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 two 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) beta band (13-30 Hz) neuromotor oscillations are present and can be readily separated by filtering from the high frequency MASK coil fields; (2) primary speech sensorimotor cortex can be reliably located in peri-rolandic regions of the left hemisphere from the beta band signals; 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 dramatically 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) (Forkel & Catani, 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), 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 can track articulator movements at rates up to 50 cm/s (Alves et al.). 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 time-frequency and space; and for establishing a mapping between kinematic parameters and brain activities. The scope of the present paper is restricted to a description of analyses carried out at the single subject level. Group-level analyses are described in Anastasopoulou et al. (2023a, b) and forthcoming publications.

Methods

Participants: Two healthy adults participated in this study: S1 (M, 31.6 years; S2 (M, aged 26.5 years). Both participants were monolingual native speakers of Australian English. All procedures were approved by the Macquarie University Human Research Ethics Committee.

Materials: 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). 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).

The occlusal plane was digitised using a plastic protractor with three sensors (Rebernik et al., 2021). MASK coils were placed at mid-sagittal positions on the vermilion border of the upper and lower lip, the tongue body (2 cm from the tongue tip) and the lower incisor sensor which were 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.

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 with 48 kHz sample rate and 8-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 five tasks, four speech production tasks and one 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 1996, Van Lieshout et al., 1997). 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., 1997). 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. 1997; 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.

With regards to the manual nonspeech task, participants were asked to rest their right hand on a response button with two buttons which were associated with the numbers ‘1’ and ‘2’. They had to press the button numbered ‘1’ with their right index finger when a ‘+’ appeared on their screen for approximately every ~2sec for 180sec (see Figure 1B). (Cheyne et al., 2008; Cheyne et al., 2014; De Nil et al., 2021; Johnson & He, 2019).

Fig 1 about here

Analysis:

Analysis of the articulatory signals and trial identification: The articulatory signals were analysed using EGUANA software (Henriques & van Lieshout, 2013; van Lieshout, 2021) to identify coil locations, orientations, and signal magnitudes. 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).

Once the analysis of the kinematics and coordination was complete, 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-written 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.

Fig 2 about here

Analysis of MEG signals:

Speech Motor Cortex Localiser: The *BrainWave* MATLAB toolbox (Jobst et al., 2018; cheynelab.utoronto.ca/brainwave) was used for analysis of MEG data. As an independent localiser, the onset of each individual trial was identified based on the onset of the closing movement of the BC gestural signal. For each production task, continuously recorded MEG signals were sequenced into 15 second epochs (5 sec pre and 10 sec post stimuli locked). All epoched data sets were low pass filtered from 0 – 100 Hz. For the source analysis, each participant's individual MRI coregistered with the MEG source localisation and normalised into standard adult MNI template space using SPM8 (Welcome Institute of Cognitive Neurology).

Multivariate pattern analysis of MEG data: Using source reconstructed MEG time series ("virtual sensor" data) extracted from the localiser voxel locations, we used the multivariate pattern analysis toolbox MVPA-light (Treder, 2020) to map the effects of speech movement variables on neuromagnetic brain signals.

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 both the MASK and EMA 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/. 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).

Fig 3 about here

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.

Fig 4 about here

Gestural phase relationships. The preceding analyses have focussed on the kinematic properties of individual speech movements, we will now shift our attention to the coordinative properties of articulator movements within and between speech gestures. Coordination in speech motor control is typically defined in terms of relative timing, as indicated by the relative phase between two articulators or two gestures [(26, 28, 39); for alternative conceptualizations of coordination see for e.g., Pearson and Pouw (40) and Vilela Barbosa et al. (41)].

van Lieshout et al. (26) distinguished between two types of coordination: "intra-gestural coordination," which is based on relative phase signals between upper and lower lip movements, and "inter-gestural coordination," which involves relative phase calculations between the TB and BC gestures. We computed TB vs. BC phase coordination for both the /ipa/ and /api/ tasks.

