Mapping speech kinematics to brain activity with concurrent and time-registered measures of speech movements and brain function

Ioanna Anastasopoulou[1]*,

Douglas O. Cheyne[2,3], Pascal van Lieshout[2],

Blake W Johnson[1]*

[1]Macquarie University, School of Psychological Sciences, Sydney, New South Wales, Australia

[2]University of Toronto, Department of Speech-Language Pathology, Toronto, Ontario, Canada.

[3]Hospital for Sick Children Research Institute, Toronto, Ontario, Canada.

* Corresponding authors

Email: ioanna.anastasopoulou@hdr.mq.edu, blake.johnson@mq.edu.au

Running head: Neuroimaging of speech motor control

Keywords: magnetoencephalography (MEG), speech movement tracking, speech production, speech motor control

Abstract

Articulography and functional neuroimaging are two major tools for studying the neurobiology of speech production. Until now, however, it has generally not been feasible to use both in the same experimental setup because of technical incompatibilities between the two methodologies. Here we describe results from a novel articulography system dubbed Magneto-articulography for the Assessment of Speech Kinematics (MASK; Alves et al., 2016), which is technically compatible with magnetoencephalography (MEG) brain scanning systems. In the present paper we describe our methodological and analytic approach for extracting brain motor activities related to key kinematic and coordination event parameters derived from time-registered MASK tracking measurements (Anastasopoulou et al., 2022). Data were collected from ten healthy adults with tracking coils on the tongue, lips, and jaw. Analyses targeted the gestural landmarks of reiterated utterances /ipa/ and /api/, produced at normal and faster rates (Anastasopoulou et al., 2022; Van Lieshout, 2007). The results show that (1) Speech sensorimotor cortex can be reliably located in peri-rolandic regions of the left hemisphere; (2) mu (8-12 Hz) and beta band (13-30 Hz) neuromotor oscillations are present in the speech signals and contain information structures that are independent of those present in higher-frequency broadband noise signals associated with overt speech movements in the MEG scanner; and (3) kinematic parameters of speech movements can be mapped on to neuromagnetic brain signals using multivariate pattern analytic techniques. These results show that MASK provides the capability, for the first time, for deriving subject-specific articulatory parameters, based on well-established and robust motor control parameters, in the same experimental setup as the brain recordings and in temporal and spatial co-register with the brain data. The co-registered MASK data improves the precision and inferential power of MEG measures of speech-related brain activity compared to previous methodological approaches. This new capability for measuring and characterising speech movement parameters, and the brain activities that control them, within the same experimental setup, paves the way for innovative cross-disciplinary studies of neuromotor control of human speech production, speech development, and speech motor disorders.

Introduction

In recent years, systematic studies of speech motor control in the human brain have significantly expanded our understanding of the neural foundations of expressive speech. Converging evidence now points to a comprehensive re-evaluation of conventional and long-held theoretical models of speech production (see recent reviews by Hickok & Venezia, 2023 and Silva et al., 2022). This re-evaluation is ongoing and rapidly evolving, but there is an emerging shift away from the traditional Wernicke-Geschwind model's emphasis on Broca's region in the left hemisphere, and towards a greater recognition of the computational roles and network connections of various premotor, motor, sensory, and insular regions of the cerebral neocortex (Hickok & Venezia, 2023).

Much of the information that informs these new models comes from non-invasive neuroimaging techniques, predominantly functional magnetic resonance imaging (fMRI) (Bohland & Guenther, 2006; Peeva et al., 2010; Pang et al., 2011; Behroozmand et al., 2015; Rong et al., 2018; Tourville et al., 2019; Heim & Specht, 2019) and associated techniques including diffusion tensor imaging (DTI) (Catani & Forkel, 2019; Chang et al., 2020; Janssen et al., 2022).

Non-invasive electrophysiology with electroencephalography (EEG) and magnetoencephalography (MEG) have added important detail regarding timing of neuronal processing events (Munding et al., 2016; Salmelin et al., 2019; Leckey & Federmeier, 2019). Finally, recent years have provided an increasing amount of very highly detailed electrophysiological evidence from invasive electrophysiological (electrocorticography; ECoG) recordings of speech motor regions in neurosurgical patients (Bouchard et al., 2013; Ramsey et al., 2018; Chartier et al., 2018; Silva et al., 2022).

Current evidence is clear that spoken language processing draws on a complex set of neural computations performed in a widely distributed set of brain regions (Levelt et al., 1998; Munding et al., 2016; Carota et al., 2022). These computations range from abstract and high-level aspects of semantics and syntactics to the low-level sensorimotor processes that directly control and modulate the overt movements of speech articulators of the peripheral vocal tract (Indefrey & Levelt, 2000; Tong et al., 2022). Experimental and clinical protocols for mapping of expressive speech centres therefore employ a wide variety of speech tasks according to their specific experimental or clinical aims (Salmelin et al., 2019). Speech tasks can be variously deployed to emphasise different aspects of spoken language processing: Story listening, object naming, rhyming, and covert word production invoke relatively high level linguistic processes and have been shown to reliably activate distributed areas of prefrontal, temporal and parietal cortex, including Broca's area in the left hemisphere (Bowyer et al., 2005; Doesburg et al., 2016; Kadis et al., 2011; Youssofzadeh & Babajani-Feremi, 2019; Correia et al., 2020); while in contrast, non-word/pseudoword tasks are intended to limit the requirements for semantic, syntactic and

attentional processing and elicit neural activity that is more restricted to brain regions associated with sensorimotor processes (Frankford et al., 2021).

The subject of the current paper is set within the context of low-level speech motor control: the phonological, phonetic, and sensorimotor processes that fairly directly control and/or modulate the neuromuscular output to the articulators. In this context, an important methodological limitation of current neuroimaging research is that, with rare exceptions (Chartier et al., 2018; Mugler et al., 2018; Ouyang et al., 2016), researchers obtain little or no information about the actual movements of said articulators. This is a fundamental limitation in light of evidence that speech (and other movements) is encoded in the form of kinematic movement trajectories in neurons in primary motor cortical neurons (Chartier et al., 2018; Conant et al., 2018; Kolasinski et al., 2020). Such essential information is technically difficult to obtain for crucially important articulators (such as the tongue) which are located out of the line of sight within the oral cavity. Unfortunately, specialised electromagnetic and ultrasound articulography techniques that are capable of non-line-of-sight speech tracking are technically incompatible with the scanner environments used for functional imaging with fMRI and MEG (Anastasopoulou et al., 2022).

In the following, we describe our method for linking speech kinematics to brain activity using a novel MEG setup. This setup enables us to simultaneously and accurately measure speech movements and brain function. The system, termed Magnetoencephalography for Assessment of Speech Kinematics (MASK), can track the independent motion of up to 12 lightweight coils that are similar in size and shape to the tracking coils used in conventional electromagnetic articulography (EMA). In contrast to the passive induction coils used in EMA, MASK coils are actively energized by sinusoidal currents, and their associated magnetic fields are measured by the MEG sensors.

To distinguish between the coil fields and brain activities, we drive the tracking coils at frequencies higher than 200 Hz, which allows us to separate coil fields from brain activities that primarily occur at frequencies lower than 100 Hz. To determine the coil positions we use the same computational algorithms used in conventional MEG to localize and track head positioning and movement. Line-of-sight is not required, permitting tracking of all oral articulators, including the tongue.

MASK evaluates coil positions every 33 ms for movement tracking at rates up to 50 cm/s (Alves et al., 2016). Spatial accuracy depends on the distance of the tracking coils from the MEG sensor array. For coils that are close to the array (like those on the tongue), the accuracy is less than 1 mm relative position error, similar to the standard MEG head position indicator coils. However, for coils that are more distant from the helmet sensor array (like those on the lower lip), spatial accuracy decreases non-linearly to approximately 1-2 mm.

In a previous paper we have described in detail movement parameters (amplitude, duration, velocity) and interarticulator phase relationships derived from direct MASK measurements of articulator movements; and have demonstrated that MASK reliably characterizes key kinematic and movement coordination parameters of speech motor control with a resolution that is comparable to standard electromagnetic articulography devices (Anastasopoulou et al., 2022). In the present work we proceed to describe our methodology for separating speech-related brain signals from MASK signals; for localising these in space and time-frequency, and for establishing a mapping between kinematic parameters and brain activities.

Methods

Participants. Ten healthy adults participated in this study (4F; mean age 32.5, range 19.7- 61.8; all right-handed as assessed by the Edinburgh Handedness Inventory). All participants were fluent speakers of English; Nine were native English speakers, one participant's first language was Mandarin. All procedures were approved by the Macquarie University Human Research Ethics Committee.

MEG scans. Speech tracking data and neuromagnetic brain activity were recorded concurrently with a KIT-Macquarie MEG160 (Model PQ1160R-N2, KIT, Kanazawa, Japan) whole-head MEG system consisting of 160 first-order axial gradiometers with a 50-mm baseline (Kado et al., 1999; Uehara et al., 2003). MEG data were acquired with analogue filter settings as 0.03 Hz high-pass, 1000 Hz low-pass, 1000 Hz sampling rate and 16-bit quantization precision. Measurements were carried out with participants in supine position in a magnetically shielded room (Fujihara Co. Ltd., Tokyo, Japan).

Structural scans. T1-weighted anatomical magnetic resonance images (MRIs) were acquired for all participants in a separate scanning session using a 3T Siemens Magnetom Verio scanner with a 12-channel head coil. Those anatomical images were obtained using 3D GR\IR scanning sequence with the following acquisition parameters: repetition time, 2000 ms; echo time, 3.94 ms; flip angle, 9 degrees; slice thickness, 0.93 mm; field of view, 240 mm; image dimensions, 512 × 512 × 208.

Procedure. Five head position indicator coils (HPI) were attached in the head in an elastic cap, and their positions were measured at the beginning and at the end of the experiment, with a maximum displacement criterion of < 5 mm in any direction. The coils' positions with respect to the three anatomical landmarks (nasion, right and left preauricular landmarks) were measured using a handheld digitiser (Polhemus FastTrack; Colchester, VT).

MASK coils were placed at mid-sagittal positions on the vermilion border of the upper lip (UL) and lower lip (LL), the tongue body (TB; 2 cm from the tongue tip) and the lower incisor (JAW) sensor which was attached to a thin thermoplastic mould. Tongue sensors were attached with surgical glue (Epiglu, MajaK Medical Brisbane; Australia), while lip sensors were attached with surgical tape.

MEG time-aligned speech was recorded in an auxiliary channel of the MEG setup with the same sample rate (1000 Hz) as the MEG recordings. An additional speech recording was obtained with an optical microphone (Optoacoustics, Or-Yehuda, Israel) fixed on the MEG dewar at a distance of 20 cm away from the mouth of the speaker; and digitised using a Creative sound blaster X-Fi Titanium HD sound card (Creative, Singapore) with 48 kHz sample rate and 24-bit quantization precision. The higher resolution recordings were aligned off-line using the MATLAB

alignsignals function to the MEG speech channel to bring them into time register with the neuromagnetic data.

Experimental protocol: Participants performed four speech production tasks and one manual button press task. Speech productions were non-word disyllabic sequences with a V1CV2 structure, /ipa/ and /api/, and each was produced in a reiteration fashion at normal and faster rates. The tongue and lip gestures for /ipa/ and /api/ are mirror reversed in phase, providing a robust behavioural contrast in terms of interarticulator coordination (Anastasopoulou et al., 2022). Variations in speech rate were used as a control variable to examine the intrinsic stability of the coordination (Kelso 1986, Van Lieshout et al., 1996). Asking participants to change their speaking rate (frequency of executed movements) is a typical characteristic of studies which investigate coordination dynamics (Kelso, 1995, Van Lieshout et al., 1996). The same reiterated stimuli have been used in previous studies investigating speech motor control strategies in normal and in disordered populations (Van Lieshout et al. 1996; Van Lieshout et al. 2002; Van Lieshout et al. 2007; Van Lieshout, 2017). Nonword stimuli with no linguistic information avoid familiarity issues (Van Lieshout, 2017) and have been widely used in the literature to investigate normal and pathological function in speech motor control (Murray et al., 2015; Case & Grigos, 2020).

