the automatic imitation route is fast, the arbitrary association route takes more time to implement.

Similarly, Tagliabue et al. (2000) invoked a dual-route model to account for the disruption of the Simon effect caused by a previous counter-spatial compatibility session. They postulated the interaction between "long-term” and “short-term” memory links. The counter-spatial training did not modify the long-term links, as their effects resulted in the short-term links. The former conceptually corresponds to the rigid visuomotor mapping in our model, while the latter corresponds to our "arbitrary association" mapping.

Although the dual-route model fits our results, a contingency-based explanation for the automatic imitation effect is possible. Indeed, we cannot rule out the possibility that a (significantly) more extended training might also affect the visuomotor mechanism responsible for the early automatic imitation effect. However, the reasons why this should happen are not provided by ASL theory. Furthermore, consider the hypothesis that lifelong Pavlovian-like visuomotor contingencies are

responsible for the automatic imitation effect. In that case, these visuomotor links should be expected to resist a relatively brief counter-imitative training. Contrary to what was hypothesized by ASL theory, our results suggest that the brief counter-imitative training employed in our experimental design does not modify such links. Importantly, our results do not imply that experience does not play a role in the development of the visuomotor mappings responsible for the automatic imitation effect, but indicate that a brief counter-imitative training is insufficient to influence them.

Our results contrast with the TMS experiment by Cavallo et al. (2014), who found that counter-imitative training impacted visuomotor facilitation at all time points. Conversely, they align with the TMS study by Barchiesi & Cattaneo (2013). A note of caution must be taken when comparing results from MEPs and behavioral experiments. While the logic of the present experimental design has been inspired by Barchiesi & Cattaneo (2013) and Cavallo et al. (2014), a direct comparison between the present results and the TMS ones could be not so immediate, as they might measure slightly different constructs (Hetu et al. 2016). For example, when considering the “test-training-test” approach, there is evidence that MEPs amplitude is reversed between effectors after counter-imitative training (see Catmur et al. 2007); however, after a similar training, no inversion of the automatic imitation effect is detected (see also Heyes et al. 2005). Another important distinction is that the TMS experiments mostly rely on passive observation, while in automatic imitation studies, participants actively prepare actions that will eventually be produced at cue appearance.

Finally, it could be argued that the absence of counter-training modulation at 0 ms delay, might be due to a lack of statistical power; while we cannot completely exclude this interpretation, the results from the bayesian factor analysis suggest that the absence of modulation is a true null-effect, with increasing evidence toward for fast responses (10th and 30th percentiles).

Automatic imitation exhibits a specific time course. Exploring this time course allows us to better understand whether and to what extent experience can modulate the automatic imitation effect. We provided evidence that such an effect cannot be reduced to mere sensorimotor contingencies, at least not in the way it has been understood thus far.

REFERENCES

Barchiesi, G., & Cattaneo, L. (2013). Early and late motor responses to action observation. *Social Cognitive and Affective Neuroscience*, *8*(6), 711–719. https://doi.org/10.1093/scan/nss049

Barchiesi, G., & Cattaneo, L. (2015). Motor resonance meets motor performance. *Neuropsychologia*, *69*, 93–104. https://doi.org/10.1016/j.neuropsychologia.2015.01.030

Barchiesi, G., Zazio, A., Marcantoni, E., Bulgari, M., Barattieri di San Pietro, C., Sinigaglia, C., & Bortoletto, M. (2022). Sharing motor plans while acting jointly: A TMS study. *Cortex*, *151*, 224–239. https://doi.org/10.1016/j.cortex.2022.03.007

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, *106*(1–2), 3–22. https://doi.org/10.1016/s0001-6918(00)00024-x

Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between Observed and Executed Finger Movements: Comparing Symbolic, Spatial, and Imitative Cues. *Brain and Cognition*, *44*(2), 124–143. https://doi.org/10.1006/brcg.2000.1225

Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*(4), 1255–1265. https://doi.org/10.3758/s13428-012-0186-0(Campbell & Thompson, 2012)

Catmur, C., Walsh, V., & Heyes, C. (2007). Sensorimotor Learning Configures the Human Mirror System. *Current Biology*, *17*(17), 1527–1531. https://doi.org/10.1016/j.cub.2007.08.006

Catmur, C., Walsh, V., & Heyes, C. (2009). Associative sequence learning: The role of experience in the development of imitation and the mirror system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2369–2380. https://doi.org/10.1098/rstb.2009.0048