We would expect the relative timing of the UL and LL to remain consistent across different speaking rates to ensure intelligibility. As expected, the intra-gestural coordination of UL and LL in the three speech tasks (Figure 5, top panel) showed highly similar relative phase relationships across normal and faster speech rates in both participants. These findings suggest that the relative timing of lip movements is highly stable for a given gesture in a specific phonetic context within a particular speaker (see also Anastasopoulou et al., 2022).

As mentioned earlier in the raw tracking results, the gestural movements of /ipa/ and /api/ can be seen as mirror images, with the relative motions of the tongue body and BC gestures reversed. This reversal is evident in the inter-gestural timing plots of Figure 5 (bottom panel) for both participants. Notably, both participants exhibit highly similar relative phase values for individual productions, which are consistent across speaking rates. Therefore, the patterns of relative phase across the four gestures plotted are highly comparable. These inter-gestural patterns are also comparable to the inter-gestural patterns described for the same gestures by van Lieshout et al. [(26), Figure 9]. Overall, our results and the findings of van Lieshout et al. (26) suggest that inter-gestural coordination is a highly stable motor control parameter both within and between participants for these types of tasks.

Fig 5 about here

Speech-related brain signals. We now turn our attention to extraction of the neural signals contained in the MEG-MASK measurements. As noted previously, the bulk of known EEG/MEG brain responses reside at relatively low frequencies (sub-100 Hz) of the overall frequency spectrum and can thus be readily separated from the (> 250 Hz) MASK tracking signals by low pass filtering.

This component of our analysis pipeline proceeds in two successive phases: *Localisation*, followed by *classification*. In the localisation phase we begin with a formidably high dimensional analysis space (several thousand spatial voxels) and aim for a highly reduced subset of brain regions to provide the focus for subsequent analyses. In the present context our spatial region of interest (speech motor cortex) is localised by applying synthetic aperture (SAM) beamforming to 15 seconds duration MEG segments encompassing 10 seconds prior to and 5 seconds after speech trial onsets. 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). 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 nicely 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 (Cheyne, 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).

Finally, we can assert that the results of the localiser procedure are highly robust and reliable across participants. In separate reports (Anastasopoulou et al., 2023a; Anastasopoulou et al., 2023b) we have reported group localiser results for a sample of 10 healthy adults.

Fig 6 about here

Mapping speech movement parameters to brain activity. At this stage of the analysis pipeline, we have separate sets of well-defined movement and brain parameters, and the remaining task is to reconcile these into an integrated representation of the relationship between brain activities and overt speech movements. As our goal here is to probe for and identify salient neural activity patterns, the underlying logic of our analytic approach is to treat the movement parameters as independent variables and the temporally associated MEG measurements as the outcome variables.

Figure 7 shows the patterning of MEG neural activity in participant S1133 resulting from application of a simple binary classifier, using the production variable (/ipa/ versus /api/) as the classifier variable. At the normal speaking rate, classification accuracy is clearly maximally clustered within the beta band, supporting the relevance of this neuromotor oscillation to the

speech movement variables (as noted previously, the /ipa/ and /api/ productions require mirror symmetric relative phases of the TB and BC gestures). Interestingly, at the faster speaking rate, the classification accuracy cluster maxima shifts to a distinctly higher frequency within the beta band, and with an increase in the prominence of a frequency cluster in the mu-band, another prominent and well-established rhythm of the central sensorimotor cortices (Cheyne, 2014).

Fig 7 about here

Conveniently, the multivariate pattern analysis approach can readily accommodate high dimensional classifications as well. To extend analyses from the simple binary classifications described above, techniques such as Representational Similarity Analysis (RSA) are now commonly employed in neuroimaging analyses. In RSA pattern classification is based on dissimilarities in brain activity obtained between each pair of experimental conditions in a multi-level experimental design (Kriegeskorte, 2008; Kriegeskorte & Kievit, 2013). 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).