Participants performed manual button presses on a response pad with the index finger of their dominant hand at a self-paced rate of about 1 per 2 seconds.

Participants were presented with a fixation cross on a display screen and instructed to take a deep breath. The stimulus nonword then appeared on the screen for 12 sec. For the normal rate production, participants were required to utter productions at a normal, comfortable rate as they would do while conversing with a friend, until the stimulus nonword disappeared from the screen. For the faster rate, they were instructed to produce the stimuli as fast as possible while maintaining accuracy (Van Lieshout et al. 2002). Following Van Lieshout (2007), we refer to the reiterated productions generated within the span of a breath intake as a “trial set”. A short break was provided after each trial set. Participants generated about 10 individual productions in each normal rate trial set and about 12 individual productions in each faster rate trial set. Since 100+ individual trials (in this case, individual nonword productions) are typically required for downstream analyses of MEG data, the number of trials was increased to 10 trial sets at each rate. The recorded productions of a female native Australian speaker were used to train the participants before data acquisition began. Participants were trained to avoid incorrect speech productions or head movements, were required to produce each task correctly at the correct rate before data acquisition began and to avoid eye blinking during the speech production trial sets. They were allowed to blink their eyes between the trial sets but were instructed to avoid these during trial sets.

For the manual nonspeech task, participants performed self-paced button presses at a rate of about 1/ 2 seconds on a fibre optic response pad (Current Designs, Philadelphia) for 180sec (see Figure 1B). (Cheyne et al., 2008; Cheyne et al., 2014; De Nil et al., 2021; Johnson & He, 2019).

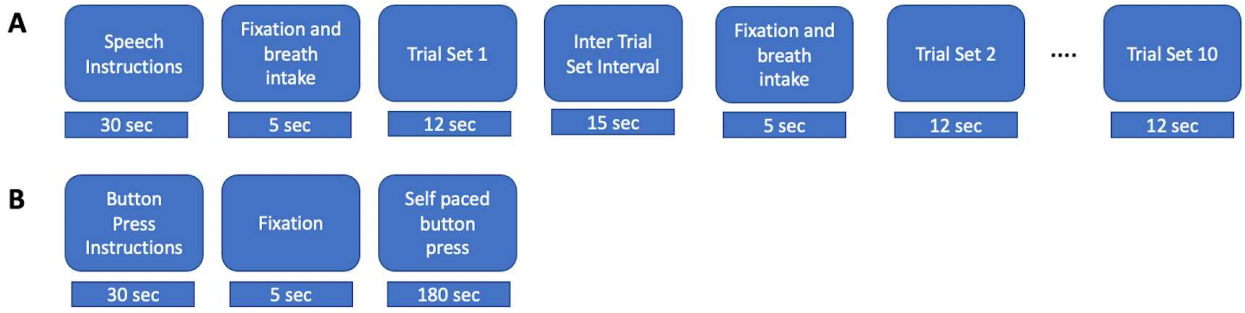


Figure 1. Experimental procedures. **A.** Speech task. Instructions were displayed for 30s, followed by a 5s fixation cross '+' and breath intake in preparation for the speech production trial set. During a trial set participants produced the indicated nonword in a reiterated fashion for 12s. 10 consecutive trial sets were performed for each nonword stimulus. **B.** Button press task. Instructions were displayed for about 30s followed by a fixation cross, during which participants performed self-paced button pressed with the index finger of their dominant (right) hand at a rate of about 1 per 2 seconds for a total of about 90 trials.

Analyses: Data analyses proceeded in four main phases:

(1) Analyses of MASK speech movement signals to characterise speech kinematic profiles;

(2) MEG source reconstruction to identify location of speech motor cortex;

(3) Extraction of MEG-time frequency spectrograms from source-localised speech motor cortex, followed by and multivariate pattern analysis of speech-relevant brain rhythms.

(4) Mapping of speech kinematic profiles onto source- and frequency-constrained MEG data, via representational similarity analysis (RSA).

For the purposes of clarity, we present details of each set of analytic methods along with their results, organised according to these four analytic phases.

1. Analysis of MASK-derived speech movement signals.

1A. Methods. The raw MASK position data were head motion corrected using MASK coils placed at fiducial landmarks (nose, left and right ear) and transformed from the MEG coordinate system into the occlusal plane frame of reference such that motion signals could be measured relative to a midsagittal plane defined by the x (posterior-anterior) and z (inferior-superior) axes relative to the bite plane origin. These transformed signals were then analysed using EGUANA software (Henriques & Van Lieshout, 2013; Van Lieshout, 2021) to derive signal amplitude and phase for selected articulators and speech gestures. Movement artifacts were initially screened out, and subsequent analyses were focused on accurate productions (Case & Grigos, 2020), with errors such as substitutions and lengthy pauses being excluded.

/ipa/ and /api/ productions involve specific movements of the lips and tongue. To create the voiceless stop /p/ sound, a bilabial closure (BC) gesture is used. The two tongue body constriction gestures (TB) are used to produce the sounds /i/ and /a/. The BC gesture was calculated using the two-dimensional (x-y) Euclidian distance of the upper and lower lip positions, while the tongue body gesture was derived from the two-dimensional (x-y) Euclidian distance of the tongue body and the nasal reference coil, as described in Van Lieshout et al. (2007). The kinematic and coordination parameters were computed using the methods described in (Anastasopoulou et al., 2022; van Lieshout et al., 2002; van Lieshout et al., 2007).

The opening and closing movements of each cycle were identified using the minimum and maximum vertical position of the gestural and articulatory signals (Van Lieshout, 2017). The amplitude levels of the opening and closing movements at 10% and 90% were determined for each individual cycle. Custom MATLAB scripts were then utilized to determine the times corresponding to the 10% and 90% amplitude levels of each opening and closing movement of each individual gestural signal of BC and TB. Finally, the individual times of the onset and offset of the opening and closing movements were brought into time register with MEG data by aligning the acoustic signal of the MASK acquisition signal and the acoustic signal recorded in the MEG adult acquisition computer, as shown in Figure 2.

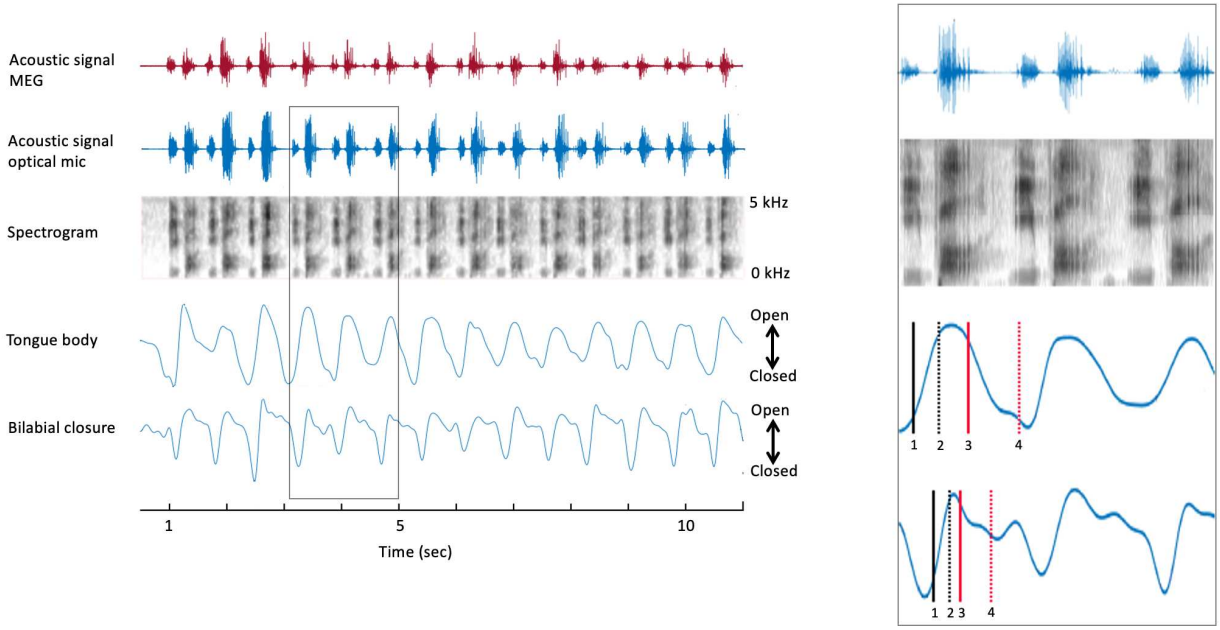


Figure 2: Temporal alignment of MEG and MASK signals. MASK articulatory signals are brought into register with the MEG brain data using the MATLAB *alignsignals* function on the MEG and PsiFi acoustic signals (top and second rows). **Inset** shows enlargement of rectangle bounded area in main figure. 1 = onset of opening movement; 2 = offset of opening movement; 3 = onset of closing movement; 4 = offset of closing movement.

1B. Results

MASK tracking signals. Figure 3 displays acoustic recordings and tracking signals of the tongue body (TB) and bilabial closure (BC) from Participant 1 for a single trial of each of the four speech production tasks. The participant produced 14 utterances of /ipa/ and /api/ at a normal speaking rate and 18-19 utterances at a faster rate.

The contrast between the mirrored positions of the tongue and lips in /ipa/ and /api/ is clearly observed in the MASK measurements of tongue and lip gestures. Peaks and valleys in Figure 3 indicate the high and low positions achieved by the BC and TB gestures during the production of /api/ and /ipa/. Valleys occur during the bilabial constriction gesture and the tongue body gesture for /i/, while peaks occur for the tongue body gesture of /a/. (We note that these positions are Euclidean distances relative to the nasion. Within this reference frame “low a” is a peak, and “high /i/” is a valley). In /api/, the /p/ closure occurs during the upward motion of the TB, going from the low /a/ to the high /i/ position. On the other hand, in /ipa/, the /p/ closure occurs during the downward motion of the TB, going from the high /i/ to low /a/ position. The gestural movements of /ipa/ and /api/ are mirror images, with the relative timing of the motions of TB and BC gestures reversed.

Overall, the TB and BC tracking signals measured with MASK are entirely comparable in morphology and quality with those obtained from a conventional electromagnetic articulography setup (please see Anastasopoulou et al., 2022, for a direct comparison of MASK and EMA signals measured during the same utterances described here).