Cattaneo, L., Maule, F., Tabarelli, D., Brochier, T., & Barchiesi, G. (2015). Online repetitive transcranial magnetic stimulation (TMS) to the parietal operculum

disrupts haptic memory for grasping: Haptic Working Memory in the Parietal Operculum. *Human Brain Mapping*, *36*(11), 4262–4271. https://doi.org/10.1002/hbm.22915

Cavallo, A., Heyes, C., Becchio, C., Bird, G., & Catmur, C. (2014). Timecourse of mirror and counter-mirror effects measured with transcranial magnetic stimulation. *Social Cognitive and Affective Neuroscience*, *9*(8), 1082–1088. https://doi.org/10.1093/scan/nst085

Chong, T. T.-J., Cunnington, R., Williams, M. A., & Mattingley, J. B. (2009). The role of selective attention in matching observed and executed actions. *Neuropsychologia*, *47*(3), 786–795. https://doi.org/10.1016/j.neuropsychologia.2008.12.008

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: From origin to function. *Behavioral and Brain Sciences*, *37*(2), 177–192. https://doi.org/10.1017/S0140525X13000903

Cook, R., Press, C., Dickinson, A., & Heyes, C. (2010). Acquisition of automatic imitation is sensitive to sensorimotor contingency. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(4), 840–852. https://doi.org/10.1037/a0019256

Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., Radkova, I., Deschrijver, E., & Brass, M. (2018). Automatic imitation: A meta-analysis. *Psychological Bulletin*, *144*(5), 453–500. https://doi.org/10.1037/bul0000143

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323; author reply 323-324. https://doi.org/10.1016/j.tics.2007.06.004

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review*, *77*(2), 73–99. https://doi.org/10.1037/h0028689

Guidali, G., Zazio, A., Lucarelli, D., Marcantoni, E., Stango, A., Barchiesi, G., & Bortoletto, M. (2023). Effects of transcranial magnetic stimulation (TMS) current direction and pulse waveform on cortico-cortical connectivity: A registered report TMS-EEG study. European Journal of Neuroscience, 58(8),

3785–3809. https://doi.org/10.1111/ejn.16127

Hétu S, Taschereau-Dumouchel V, Meziane HB, Jackson PL and Mercier C (2016) Behavioral and TMS Markers of Action Observation Might Reflect Distinct Neuronal Processes. Front. Hum. Neurosci. 10:458. doi: 10.3389/fnhum.2016.00458

Heyes, C. (2010). Where do mirror neurons come from? Neuroscie*nce & Biobehavioral Reviews*, *34*(4), 575–583. https://doi.org/10.1016/j.neubiorev.2009.11.007

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

Heyes, C., Bird, G., Johnson, H., & Haggard, P. (2005). Experience modulates automatic imitation. Cognitive Brain Research, 22(2), 233–240. https://doi.org/10.1016/j.cogbrainres.2004.09.009

Jamovi project. (2024). Jamovi (Version 2.5). https://www.jamovi.org/. https://www.jamovi.org/

JASP Team (2024). JASP (Version 0.19.0)[Computer software].

Jeffreys, H. (1961). Theory of Probability (Oxford University Press)

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In Heuristics and biases: T*he psychology of intuitive judgment* (pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Keysers, C., Gazzola, V., & Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. Nature Neuroscience, 23(7), 788–799. https://doi.org/10.1038/s41593-020-0660-4

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Perugini, M., Gallucci, M., & Costantini, G. (2018). A Practical Primer To Power Analysis for Simple Experimental Designs. *International Review of Social Psychology*, *31*(1), 20. https://doi.org/10.5334/irsp.181

Press, C., Gillmeister, H., & Heyes, C. (2007). Sensorimotor experience enhances automatic imitation of robotic action. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1625), 2509–2514. https://doi.org/10.1098/rspb.2007.0774

Prinz, W. (1990). A common-coding approach to perception and action. In *Relationships between perception and action: Current approaches* (pp. 167–201). Springer. https://hdl.handle.net/11858/00-001M-0000-002B-F881-9

Prinz, W. (2005). An Ideomotor Approach to Imitation. In *Perspectives on imitation: From neuroscience to social science: Vol. 1: Mechanisms of imitation and imitation in animals* (pp. 141–156). MIT Press.