Figure 8 illustrates this approach with an RSA dissimilarity matrix derived from the amplitude/velocity and duration/stiffness plots for bilabial closure gesture for participant S1119 in Figure 4. Expressed as velocity distances as a function of amplitude, the Figure 8 representation (left) retains the roughly linear increase in magnitude as a function of amplitude. This linearity is visually obvious when one considers amplitude excursions between about 5-10 mm relative to all other amplitudes since the bulk (about 55%) of data points fall within this range. As may be anticipated from the curvilinear stiffness/duration X/Y plot of Figure 4, the stiffness dissimilarity representation is strikingly different, with a roughly linear increase in stiffness distances for durations less than about 50ms; and differences approaching zero at longer durations, reflecting the plateau portion of stiffness duration in Figure 4.

From this point the implementation of RSA multivariate pattern analysis to the time-registered neuromagnetic data is straightforward: in simple terms, the ensuring classification of neural data will attempt to find the best match of neural activity patterns to the behavioural patterns represented in the dissimilarity matrices. According to the specific aims of a given experiment, multivariate pattern analysis can be deployed to identify matching patterns of activity in space (“spatial spotlight” analysis); or alternatively, to extract time series from a

predefined region of interest (specified by the localiser analysis in the preceding) and identify matching patterns of activity in time ("temporal spotlight" analysis).

Since multivariate analyses typically require higher levels of physiological signal to noise ratios relative to simple binary contrasts, and since the scope of the present technical paper is restricted to single subject (level 1) analysis, we defer a description of the application of the RSA classification analyses to neuromagnetic data to a forthcoming report on a statistically tractable group of participants.

Fig 8 about here

Discussion

In the current paper we have outlined and described the technical capabilities of MASK-MEG at an individual participant level, for measuring and characterising speech movement kinematics and coordinative dynamics; for spatial localisation of anatomically pertinent regions of interest within the widely distributed brain language network; for frequency localisation of physiologically relevant sensorimotor brain signals; and for mapping of the high dimensional space of derived kinematic parameters on to temporally co-registered physiological signals.

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. These authors reported that individual electrodes preferentially encoded *articulatory kinematic trajectories*, reflecting coordinated articulator movements toward specific vocal tract shapes. Such observations obviously converge neatly with and provide strong support for current theoretical conceptualisations of articulatory phonology and task dynamics (Saltzman, 1986; Browman & Goldstein, 1992; Goldstein & Fowler, 2003). 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. 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.

Figure Captions

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.

Figure 2: Temporal alignment of MEG and MASK signals. MASK (PsiFi) 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).

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.

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.

Figure 5. Intra- and inter-gestural coordination in /ipa/ and /api/ productions, indexed by relative phase. UL = upper lip, LL = lower lip, BC = bilabial closure, TB = tongue body.

Figure 6. 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.

Figure 7. Binary classification (/ipa/ vs. /api/) for time-frequency data for S1133. Classification accuracy is plotted as a function of time and frequency.

Figure 8. Representational Similarity Analysis (RSA) matrices of bilabial closure kinematic parameters for S1119. Left: Velocity distance matrix. Right: Stiffness distance matrix. These distance matrix representations can be readily interpreted with respect to xy plots of Fig. 4: Velocity distances increase as a linear function of amplitude distances. In contrast, stiffness distances are largest at smaller durations and plateau thereafter.

References

Anastasopoulou, I., Cheyne, D. O., van Lieshout, P., & Johnson, B. W. (2023a). Non-invasive mapping of speech motor control regions during speech production. Paper to be presented at biannual meeting of *International Society for Advancement of Clinical Magnetoencephalography*, 24-27 May 2023, Osaka, Japan.

Anastasopoulou, I., Cheyne, D. O., van Lieshout, P., & Johnson, B. W. (2023b). A novel protocol for non-invasive mapping of speech motor control regions during speech production. Paper to be presented at annual meeting of *Organisation for Human Brain Mapping* (OHBM2023), Montreal, Canada, 22-26 July 2023.

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

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

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

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.