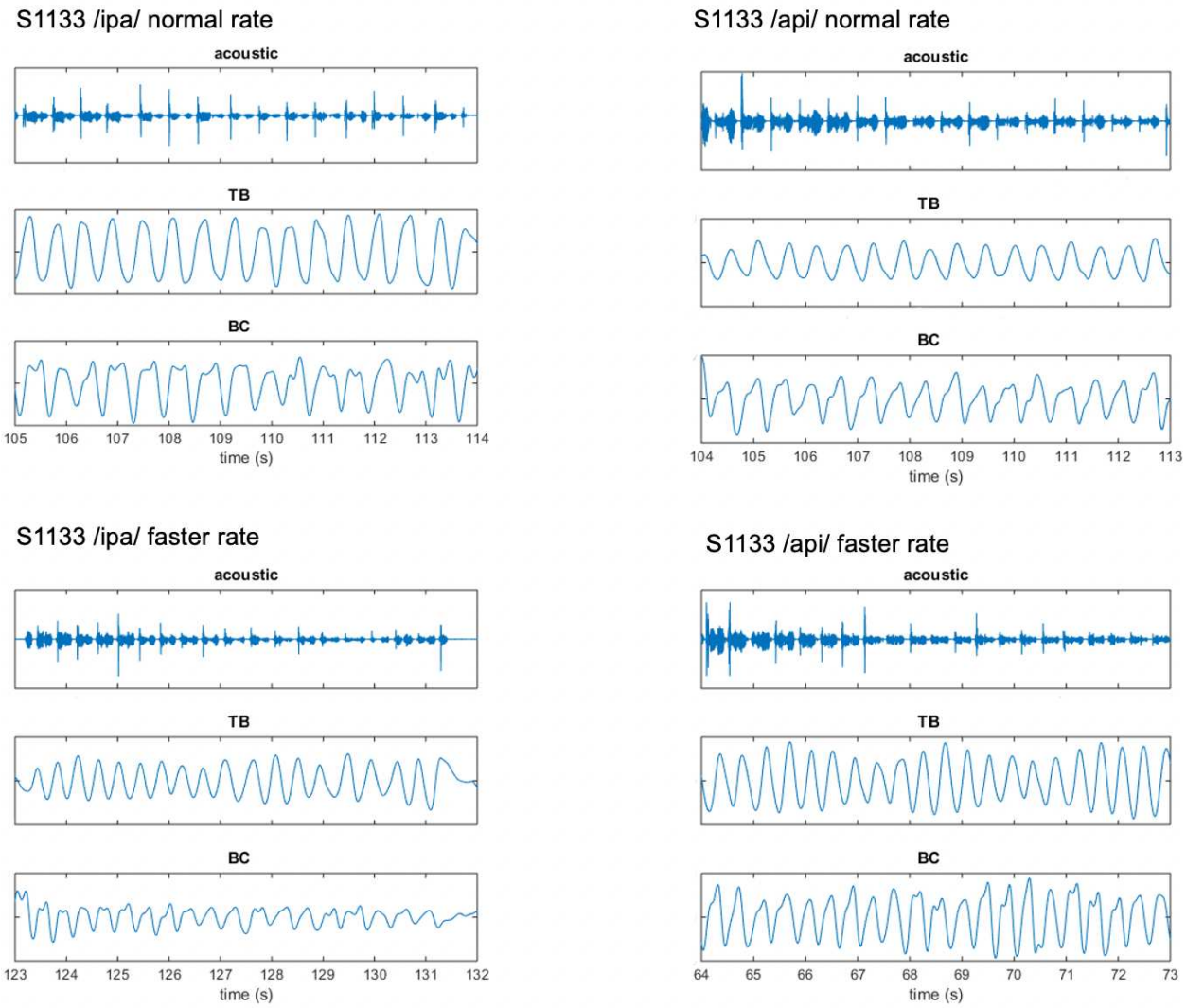


Figure 3. **Representative acoustic and kinematic measurements from MASK**. Data are shown for two participants for a single /ipa/ trial set at normal and faster speaking rates. Shown are (from top to bottom) waveforms for the audio signal, tongue body (TB) gesture, and bilabial constriction (BC) gesture.

Derived kinematic profiles. The next stage in our analysis pipeline involves generating profiles that capture the relationships between key kinematic parameters of BC and TB gestures. Specifically, we examine the amplitudes, durations, velocities, and stiffnesses of gestural movements, as these parameters are known to covary in highly consistent ways and reflect "invariant" properties of speech kinematic movements. These invariant properties are crucial in understanding the motor control of human speech.

Figure 4 illustrates the covariation of these kinematic parameters for /ipa/ and /api/ for two participants. Our analysis shows that movement peak velocity increases as a linear function of movement amplitude, indicating that larger movement distances are associated with higher peak speeds. Furthermore, we observe comparable amplitude/velocity relationships for opening and

closing movements, suggesting that these parameters are controlled similarly regardless of movement direction. This roughly linear relationship between amplitude and peak velocity is a well-established characteristic of speech kinematics, and has been described for a variety of articulators, gestures, and utterances (33,38). Regarding the stiffness vs. duration relationship, our results indicate that stiffness systematically decreases as a curvilinear function of durations less than 200 ms, after which the relationship plateaus into a relatively flat line.

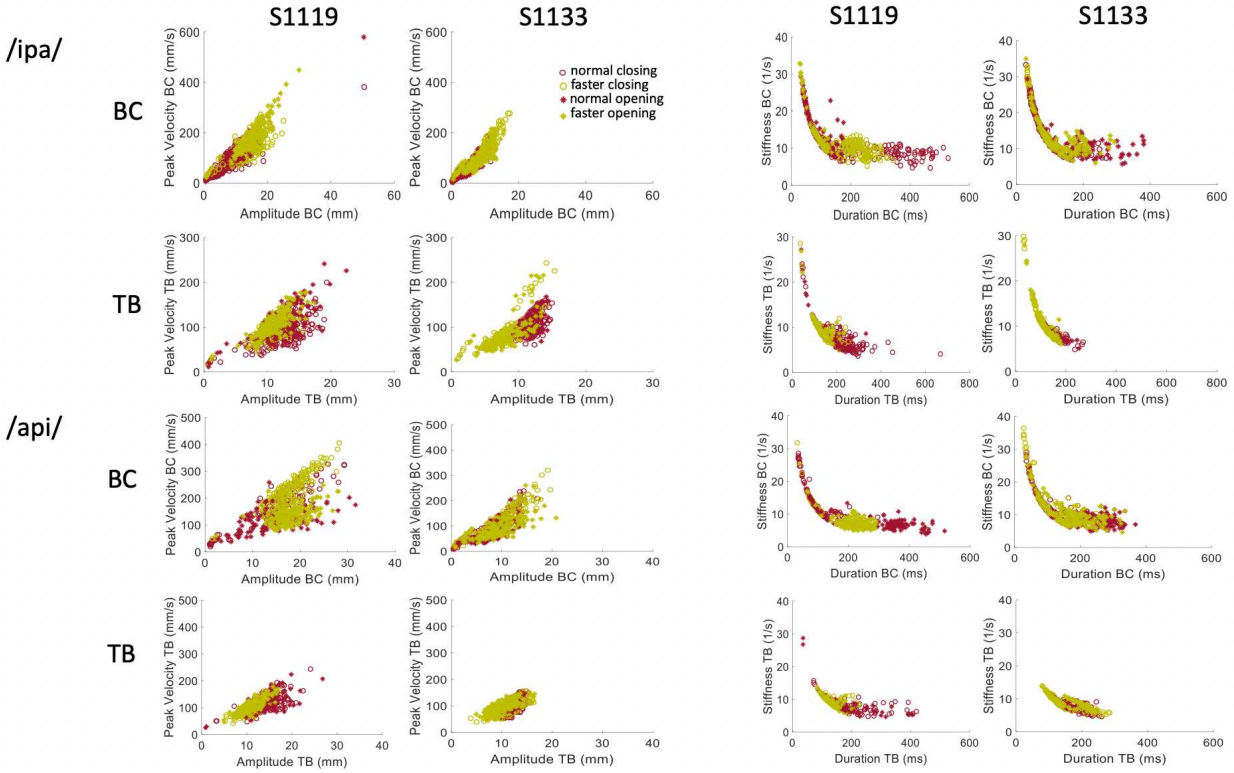


Figure 4. Covariation of kinematic parameters of speech movements for two participants. Left columns: Velocity versus amplitude. Right columns: Stiffness versus duration. BC = bilabial closure. TB = tongue body.

2. MEG source reconstruction of speech motor cortex

***2A. Methods*:** Source reconstruction of brain activity was carried out using the synthetic aperture magnetometry (SAM) beamformer algorithms implemented in the *BrainWave* MATLAB toolbox (Jobst et al., 2018; cheynelab.utoronto.ca/brainwave). The raw KIT/Yokogawa data files were initially converted to CTF format and transformed to the CTF head coordinate system using the fiducial coil positions relative to the sensor array. Each participant's structural MRI was then spatially coregistered with the MEG data and normalised into standard adult MNI template space using SPM12 (Wellcome Institute of Cognitive Neurology).

For the speech motor cortex localiser analyses, the onset of each speech trial set was marked according to the onset of the acoustic signal and raw data were pre-filtered using a 100 Hz low pass bidirectional zero phase-shift Butterworth filter and epoched into 15 sec segments, from 10 sec prior to speech onset to 5 seconds after speech onset. Each 15 sec segment encompassed three distinct task periods: the last five seconds of the preceding trial set (-10 to – 5 sec); the inter-trial set rest period (-5 to 0 sec); and the first five seconds of the current trial set (0 to +5 sec), thereby providing maximal contrast between active (speech) and rest periods.

The current trial set and inter-trial rest period intervals were used for the SAM pseudo-T analysis window and baseline window respectively. We used a sliding active window of 1 second duration starting from 0-1000 ms (step size 200 ms, 10 steps), and a fixed baseline window of 2 seconds duration extending from -5 to -3 seconds relative to speech movement onset and a bandpass of 18–22 Hz (centre of the beta frequency range). The full 15 second time window was used to compute the data covariance matrix for beamformer weight calculations. SAM pseudo-T images were volumetrically reconstructed using a 4 mm resolution grid covering the entire brain.

In all individuals SAM source reconstruction resulted in robust peaks centred on the left middle precentral gyrus and adjacent regions of the left middle frontal gyrus. The time-course of source activity was then computed as the output of the beamformer with optimized orientation ("virtual sensor") and plotted as time-frequency spectrograms (encompassing the entire 15 sec data epoch) to assess the temporal correspondence of beta activity with active and rest periods. To maximise the number of trials (and consequently, the signal to noise ratio) in this analysis we use trial sets from all four speech tasks (for a total of 40 trial sets).

For the hand knob localiser analysis, trials were prefiltered with a bandpass of 0-100 Hz and epoched with respect to the button press onset into 1.5 sec segments (-500 to +1000 ms), encompassing the established time course of beta-band desynchronisation (several hundred ms prior to and after the button press) and "rebound" synchronisation (several hundred ms starting about 500 ms after the button press) (see Cheyne, 2013; Cheyne et al., 2014; Johnson et al., 2016). Following the maximal contrast approach used for the speech analysis, the SAM pseudo-T analysis

used a sliding active window of 200 ms duration starting from 600-800 ms, (step size 10 ms, 10 steps), a fixed baseline window from 0 to 200 ms, and bandpass of 18-22 Hz. The full 1.5 second epoch was used to compute the data covariance matrix for beamformer weight calculations. Volumetric reconstruction used the same grid employed for the speech analysis.

In all individuals SAM source reconstruction resulted in robust peaks centred on the hand knob of the left precentral gyrus, and a smaller mirror source centred on the right hemisphere homologue. The left hemisphere virtual sensor source activity was then computed and plotted as a time-frequency spectrogram (encompassing the entire 1.5 sec data epoch) to assess the correspondence with the established time course of beta band activity associated with the manual button press task (Cheyne, 2013; Cheyne et al., 2014; Johnson et al., 2020).

2B. Results. Figure 6 shows that the SAM beamformer cluster maxima encompass the middle portion of the prefrontal gyrus (mPFG) and the immediately adjacent region of the middle frontal gyrus (MFG), both established areas of low-level speech motor control (Silva et al., 2022). The anatomical localisation of the mPFG is well-supported by comparison with the SAM beamformer map for the button press task, which shows a cluster maximum in the hand knob of the immediately dorsal region of precentral gyrus.

Physiological activities at the locations of the SAM beamformer cluster maxima are visualised in the "virtual sensor" time frequency plots below their respective brain maps. For the button press task the time-frequency plot shows the well-established pattern of beta-band (13-30 Hz) desynchronisation, starting several hundred ms before the button press, persisting for several hundred ms after, and followed by a "rebound" beta synchronisation at about 600-700 ms after the button press.

A comparable pattern of beta band activity is evident in the speech virtual sensor plot, keeping in mind the different time scales (1.5 sec for button press, 15 sec for speech) and movement requirements (a single punctate button press versus 10 seconds of steady state, reiterated speech) of the two tasks. Beta band desynchronisation begins several hundred ms before speech onset and persists for the duration of the speech movements. Note that in this plot the baseline of the colour scaling (the five second inter trial set rest period) was chosen to emphasise event-related desynchronisation. Baselining to the speaking portions of the epoch will emphasise the event-related synchronisation during the rest period).