R Core Team. (2022). *R: The R Project for Statistical Computing*. https://www.r-project.org/. https://www.r-project.org/

Ray, E., & Heyes, C. (2011). Imitation in infancy: The wealth of the stimulus: Imitation in infancy. *Developmental Science*, *14*(1), 92–105. https://doi.org/10.1111/j.1467-7687.2010.00961.x

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*(4), 264–274. https://doi.org/10.1038/nrn2805

Rizzolatti, G., & Sinigaglia, C. (2016). The mirror mechanism: A basic principle of brain function. *Nature Reviews Neuroscience*, *17*(12), 757–765. https://doi.org/10.1038/nrn.2016.135

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence Effects With Manual Gestures and Postures: A Study of Imitation. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1746-1759. https://doi.org/10.1037/0096-1523.26.6.1746

Tagliabue, M., Zorzi, M., Umiltà, C., & Bassignani, F. (2000). The role of long-term-memory and short-term-memory links in the Simon effect. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 648–670. Scopus. https://doi.org/10.1037/0096-1523.26.2.648

Ubaldi, S., Barchiesi, G., & Cattaneo, L. (2015). Bottom-Up and Top-Down Visuomotor Responses to Action Observation. *Cerebral Cortex*, *25*(4),

1032–1041. https://doi.org/10.1093/cercor/bht295

Wickham, H. (2016). *Ggplot2*. Springer Science+Business Media, LLC.

Zazio, A., Barchiesi, G., Ferrari, C., Marcantoni, E., & Bortoletto, M. (2022). M1-P15 as a cortical marker for transcallosal inhibition: A preregistered TMS-EEG study. *Frontiers in Human Neuroscience*, *16*. https://www.frontiersin.org/articles/10.3389/fnhum.2022.937515

Conclusions and Future Directions

The present work set out to investigate covert motor processes—specifically motor imagery and action preparation—through the development and application of a novel methodological framework: the multidimensional analysis of motor-evoked potentials (MultiMEP). Across the empirical studies presented here, a consistent theme has emerged: while traditional 1m1a designs have provided important insights, their univariate logic severely constrains the type and richness of information that can be retrieved. By contrast, a multivariate decoding approach allows us to overcome these limitations, offering a more sensitive and ecologically valid window into the structure of motor representations.

In the first study, I showed that MultiMEP decoding enables reliable classification of complex imagined actions at the single-trial level, far beyond what standard analyses typically reveal. This expands the scope of TMS research from simple finger movements to more ecological actions, and cross-classification between overt execution (MultiEMG) and motor imagery (MultiMEP) provided new evidence for shared representational substrates. The second study extended this approach to action preparation, demonstrating that MultiMEP decoding remains robust even when participants prepared multiple, highly similar actions. Although accuracy was lower than in motor imagery, results were consistently above chance, indicating that subtle preparatory states can also be captured.

The third study shifted to a behavioral perspective, examining automatic imitation in the context of action observation and visuomotor training. Findings revealed that counter-imitative training modulates—but does not abolish—the automatic imitation effect, and that this modulation unfolds dynamically over time. These results inform ongoing debates between associative learning accounts and dual-route models of automatic imitation. Although this study did not employ MultiMEP directly, it complements the neurophysiological findings by emphasizing the importance of integrating behavioral, neurophysiological, and computational approaches in the study of covert motor processes.

Taken together, the three studies advance our understanding in several ways. First, they demonstrate that covert motor processes can be investigated with greater ecological validity when moving beyond univariate approaches. Second, they validate MultiMEP decoding as a powerful and flexible tool for capturing distributed motor representations at the single-trial level. Third, they open avenues for applying this method to refine theoretical models of covert motor activity and to inform translational work in clinical and brain–computer interface contexts.

Future research should extend MultiMEP to the domain of action observation, where effector- and direction-specific modulations have been documented but only through 1m1a designs. Decoding multidimensional corticospinal patterns during the observation of complex actions would enable a more fine-grained characterization of the mirror mechanism and clarify the extent to which observation recruits shared motor representations. Moreover, combining MultiMEP with TMS–EEG or neuroimaging could bridge levels of analysis, linking corticospinal states with cortical network dynamics.

In conclusion, this thesis makes both methodological and theoretical contributions. Methodologically, it introduces and validates MultiMEP decoding as a novel tool for probing covert motor processes with ecological and analytical depth. Theoretically, it demonstrates that covert motor states are structured and decodable at a finer level of granularity than previously assumed, challenging simplistic mappings between muscles and actions. The next step will be to extend this approach to action observation and other ecologically relevant contexts, thereby advancing our scientific understanding of motor cognition.