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

Hickok, G., Venezia, J., & Teghipco, A. (2022). 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

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

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

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

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

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

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

Rebernik, T., Jacobi, J., Jonkers, R., Noiray, A., & Wieling, M. (2021). A review of data collection practices using electromagnetic articulography. *Laboratory Phonology*, *12*(1). https://doi.org/10.5334/labphon.237

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

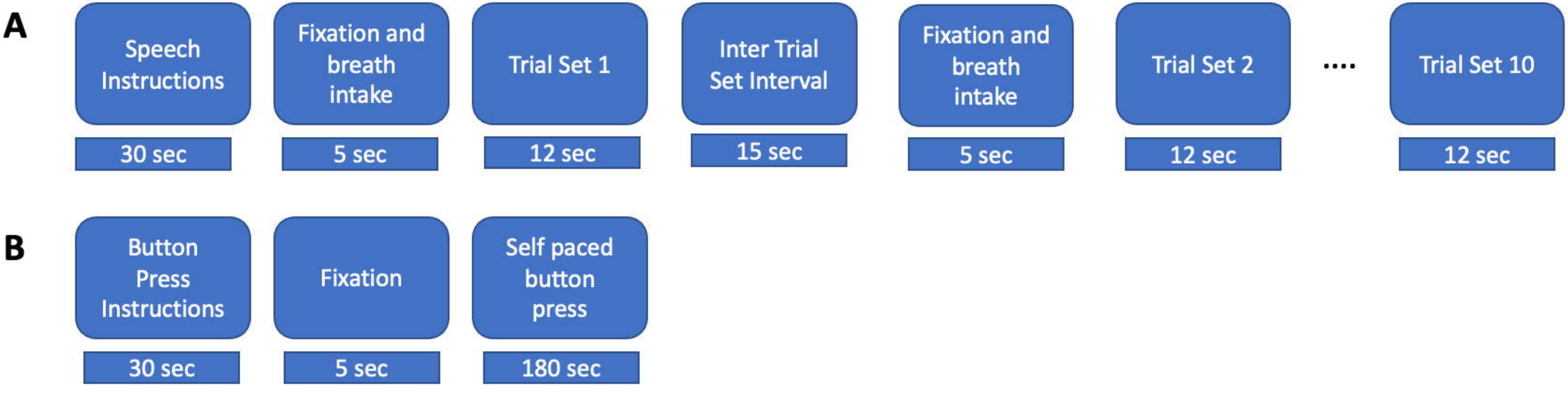
van Lieshout, P. H. H. M. (2017). Coupling dynamics in speech gestures: Amplitude and rate influences. *Experimental Brain Research*, *235*(8), 2495–2510. https://doi.org/10.1007/s00221-017-4983-7