Taken together, the results of the localisation procedure provide a focussed and well-grounded target for subsequent analyses that can incorporate the kinematic and coordination parameters derived from MASK. The plausibility of the mPFG/MFG target is well-supported by the time-frequency characteristics of the virtual sensor and its anatomic location immediately dorsal to the established landmark of the PFG hand knob, independently localised with data from

the button press task: a task that has been long established to provide highly reliable beta-band activations located in the hand regions of the sensorimotor cortices (e.g. Cheyne et al., 2014).

The speech-related plausibility of the mPFG/MFG region is also strongly supported by recent evidence from invasive neurosurgical studies of expressive speech function. The region of the precentral gyrus located immediately ventral to the hand motor region of the precentral gyrus has been functionally defined in neurosurgical studies and termed middle precentral gyrus (midPrCG). It has been posited that this region functions to coordinate complex phonological sequences into motor plans (Silva et al., 2022). Further, the coactivation of the posterior region of the middle frontal gyrus (pMFG) in our results is to be expected since this region that is tightly functionally associated with the midPrCG (Glasser et al., 2016).

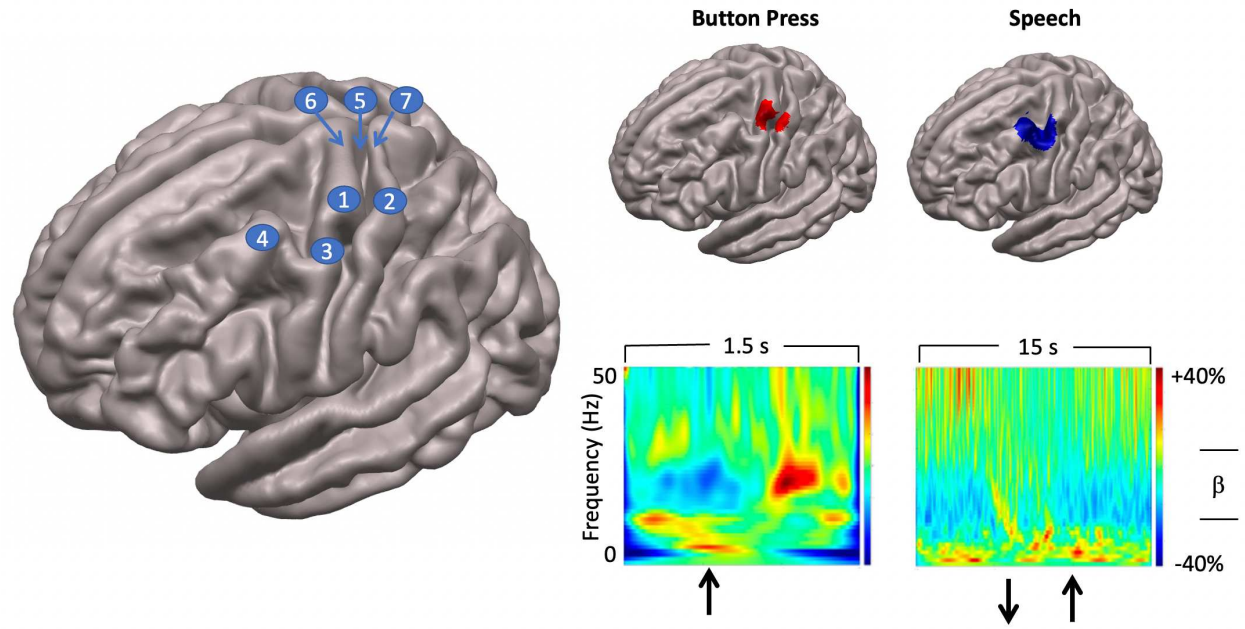


Figure 5. Localisation of speech motor cortex. *Left: Anatomical Landmarks*. 1 – Hand regions of precentral gyrus (hand knob); 2 – Hand region of postcentral gyrus; 3 – Middle precentral gyrus; 4 – Middle frontal gyrus; 5 – Rolandic fissure; 6 – precentral gyrus; 7 – postcentral gyrus. *Top right panel*: *SAM beamformer maps*. Button press task elicited activation of hand region of pre (motor) and postcentral (somatosensory) gyri. Speech task shows maximal activation in middle precentral gyrus, immediately ventral to the hand motor region of the precentral gyrus. Speech activation cluster also encompasses the middle frontal gyrus immediately adjacent to the middle precentral speech region. Bottom right panel: Time-frequency plots showing temporal evolution of oscillatory responses at virtual sensors placed at locations of cluster maxima shown above.

3. Extraction and pattern analysis of source-localised MEG time-frequency spectrograms.

3A. Methods. For each production task, continuously recorded MEG signals were pre-filtered with a bandpass of 0-100 Hz and 50 Hz notch filter and segmented into 4 second epochs (-2 sec to +2 sec) using the onset of the BC opening movement for each speech task as time zero. Data epochs were subsequently truncated to 3 seconds (-1.5 sec to + 1.5 sec) to remove edge effects from the frequency analysis. Using the speech motor cortex coordinates derived from the speech localiser for a virtual sensor, time-frequency spectrograms were generated for each individual trial and the resulting three-dimensional (time x frequency x trial) matrix was exported for classification analysis using the MVPA-Light MATLAB toolbox for classification and regression of multidimensional data (Treder, 2020).

Equivalent duration non-speaking "resting" condition epochs were derived by randomly selecting epoch-reference time-points from the inter- trial set rest periods of the MEG data. For each speaking condition and participant, an equal number of resting condition trials was epoched.

Our aim in this analysis was to perform a "time-frequency classification" to determine if the trial by trial time-frequency data derived from the speech motor cortex virtual sensor contains information that is able to discriminate between the neural activities associated with speech and rest trials; and, if so, to determine if the discriminative information is confined to a specific frequency range[1] (Treder, 2020).

We performed a searchlight analysis using a binary linear discriminant analysis (LDA) classifier and a metric of "accuracy" (fraction of correctly predicted class labels, range = 0 – 1), with training parameters of five folds and five repetitions.

Group level 2 statistics were performed using nonparametric permutation testing and cluster corrections for multiple comparisons (Maris & Oostenveld, 2007) as implemented in the MVPA-Light toolbox (Treder, 2020).

3B. Results. Figure 6 shows source-localised time-frequency spectrograms for individual participant S1119. The speech-condition spectrograms show clear speech rate related modulation of circa 20 Hz beta-band activity in all speech conditions. A clear and distinct pattern of circa 10 Hz mu band activity is also observable for both of the /api/ productions. Substantial movement-related broadband noise is also evident in the supra-beta frequencies for all speaking conditions and is especially prominent in the /api/ faster rate condition. Both the beta and mu-band rhythms are well-known and established rhythms of the central motor cortices (Cheyne, 2013; Cheyne et al., 2014).

[1] In the case of event-related experimental designs the time-frequency classification can also determine if discrimination is confined to specific times (Treder, 2020). The reiterated speech paradigm used here is akin to a system in steady state, so the analytic question at this stage simplifies to frequency discrimination alone.

In contrast to the speaking conditions, for the non-speech resting conditions mu and beta activities are manifest as relatively continuous (unmodulated) bands of activity throughout most of the epoch, and broadband movement-related noise patterns are absent from the resting spectrograms.

The MVPA classification results of Figure 9 shows that the classification show up both the mu and beta band rhythms, with a well-defined frequency boundary between the two rhythms that is clear and prominent in the cases of the /ipa/ faster rate and /api/ normal rate date. The classifier also picks up speech-movement related noise (from the speech condition), particularly in the /api/ faster rate condition. Speech movement-related noise is evident as high-frequency broad-band patterns extending to circa 50-60 Hz. The broadband noise in the classification patterns is well separated in frequency from the beta/mu classifier signals and confined to frequencies above 50 Hz.

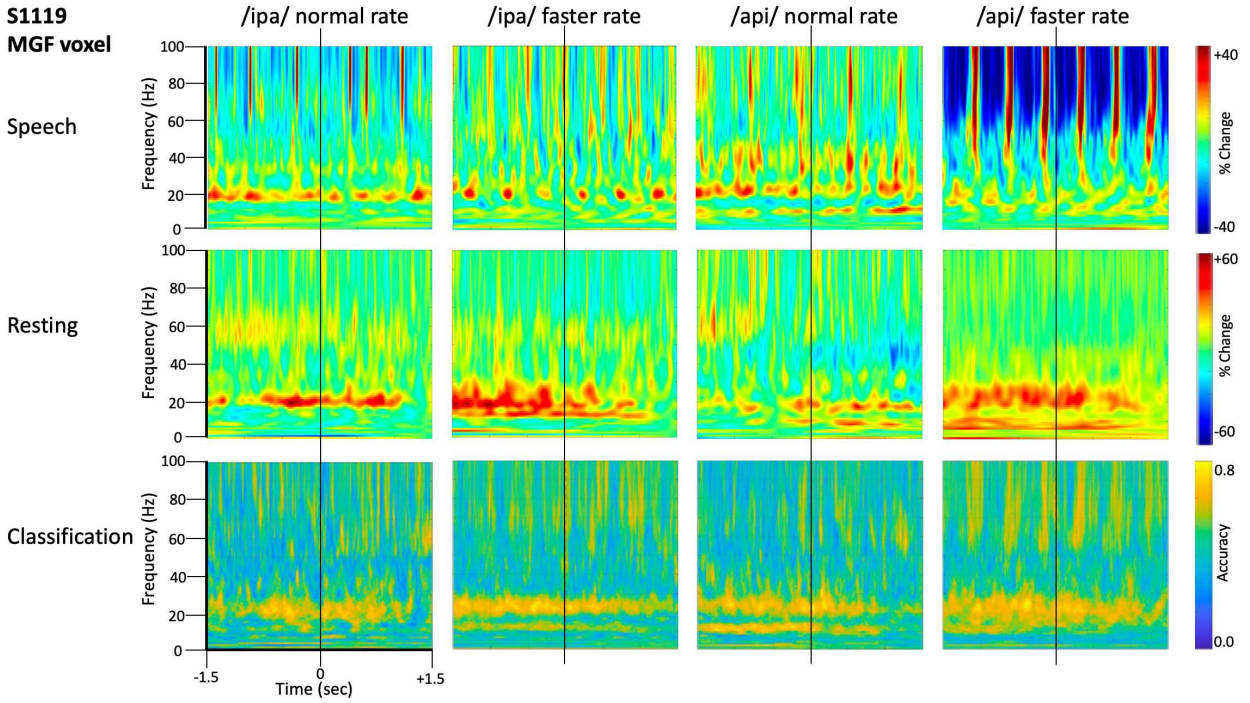


Figure 6. Time-frequency characteristics of speech and resting conditions for an individual participant. All plots show three seconds of MEG data derived from the medial frontal gyrus voxel. *Top row*. Time-frequency spectrograms during speech. Data are epoched relative to the onset of the bilabial closure opening movement. Speech rate modulated beta-band (circa 20 Hz) activity is evident in all plots, and mu-band activity is evident in several, especially the /api/ normal rate condition. *Middle row*. Spectrograms derived from the inter-trial set rest periods. Relatively continuous beta-band ERS is evident in all plots, as well as mu-band ERS in the /api/ conditions. *Bottom row*. MVPA classification results for speech versus resting conditions. High classification accuracy is quite tightly constrained to circa 20 Hz beta band, and a well-defined mu-frequency band is evident in the /ipa/ faster rate and /api/ normal rate conditions.

Level 2 group analysis of the speaking/resting classifier results are shown in Figure 7. The group results are entirely consistent with the individual results described above and provide clear

statistical support for high classification accuracy for the mu and beta motor rhythms, as well as the high-frequency movement-related noise region. The mu/beta frequencies are well separated by by a region of low-classifier accuracy for circa 30-50 Hz frequencies, suggesting at least a lack of continuity between these frequency regions, and possibly that the underlying informational structures of the motor rhythms and the noise are functionally independent. We consider this issue more formally in the frequency generalisation analyses below.

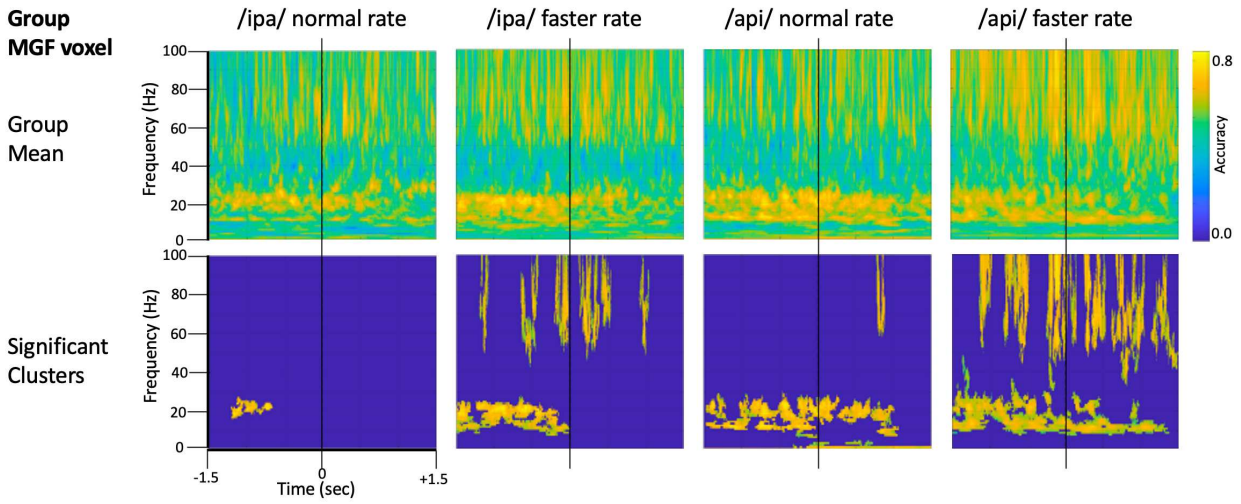


Figure 7. Group analysis of classifier performance for speech versus resting conditions. All plots show three seconds of MEG data derived from the medial frontal gyrus voxel. As seen in Figure 7, classifier performance is high for both speech-related high-frequency noise and beta/mu signals.

Frequency generalisation (cross-frequency decoding).

In addition to providing an estimate of decodability for the task contrasts described above, time/frequency series decoding can be applied to provide a picture of the continuity (or discontinuity) of decoding estimates over time or frequency. This provides an important inferential advantage for further interpretation of the timing or frequency specificity of experimental effects. This "cross-decoding approach" involves training the classifier on a given time or frequency and then testing classifier performance on different times or frequencies. The logic of this approach relies on the classifier's ability to partition multidimensional space as a basis for discriminating between experimental conditions: hence, where a classifier trained on a given time or frequency can successfully discriminate experimental classes based on other time or frequency points, one can infer that the structure of the underlying multidimensional space is similar for those two points. Conversely, in the case where cross-point decoding is unsuccessful, one can infer that the underlying multidimensional patterns are sufficiently different that the distinction between class labels determined at one point are not meaningful for discrimination at the second point (Grootswagers et al., 2017; Treder, 2020).

In the present context cross-frequency decoding enables us to more precisely address questions about possible relations between frequency bands identified by the basic speech/rest classificational analyses described in Figures 7 and 8: (1) Do the beta/mu band signals rely on the same classification information as the high frequency signals, which we interpret to reflect non-neural noise associated with speech movements? In this case redundancy would suggest the beta/mu signals simply contain some level of speech movement noise that is the basis for class discrimination. On the other hand, a lack of cross frequency generalisation supports the conclusion that they are independent signals and further, that the beta/mu signals are not significantly contaminated by broadband noise. (2) In a similar fashion, it is of interest to assess the cross-frequency generalisation between the mu and beta bands, two motor rhythms that have been frequently observed to co-occur in electrophysiological studies and can be assumed to have some functional inter-relationship (Cheyne et. al., 2014).

The frequency generalisation results of Figure 8 provide clear answers to both questions. First, there no evidence for frequency generalisation between mu/beta and the high frequency noise region, at either the individual or group level. To the contrary, frequency generalisation (observable as off-diagonal clustering) occurs within the sub-30 Hz mu/beta frequencies, and within the supra-60 Hz frequencies (particularly within the range of about 60-80 Hz); but the intermediate zone between mu/beta and high frequency noise (circa 30-60 Hz) exhibits a fairly strictly diagonal trajectory (for example, see group means for /api/ normal and faster rates). Classification of speech and resting conditions can clearly rely on either high frequencies associated with speech movements, or mu/beta frequency information: both frequency regions are prominent in the classification plots of Figures 7-9. However, the cross-frequency coding results provide clear support for the conclusion that the high frequency (noise) band and the mu/beta bands are discontinuous and rely on distinct patterns of multidimensional structure within their data to achieve discrimination between speech and resting data conditions.

On the second question, the group results show clear frequency generalisation between beta frequencies circa 15-Hz and also suggest a possibly weaker generalization for beta training frequencies and mu test frequencies (circa 8-12 Hz; see group statistical results for /ipa/ and /api/ faster rates. Although the mu/beta clusters do not achieve statistical significance for the slower speech rates, comparable clusters are evident in their group mean data. The individual results for S1119 are entirely comparable to the group mean data but show a much clearer distinction between the mu and beta bands (see especially plot for /api/ normal rate): These data show that mu training frequencies about 8-12 Hz generalise to beta band frequencies; and that beta training frequencies circa 20-30 Hz generalise to mu test frequencies. However, the plots also show a clear mu/beta discontinuity, with low classification accuracy for frequencies between about 13-20 Hz. This

mu/beta discontinuity is not as evident in the group results, presumably due to individual differences in the precise frequency ranges of the mu rhythm (Pfurtscheller et al., 1997).

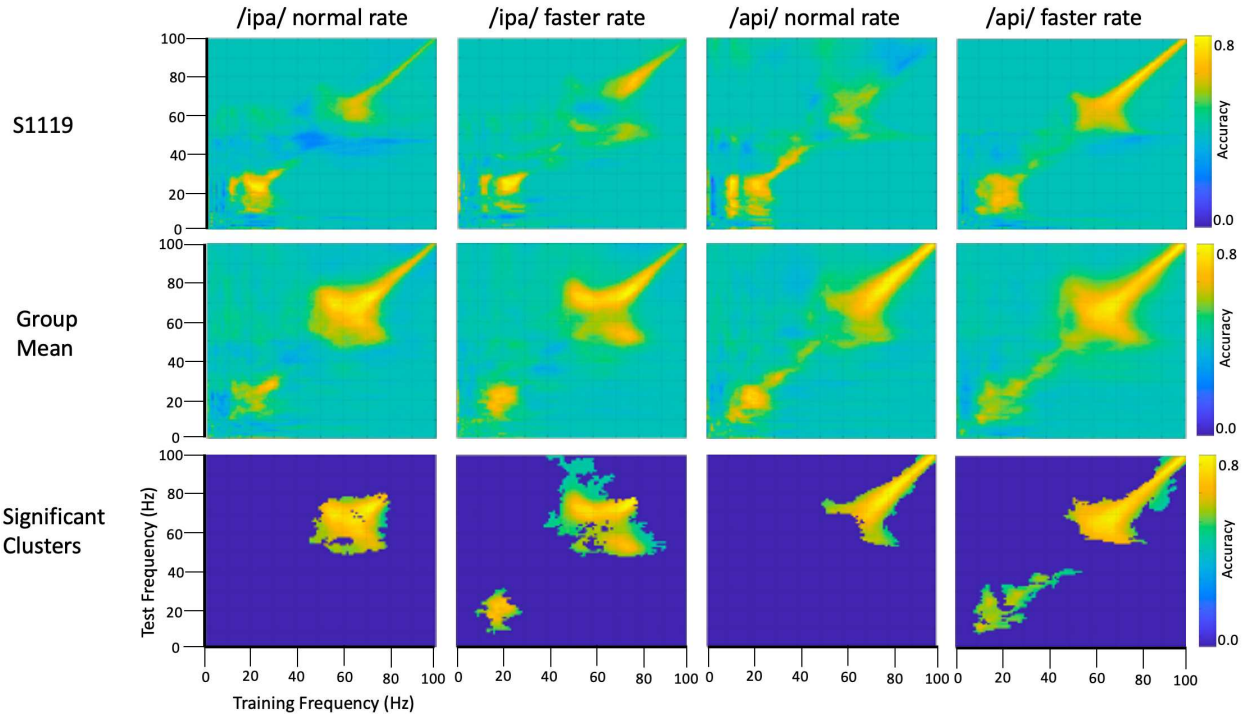


Figure 8. Frequency generalisation. In this analysis the classifier is trained on a given frequency and decoding performance is tested on a different frequency. This is repeated for all possible frequency pairs. The classifier results show that beta frequencies generalise to each other and to some extent to mu frequencies (bottom row, /ipa/ faster and /api/ faster). Importantly, beta/mu frequencies do not generalise to the higher frequency noise band, and conversely the noise band does not generalise to the beta/mu frequencies. Permutation-based significance tests used 500 permutations, Wilcoxin signed rank test (alpha < .05), controlled for multiple comparisons using FDR.

4. Mapping of speech kinematic parameters onto source- and frequency-constrained MEG data.

In the preceding analyses we have used standard MVPA classification of speech versus resting conditions to demonstrate that the neural signals derived from speech motor cortex contain information that is capable of discriminating between speaking and resting conditions; and then employed cross-frequency classification to determine that the mu-beta motor rhythms contain the informational basis of speech-rest discrimination. Importantly, cross-frequency generalisation also shows that the informational structure of the mu/beta rhythm is independent of the high frequency broadband noise that is an inevitable confound for electrophysiological recordings during overt speech.

In subsequent analyses we attempt to derive a more detailed picture of the information structures contained within the neural data, by performing classification between data partitions within the speech condition, rather than between speech and rest conditions. Representational Similarity Analysis (RSA; Kriegeskorte, 2008; Kriegeskorte & Kievit, 2013) is an MVPA technique based on the simple logic that classes of neural data with more similar informational (representational) structures should be more difficult to classify, relative to classes with more distinct representational structures. Previous studies have successfully applied RSA to tracking data of hand movements (Kolasinski et al., 2020), articulator movements during vowel production (Carey et al., 2017), and acoustic measurements during speech production (Zhang et al., 2020).

We follow this logic to test specific hypotheses about potential representational structures in speech motor cortex activity as follows (see Figure 9 for a summary of the computational steps):

(1) Our starting hypotheses concerning candidate representational structures within speech motor cortex activity come from the well-behaved kinematic profiles derived from direct MASK measurements of speech articulator movements (Figure 4): both the strikingly linear relationship between amplitude and velocity, and the orderly curvilinear relationship between duration and stiffness have been proposed to reflect "control parameters" that are relatively tightly specified at some level within the speech motor system;

(2) Within a given kinematic profile, we divide the behavioural data points into partitions that reflect different (Euclidean) distances between the coordinates within each partition. Here we have used 10 partitions to provide a reasonable spread of inter-partition distances.

(3) Data points are averaged within each partition to provide a representation of the central tendency of each partition.

(4) A "behavioural dissimilarity matrix" is generated based on the Euclidean distances between averaged data points in all possible pairs of partitions.

(5) The MEG data (consisting of N trials by 3000-time points x 100 frequencies) is similarly divided into sets of individual trials that correspond to the behavioural data points within a partition.