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

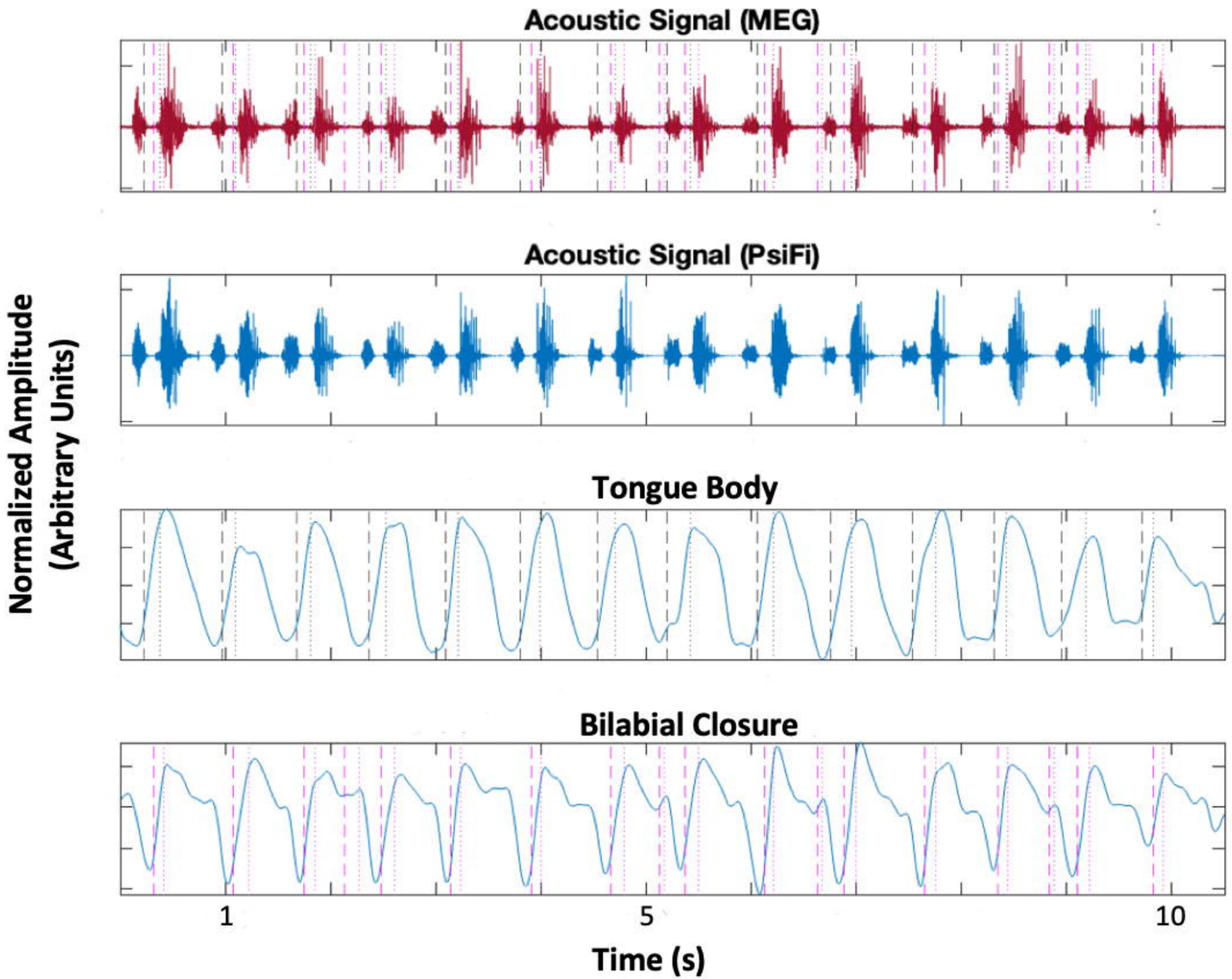
van Lieshout, P. H. H. M., Rutjens, C. A. W., & Spauwen, P. H. M. (2002). The dynamics of interlip coupling in speakers with a repaired unilateral cleft-Lip history. *Journal of Speech, Language, and Hearing Research*, *45*(1), 5–19. https://doi.org/10.1044/1092-4388(2002/001)