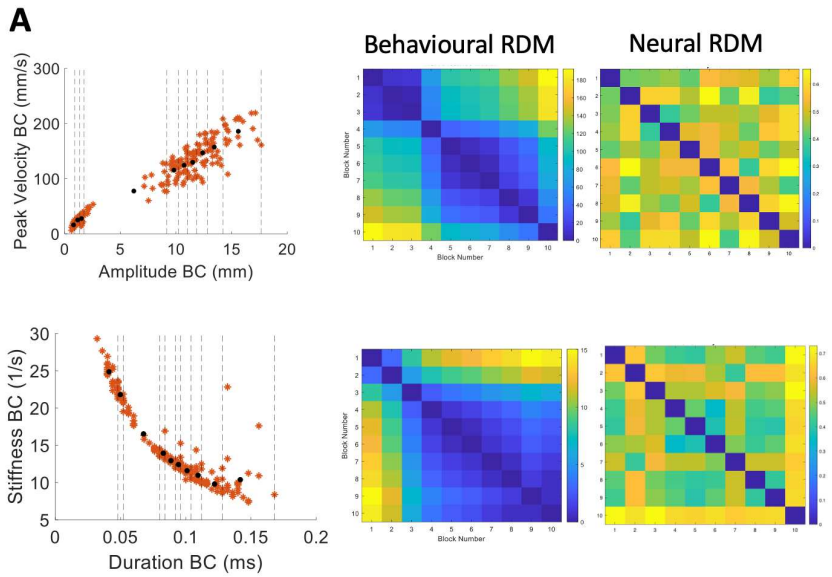
(6) Classification analysis is performed for all possible pairs of MEG trial partitions.

(7) A "neural dissimilarity matrix" (for each time and frequency point) is generated based on the Euclidean distances between the classification accuracy scores.

(8) An "RSA time-frequency plot" is generated containing the correlations between the behavioural dissimilarity matrix and the neural matrices for each time and frequency point.

Model evaluation is restricted in *time* to a 1 second epoch centred on the onset of the BC opening movement, as this event is the reference for the epoching of the MEG data. Model evaluation is further restricted to the *frequencies* of the mu and beta speech motor frequency bands of interest defined by the analyses described in the frequency localisation sections above; and for comparison purposes, a third high frequency region (60-80 Hz) dominated by speech movement noise and which we therefore do not expect to contain useable information concerning the representational structures of speech neuromotor activity.

Group data were statistically evaluated with cluster-based permutation analyses using 500 permutations, alpha < .05, and controlled for multiple comparisons using FDR.



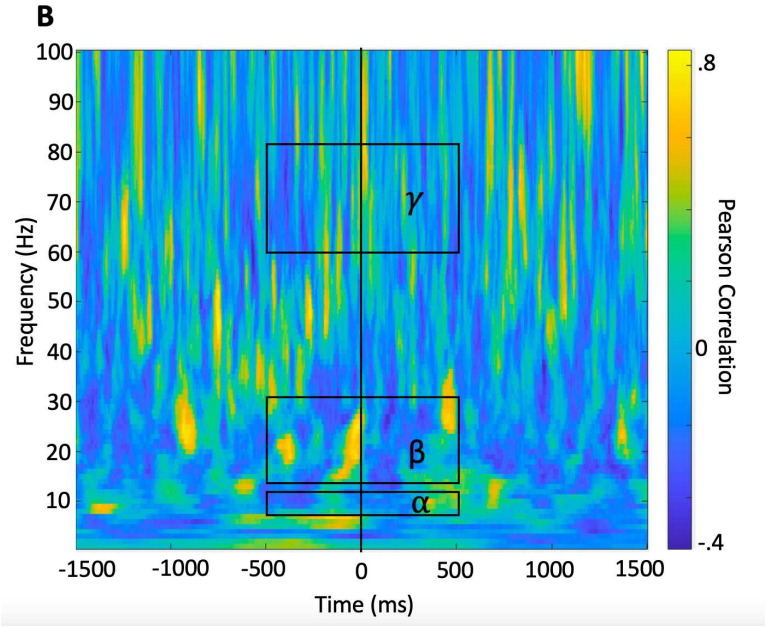


Figure 9. **Procedure for calculation of behavioural and neural correlations**. **A**. Velocity versus amplitude and stiffness versus duration profiles are partitioned into 10 blocks containing equal numbers of trials. Distances between each partition based on the partition mean coordinates are used to generate the behavioural RDMs. For each partition, the same trials of MEG data are input to MVPA classification analysis. The resulting classification (accuracy) metric is used to generate a neural RDM, for each time-frequency point. **B**. A time-frequency correlation matrix resulting from correlating the (in this case, stiffness-duration) behavioural RDM with the neural RDM for each time-frequency point. Group statistics/model evaluation are performed for the mean alpha, beta, and gamma frequency bands within the time range of -500 to +500 ms from onset of the BC opening movement.

4B. Results. The statistically significant results of model evaluation are shown in Figure 10. Of the two behavioural models, four speech conditions, and three frequency bands evaluated, only the /api/ normal rate condition shows showed statistically significant correlations, for beta band and for the stiffness-duration model. Three observations are relevant from these results: (1) The group mean correlations are overall very weak, with peaks restricted to a range of less than - .2 to .2; (2) The temporal structure of the significant positive clusters is appropriate, beginning at a latency of about 100 ms prior to the onset of the BC opening movement. This timing is in good accordance with what one would expect for neural activity associated with behavioural movements. (3) The post-movement cluster of correlations is oppositely (negatively) correlated to the premovement cluster.

Figure 10 also shows comparable (though non-significant after cluster-correction) results for /ipa/ normal rate for the velocity-amplitude model. While we do not wish to over-interpret a non-significant result, the circa -90 ms timing of the peak positive correlation cluster is entirely comparable to that observed for the /api/ normal rate stiffness-duration data described above.

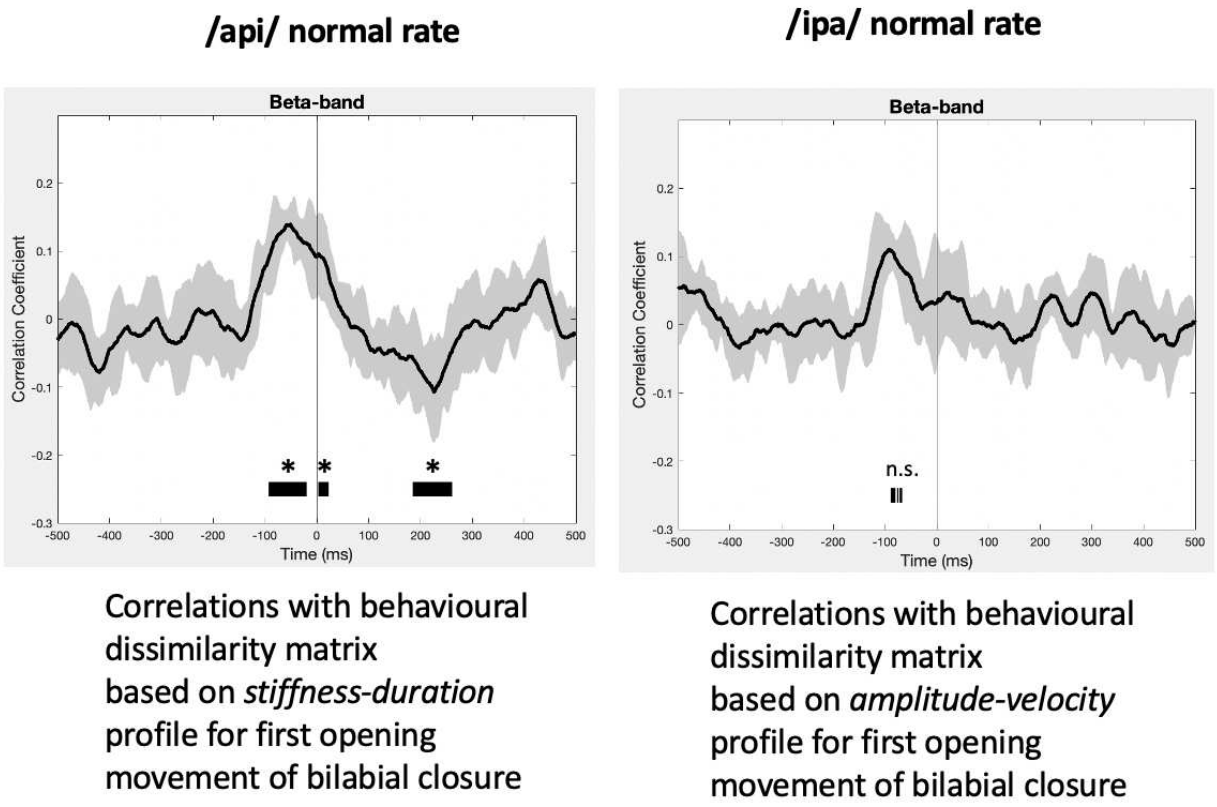


Figure 10. RSA model evaluation. Black lines show group mean correlations between behavioural and neural RDMs; shading shows standard errors. *Left panel*: The beta-band profile for /api/ normal rate shows three significant correlational clusters against the stiffness-duration RDM. The first positive cluster begins about 90 ms prior to onset of the first opening movement of BC. A second positive cluster occurs beginning at time zero, and a third cluster of weak *negative* correlations begins about 180 ms post movement-onset. *Right panel*: Beta correlation against amplitude-velocity time profile shows a similar positive peak circa -90 ms, although the t-value clusters do not survive cluster correction for FDR rate.

Group mean behavioural-neural correlation time series are shown for all speech conditions, frequency bands, and behavioural models in Figure 11. As expected, in the gamma frequency band no significant results were obtained for any speech condition or behavioural model, and there is no discernible, consistent temporal structure in any of the plots. The mu band similarly shows no discernible or consistent temporal structure, and no significant results were obtained.