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

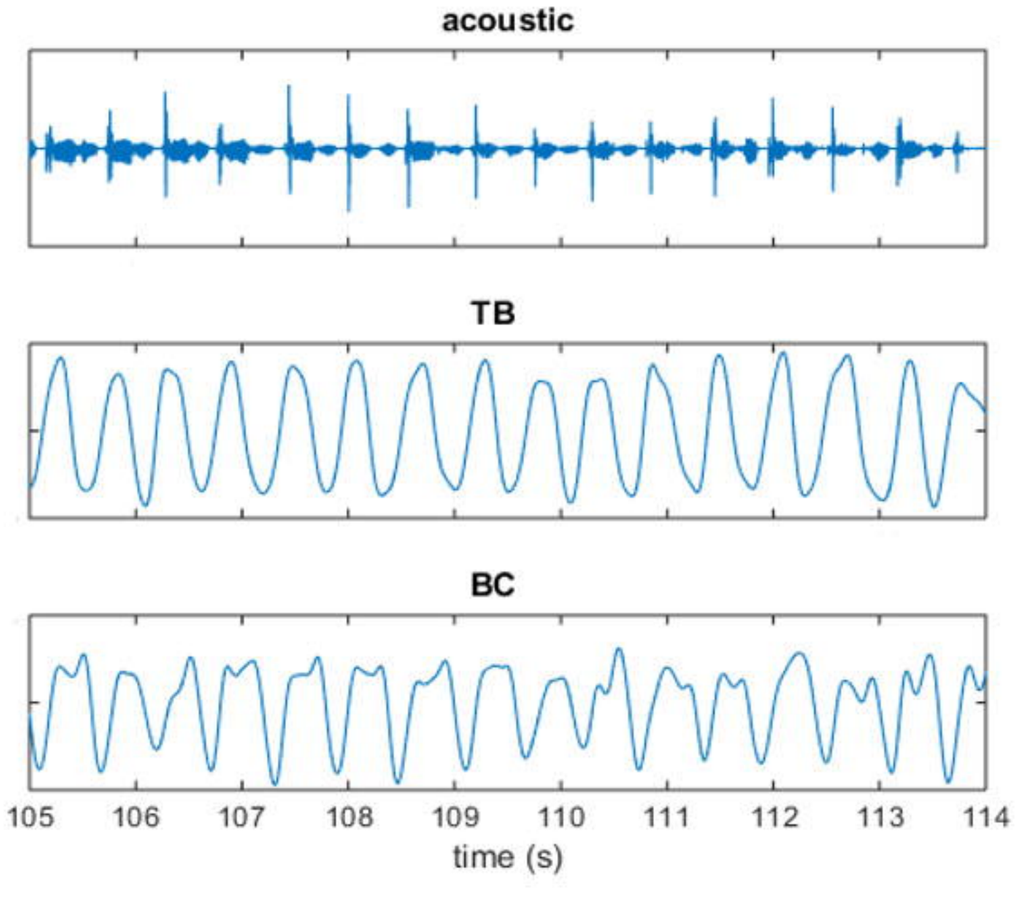

A
Speech Instructions
30 sec
Fixation and breath intake
5 sec
Trial Set 1
12 sec
Inter Trial Set Interval
15 sec
Fixation and breath intake
5 sec
Trial Set 2
12 sec
....
Trial Set 10
12 sec
B
Button Press Instructions
30 sec
Fixation
5 sec
Self paced button press
180 sec