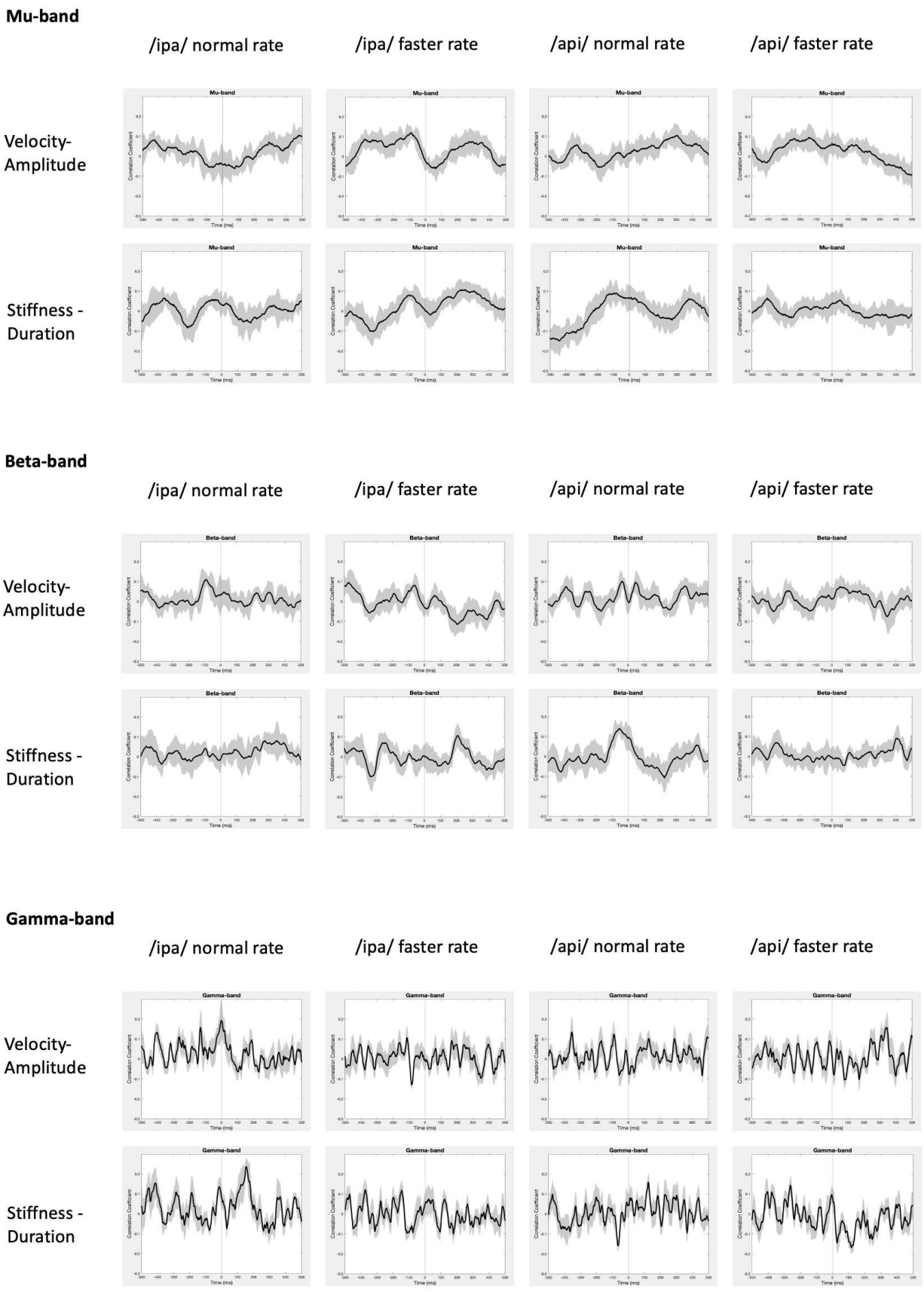


Figure 11. Group mean behavioural-neural correlations for all speech conditions, frequency bands, and behavioural models. Black lines show group mean correlations; shading shows standard errors.

In summary, the present results provide support for very weak but statistically significant encoding of the stiffness-duration relationship only in the beta motor rhythm. This encoding is statistically robust in only one speaking condition and is very weak in terms of magnitude of correlation. However, it is well-structured in time and shows an appropriate and expected temporal relationship (about -90 ms) with respect to movement onset.

Discussion

In the current paper we have presented an analytic framework for MASK-MEG: for measuring and characterising speech movement kinematic parameters and relationships; for spatial localisation of anatomically pertinent regions of interest within the widely distributed brain language network; for frequency localisation of speech-related sensorimotor brain signals; and for mapping of MASK-derived kinematic movement parameters on to temporally co-registered neurophysiological signals from MEG. This analytic framework has the following features:

(1) It provides detailed profiles of speech articulator movements (including non-line-of-sight articulators including the tongue) movements that are demonstrably comparable to those obtained by conventional and established speech movement tracking setups in motor control research laboratories.

(2) It reduces the *spatial dimensionality* of the overall analysis problem, in this case from 160 MEG channels (or alternatively, thousands of source-reconstructed voxels) to a single virtual sensor. Further, the virtual sensor is centred in a region of speech motor cortex encompassing the middle central gyrus/medial frontal gyrus; As we have noted, there is now substantial evidence that this region plays a central role in control and coordination of integrative speech movements.

(3) It reduces the *frequency dimensionality* of the analysis problem to the mu/beta band rhythms, which are well-established rhythms of the sensorimotor cortices of the brain.

(4) It further demonstrates that these sensorimotor rhythms are well-separated and structurally independent (in terms of information content that determines speech-nonspeech classification) from higher broadband frequencies that reflect speech movement noise. This is an important consideration for electrophysiological measurements of overt speech, which will almost inevitably suffer from contamination from movement-related broadband noise generated by the muscles that move the articulators.

(5) Finally, it provides a means of mapping overt speech behaviours onto neural activities, in a manner that allows for direct evaluation of hypotheses concerning the types of information that may be contained within these neural activities; and that further allows for strong inferences concerning the *timing* of relevant neural activations with respect to behavioural outputs.

We note that while our analytic framework dramatically and effectively reduces the dimensionality of the MEG analytic problems, there remains a large decision space concerning the selection of speech behaviours for input to the overall analyses. Here we have fairly arbitrarily focussed on a single articulator metric (the first opening movement of the bilabial closure gesture) and have

tested 2 simple models (derived from the observed kinematic profiles) of the neural informational structures that may underlie this movement. Our results nonetheless provide support for the conclusion that the stiffness-duration relationship of the first opening movement of the bilabial closure may be (very weakly) encoded in the beta-band sensorimotor rhythm, with a timing beginning approximately 90 ms before the onset of the movement. The reason for the occurrence of a second period of significant (but oppositely valanced) association with stiffness-duration at a latency of about 200 ms post movement is presently unclear: one possibility is that reflects a sensory reafference event that provides a check on the motor commands.

Overall, since we have obtained such a very weak association between behaviour and neural activity (and in only one of four speech conditions), it is clear that future work should more systematically probe the sets of possible models of speech movement encoding, including models that describe relationships between articulators that are likely required for integrative speech behaviours (e.g., the Linguistic Gestural Model (LGM) which is a combination of Articulatory Phonology and Task Dynamics (Saltzman & Munhall, 1989; Browman & Goldstein, 1992; Browman & Goldstein, 1997, the speed-accuracy trade off known as Fitts' law (Fitts, 1954; Gafos & van Lieshout, 2021; Kuberski & Gafos, 2021). The current analytic approach provides a framework for just such a systematic assessment of models of speech neuromotor control.

Conclusions

MASK-MEG addresses an important gap in current neuroscientific capabilities for studying expressive language function in the human brain. While we possess robust and well-established methods for measuring and characterising overt movements of the speech articulators, and highly sophisticated equipment and methods for defining the brain activities that control these movements, the two methodologies are not readily or easily combined within a single experimental setup. As a result, speech movement tracking and speech neuroimaging methods have largely evolved within separate laboratories -- even separate disciplines -- and there remains no easy way to co-register and reconcile the different types of information that are derived from them. The advent of neuroimaging-compatible speech tracking technologies such as MASK opens a new window for integrative studies of human speech motor control, combining precision measures of overt speech behaviours with temporally co-registered and spatially localised measures of brain function.

Recent results from invasive ECoG studies of human neurosurgical patients provide compelling reasons to believe that such integrative capabilities will be important for future progress in understanding speech motor control. For example, Chartier et al. (2018) obtained direct cortical recordings of human speech sensorimotor cortex together with (inferred) articulatory kinematics derived from a recurrent neural network based articulatory inversion technique which

learned a mapping from produced speech acoustic to a speaker generic articulator space. This study showed that articulator movements were reflected significantly better in measured neural activity than were either acoustic or phonemic features of speech; that encoding is more related to coordinated movements of multiple articulators than to movements of single articulators; and that the behaviours of encoded movements were governed by damped oscillatory dynamics. These authors concluded that these coordinative and dynamical properties align neatly with the properties of articulatory units of speech (vocal tract gestures) as conceived within the theoretical framework of articulatory phonology and its associated task dynamics model (Browman & Goldstein, 1992; Goldstein & Fowler, 2003; Saltzman, 1986). As such, it seems clear that concurrent speech movement tracking and non-invasive neuroimaging should provide richer datasets with mutually reinforcing inferential power and precision relative to experiments that currently are largely conducted with only one or the other measure of speech motor control.

These new technical capabilities also have clear clinical relevance for advancing our understanding and treatment of developmental and acquired disorders of speech. Speech-sound difficulties are the most common problems encountered by paediatricians and present formidable social, educational and employment obstacles in cases where these problems cannot be readily treated and resolved (Morgan, 2018). Childhood apraxia of speech (CAS) is an intriguing example of a highly debilitating and persistent disorder of speech development whose origins are considered to lie within the brain mechanisms responsible for coordinating and sequencing speech movements, but whose study with conventional neuroimaging approaches has so far proved highly resistant to establishing any clear connection to any particular brain region. In such cases, the capability to directly map speech kinematic and coordination function in speech motor control centres within highly focal and specific brain regions promises to provide more powerful insights into the origins of speech problems in CAS (and conversely, into why speech development proceeds more smoothly in most other children). Similarly, acquired apraxias of speech are a common and debilitating outcome of strokes and other brain injuries. The greater functional specificity of MASK-MEG has a clear bearing on studies aimed at understanding the nature and degree of functional compromise and plastic capabilities in the brain of these patients.

Author Contributions

IA conducted the experiment. IA and BJ analyzed the data. All authors conceived, designed the experiment, discussed the results, wrote, and edited the manuscript.

Data Availability Statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Funding

This work was supported by a Child Development Fund Research Grant (Ref. no. 2532 – 4758) and a Discovery Project Grant (DP170102407) from the Australian Research Council.

References

Alves, N., Jobst C., Hotze F., Ferrari P., Lalancette M., Chau T., Van Lieshout P., Cheyne D. An MEG-compatible electromagnetic-tracking system for monitoring orofacial kinematics. *IEEE Transactions on Biomedical Engineering, 63*, 1709–17. https://doi.org/10.1109/TBME.2015.2500102

Anastasopoulou, I., Van Lieshout, P., Cheyne, D. O., & Johnson, B. W. (2022). Speech kinematics and coordination measured with an MEG-compatible speech tracking system. *Frontiers in Neurology*, *13*, 828237. https://doi.org/10.3389/fneur.2022.828237

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard, M. A., & Greenlee, J. D. W. (2015). Sensory–motor networks involved in speech production and motor control: An fMRI study. *NeuroImage*, *109*, 418–428. https://doi.org/10.1016/j.neuroimage.2015.01.040

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage*, *32*(2), 821–841 https://doi.org/10.1016/j.neuroimage.2006.04.173

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*, *495*(7441), 327–332. https://doi.org/10.1038/nature11911

Bowyer, S. M., Moran, J. E., Weiland, B. J., Mason, K. M., Greenwald, M. L., Smith, B. J., Barkley, G. L., & Tepley, N. (2005). Language laterality determined by MEG mapping with MR-FOCUSS. *Epilepsy & Behavior*, *6*(2), 235–241. https://doi.org/10.1016/j.yebeh.2004.12.002

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. Phonetica, 49(3-4), 155-180.

Catani, M., & Forkel, S. J. (2019). *Diffusion imaging methods in language sciences* (pp. 212-230). Oxford University Press, Oxford, United Kingdom.

Carey, D., Miquel, M. E., Evans, B. G., Adank, P., & McGettigan, C. (2017). Vocal tract images reveal neural representations of sensorimotor transformation during speech imitation. Cerebral Cortex, 27(5), 3064–3079. https://doi.org/10.1093/cercor/bhx056

Carota, F., Schoffelen, J.-M., Oostenveld, R., & Indefrey, P. (2022). The time course of language production as revealed by pattern classification of MEG sensor data. *The Journal of Neuroscience*, *42*(29), 5745–5754. https://doi.org/10.1523/JNEUROSCI.1923-21.2022

Case, J., & Grigos, M. I. (2020). A framework of motoric complexity: An investigation in children with typical and impaired speech development. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3326-3348. https://doi.org/10.1044/2020_JSLHR-20-00020

Chang, E. F., Kurteff, G., Andrews, J. P., Briggs, R. G., Conner, A. K., Battiste, J. D., & Sughrue, M. E. (2020). Pure apraxia of speech after resection based in the posterior middle frontal gyrus. *Neurosurgery*, *87*(3), E383–E389. https://doi.org/10.1093/neuros/nyaa002

Chartier, J., Anumanchipalli, G. K., Johnson, K., & Chang, E. F. (2018). Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. *Neuron*, *98*(5), 1042-1054.e4

Cheyne, D. (2008). Imaging the neural control of voluntary movement using MEG. In *Coordination: Neural, Behavioral and Social Dynamics* (pp. 137-160). Berlin, Heidelberg: Springer Berlin Heidelberg.

Cheyne, D. O. (2013). MEG studies of sensorimotor rhythms: A review. *Experimental Neurology*, *245*, 27–39. https://doi.org/10.1016/j.expneurol.2012.08.030

Cheyne, D., Jobst, C., Tesan G., Crain, S., Johnson, BW. (2014). Movement-related neuromagnetic fields in preschool age children. *Human Brain Mapping*, *35*, 4858–75. https://doi.org/10.1002/hbm.22518

Conant, D. F., Bouchard, K. E., Leonard, M. K., & Chang, E. F. (2018). Human sensorimotor cortex control of directly measured vocal tract movements during vowel production. *The Journal of Neuroscience*, *38*(12), 2955–2966. https://doi.org/10.1523/JNEUROSCI.2382-17.2018

Correia, J. M., Caballero-Gaudes, C., Guediche, S., & Carreiras, M. (2020). Phonatory and articulatory representations of speech production in cortical and subcortical fMRI responses. *Scientific Reports*, *10*(1), 4529. https://doi.org/10.1038/s41598-020-61435-y

De Nil, L., Isabella, S., Jobst, C., Kwon, S., Mollaei, F., & Cheyne, D. (2021). Complexity-dependent modulations of beta oscillations for verbal and nonverbal movements. *Journal of Speech, Language, and Hearing Research*, *64*(6S), 2248-2260.

Doesburg, S. M., Tingling, K., MacDonald, M. J., & Pang, E. W. (2016). Development of network synchronization predicts language abilities. *Journal of Cognitive Neuroscience*, *28*(1), 55–68. https://doi.org/10.1162/jocn_a_00879

Frankford, S. A., Nieto-Castañón, A., Tourville, J. A., & Guenther, F. H. (2021). Reliability of single-subject neural activation patterns in speech production tasks. *Brain and Language*, *212*, 104881. https://doi.org/10.1016/j.bandl.2020.104881

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391. https://doi.org/10.1037/h0055392

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Essen, D. C. V. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. https://doi.org/10.1038/nature18933

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. *Phonetics and phonology in language comprehension and production: Differences and similarities*, 159-207.

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of cognitive neuroscience*, *29*(4), 677-697.

Heim, S., & Specht, K. (2019). Studying language with functional magnetic resonance imaging (fMRI). In G. I. de Zubicaray & N. O. Schiller (Eds.), *The Oxford Handbook of Neurolinguistics* (pp. 71–93). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190672027.013.4

Henriques, R., Van Lieshout P. (2013). A comparison of methods for cecoupling tongue and lower lip from jaw movements in 3D articulography. *Journal of Speech, Language, and Hearing Research, 56*, 1503–16. https://doi.org/10.1044/1092-4388(2013/12-0016)

Hickok, G., Venezia, J., & Teghipco, A. (2023). Beyond Broca: neural architecture and evolution of a dual motor speech coordination system. *Brain*, awac454. https://doi.org/10.1093/brain/awac454

Indefrey, P., & Levelt, W. J. (2000). The neural correlates of language production. *The new Cognitive Neurosciences; 2nd ed.*, 845-865.

Janssen, N., Kessels, R. P. C., Mars, R. B., Llera, A., Beckmann, C. F., & Roelofs, A. (2022). Dissociating the functional roles of arcuate fasciculus subtracts in speech production. *Cerebral Cortex*, bhac224. https://doi.org/10.1093/cercor/bhac224

Johnson, B. W., & He, W. (2019). MEG studies on the connectivity of brain networks in children. *Magnetoencephalography: From signals to dynamic cortical networks*, 733-756.

Johnson, B.W., Jobst, C., Al - Loos, R., He, W., Cheyne, D. (2020). Individual differences in movement related brain activity in early childhood: An MEG study." Developmental Science, 23, e12935. https://doi-org.simsrad.net.ocs.mq.edu.au/10.1111/desc.12935

Jobst, Cecilia, Paul Ferrari, Silvia Isabella, and Douglas Cheyne. BrainWave: A Matlab Toolbox for beamformer source analysis of MEG data (2018). *Frontiers in Neuroscience, 12* (August 22, 2018). https://doi.org/10.3389/fnins.2018.00587

Kadis, D. S., Pang, E. W., Mills, T., Taylor, M. J., McAndrews, M. P., & Smith, M. L. (2011). Characterizing the normal developmental trajectory of expressive language lateralization using magnetoencephalography. *Journal of the International Neuropsychological Society*, *17*(05), 896–904. https://doi.org/10.1017/S1355617711000932

Kado, H., Higuchi, M., Shimogawara, M., Haruta, Y., Adachi, Y., Kawai, J., Ogata, H., & Uehara, G. (1999). Magnetoencephalogram systems developed at KIT. *IEEE Transactions on Applied Superconductivity*, *9*(2), 4057–4062. https://doi.org/10.1109/77.783918

Kelso, J. A., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, *14*(1), 29-59.

Kelso, J. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. MIT press.

Kolasinski, J., Dima, D. C., Mehler, D. M. A., Stephenson, A., Valadan, S., Kusmia, S., & Rossiter, H. E. (2020). Spatially and temporally distinct encoding of muscle and kinematic information in rostral and caudal primary motor cortex. *Cerebral Cortex Communications*, *1*(1), tgaa009. https://doi.org/10.1093/texcom/tgaa009

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. Trends in Cognitive Sciences, 17(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Kuberski, S. R., & Gafos, A. I. (2021). Fitts' law in tongue movements of repetitive speech. *Phonetica*, *78*(1), 3-27.

Leckey, M., & Federmeier, K. D. (2019). Electrophysiological methods in the study of language processing. In G. I. de Zubicaray & N. O. Schiller (Eds.), *The Oxford Handbook of Neurolinguistics* (pp. 41–71). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190672027.013.3

Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, *10*(5), 553–567. https://doi.org/10.1162/089892998562960

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods, 164(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Morgan, A. T., & Webster, R. (2018). Aetiology of childhood apraxia of speech: A clinical practice update for paediatricians. *Journal of paediatrics and child health*, *54*(10), 1090-1095.

Mugler, E. M., Tate, M. C., Livescu, K., Templer, J. W., Goldrick, M. A., & Slutzky, M. W. (2018). Differential representation of articulatory gestures and phonemes in precentral and inferior frontal gyri. *The Journal of Neuroscience*, *38*(46), 9803–9813. https://doi.org/10.1523/JNEUROSCI.1206-18.2018

Munding, D., Dubarry, A.-S., & Alario, F.X. (2016). On the cortical dynamics of word production: A review of the MEG evidence. *Language, Cognition and Neuroscience*, *31*(4), 441–462. https://doi.org/10.1080/23273798.2015.1071857

Murray, E., McCabe, P., & Ballard, K. J. (2015). A randomized controlled trial for children with childhood apraxia of speech comparing rapid syllable transition treatment and the Nuffield Dyspraxia Programme–Third Edition. *Journal of Speech, Language, and Hearing Research*, *58*(3), 669–686. https://doi.org/10.1044/2015_JSLHR-S-13-0179

Ouyang, G., Sommer W., Zhou C., Aristei S., Pinkpank T., Rahman, R.A. (2016). Articulation artifacts during overt language production in event-related brain potentials: Description and correction." *Brain Topography*, *29*, 791–813. https://doi.org/10.1007/s10548-016-0515-1.

Pang, E. W., Wang, F., Malone, M., Kadis, D. S., & Donner, E. J. (2011). Localization of Broca's area using verb generation tasks in the MEG: Validation against fMRI. *Neuroscience Letters*, *490*(3), 215–219. https://doi.org/10.1016/j.neulet.2010.12.055

Peeva, M. G., Guenther, F. H., Tourville, J. A., Nieto-Castanon, A., Anton, J. L., Nazarian, B., & Alario, F. X. (2010). Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage*, *50*(2), 626-638.

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

Ramsey, N. F., Salari, E., Aarnoutse, E. J., Vansteensel, M. J., Bleichner, M. G., & Freudenburg, Z. V. (2018). Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *NeuroImage*, *180*, 301–311. https://doi.org/10.1016/j.neuroimage.2017.10.011

Rong, F., Isenberg, A. L., Sun, E., & Hickok, G. (2018). The neuroanatomy of speech sequencing at the syllable level. *PloS one*, *13*(10), e0196381.

Salmelin, R., Kujala, J., & Liljeström, M. (2019). Magnetoencephalography and the cortical dynamics of language processing. In G. I. de Zubicaray & N. O. Schiller (Eds.), *The Oxford Handbook of Neurolinguistics* (pp. 114–153). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190672027.013.6

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research Series*, *15*, 129-144.

Silva, A. B., Liu, J. R., Zhao, L., Levy, D. F., Scott, T. L., & Chang, E. F. (2022). A Neurosurgical functional dissection of the middle precentral gyrus during speech production. *Journal of Neuroscience*, *42*(45), 8416–8426. https://doi.org/10.1523/JNEUROSCI.1614-22.2022

Tong, J., Binder, J. R., Humphries, C., Mazurchuk, S., Conant, L. L., & Fernandino, L. (2022). A distributed network for multimodal experiential representation of concepts. *The Journal of Neuroscience*, *42*(37), 7121–7130. https://doi.org/10.1523/JNEUROSCI.1243-21.2022

Tourville, J. A., Nieto-Castañón, A., Heyne, M., & Guenther, F. H. (2019). Functional parcellation of the speech production cortex. *Journal of Speech, Language, and Hearing Research*, *62*(8S), 3055–3070. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0442

Treder, M. S. (2020). MVPA-Light: A classification and regression toolbox for multi-dimensional data. *Frontiers in Neuroscience*, *14*, 289. https://doi.org/10.3389/fnins.2020.00289

Uehara, G., Adachi, Y., Kawai, J., Shimogawara, M., Higuchi, M., Haruta, Y., ... & Kado, H. (2003). Multi-channel SQUID systems for biomagnetic measurement. *IEICE transactions on electronics*, *86*(1), 43-54.

Van Lieshout, P. (2021). Electromagnetic articulography. *Manual of Clinical Phonetics*, pp. 356-74. Routledge. https://doi-org.simsrad.net.ocs.mq.edu.au/10.4324/9780429320903

Van Lieshout, P. H. H. M. (2017). Coupling dynamics in speech gestures: Amplitude and rate influences. *Experimental Brain Research*, *235*(8), 2495–2510. https://doi.org/10.1007/s00221-017-4983-7

Van Lieshout, P. H. H. M., Bose, A., Square, P. A., & Steele, C. M. (2007). Speech motor control in fluent and dysfluent speech production of an individual with apraxia of speech and Broca's aphasia. *Clinical Linguistics & Phonetics*, *21*(3), 159–188. https://doi.org/10.1080/02699200600812331

Van Lieshout, P. H., Hulstijn, W., Alfonso, P. J., & Peters, H. F. (1996). Higher and lower order influences on the stability of the dynamic coupling between articulators.

Van Lieshout, P. H. H. M., Rutjens, C. A. W., & Spauwen, P. H. M. (2002). The dynamics of interlip coupling in speakers with a repaired unilateral cleft-Lip history. *Journal of Speech, Language, and Hearing Research*, *45*(1), 5–19. https://doi.org/10.1044/1092-4388(2002/001)

Youssofzadeh, V., & Babajani-Feremi, A. (2019). Mapping critical hubs of receptive and expressive language using MEG: A comparison against fMRI. *NeuroImage*, *201*, 116029. https://doi.org/10.1016/j.neuroimage.2019.116029

Zhang, W., Liu, Y., Wang, X., & Tian, X. (2020). The dynamic and task-dependent representational transformation between the motor and sensory systems during speech production. Cognitive Neuroscience, 11(4), 194–204. https://doi.org/10.1080/17588928.2020.1792868