S1119 /ipa/ normal

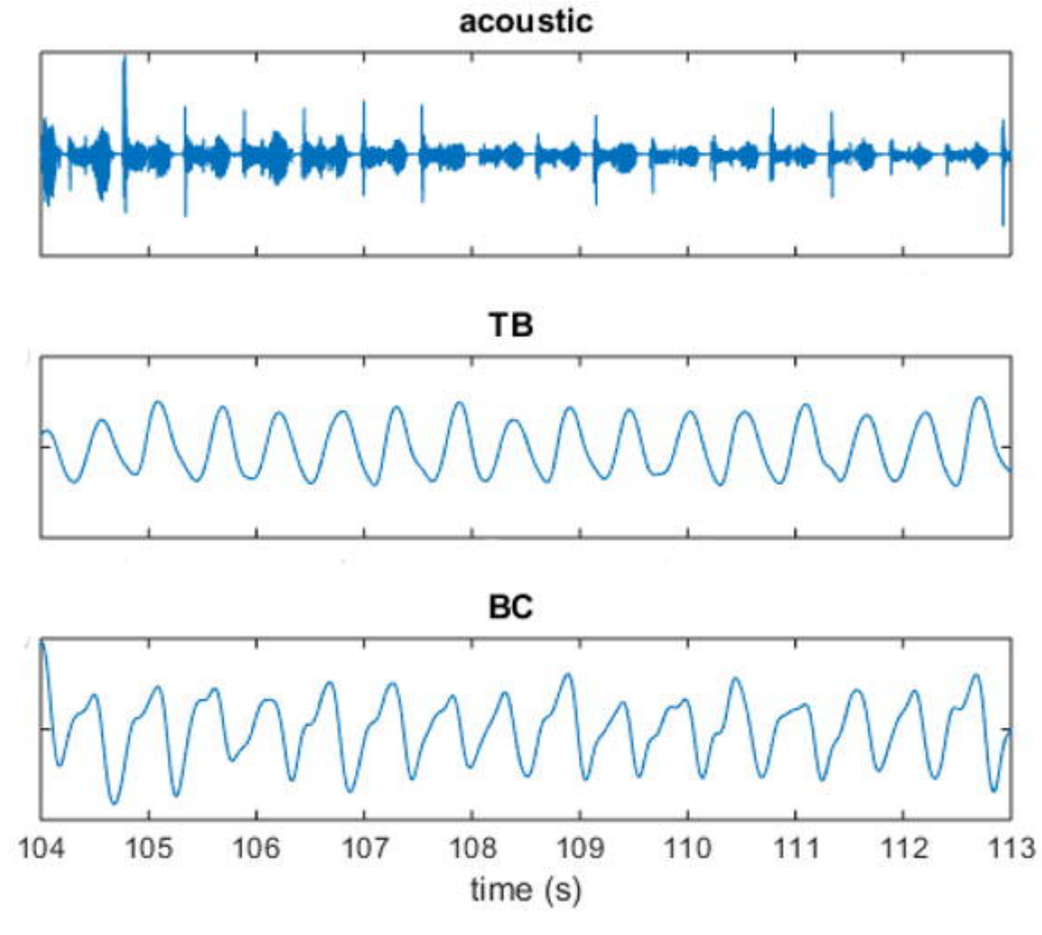


'--' BC onset of closing movement
':' BC offset of closing movement
'--' TB onset of closing movement
':' TB offset of closing movement

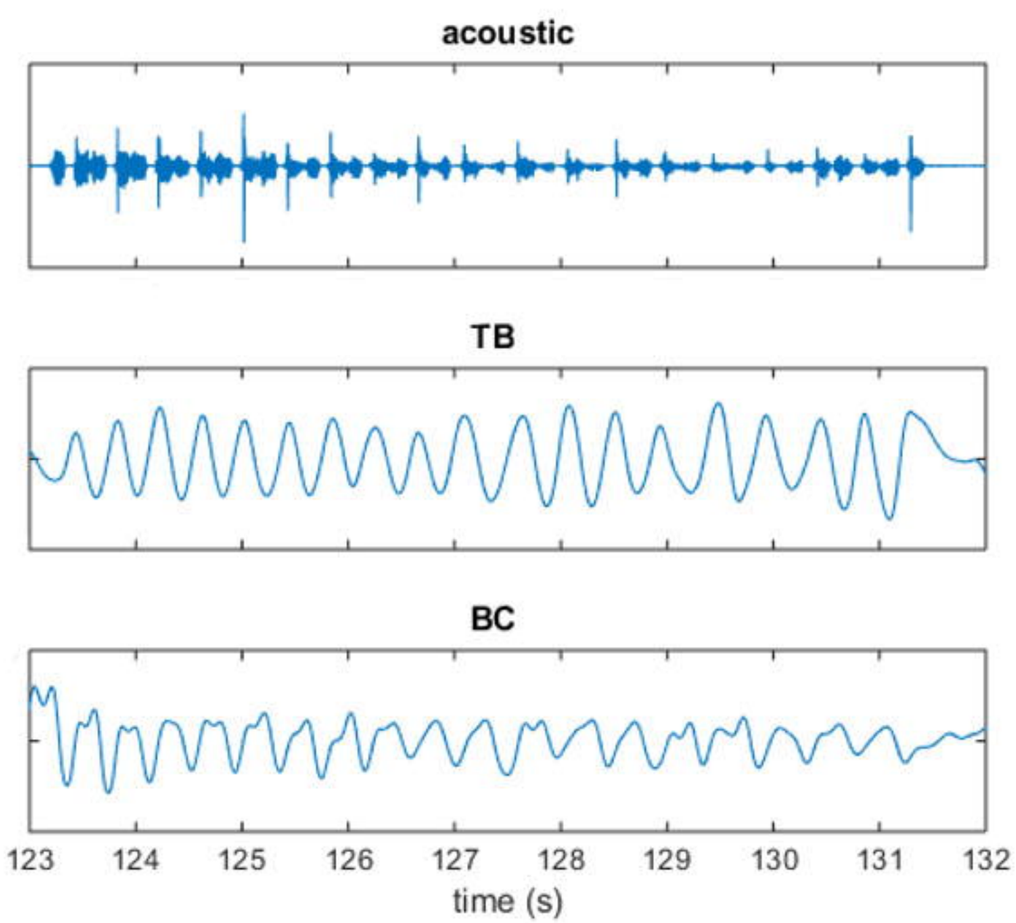
S1133 /ipa/ normal rate



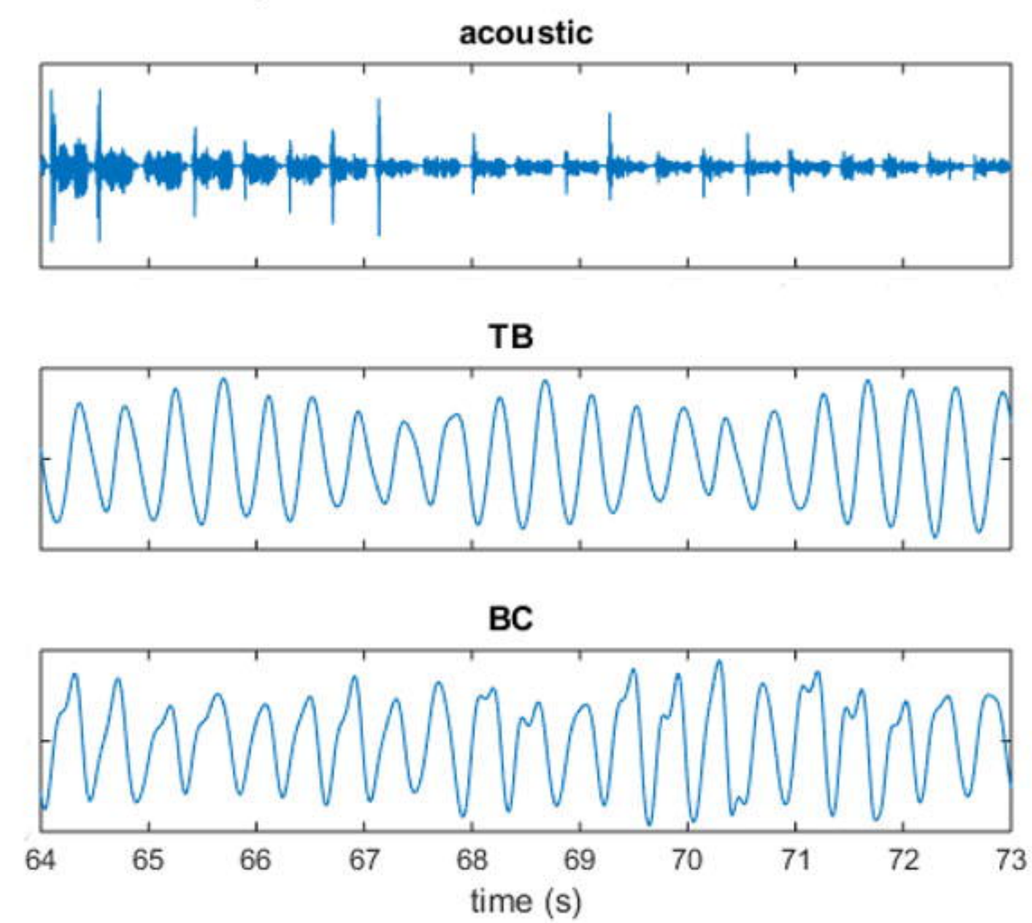
S1133 /api/ normal rate

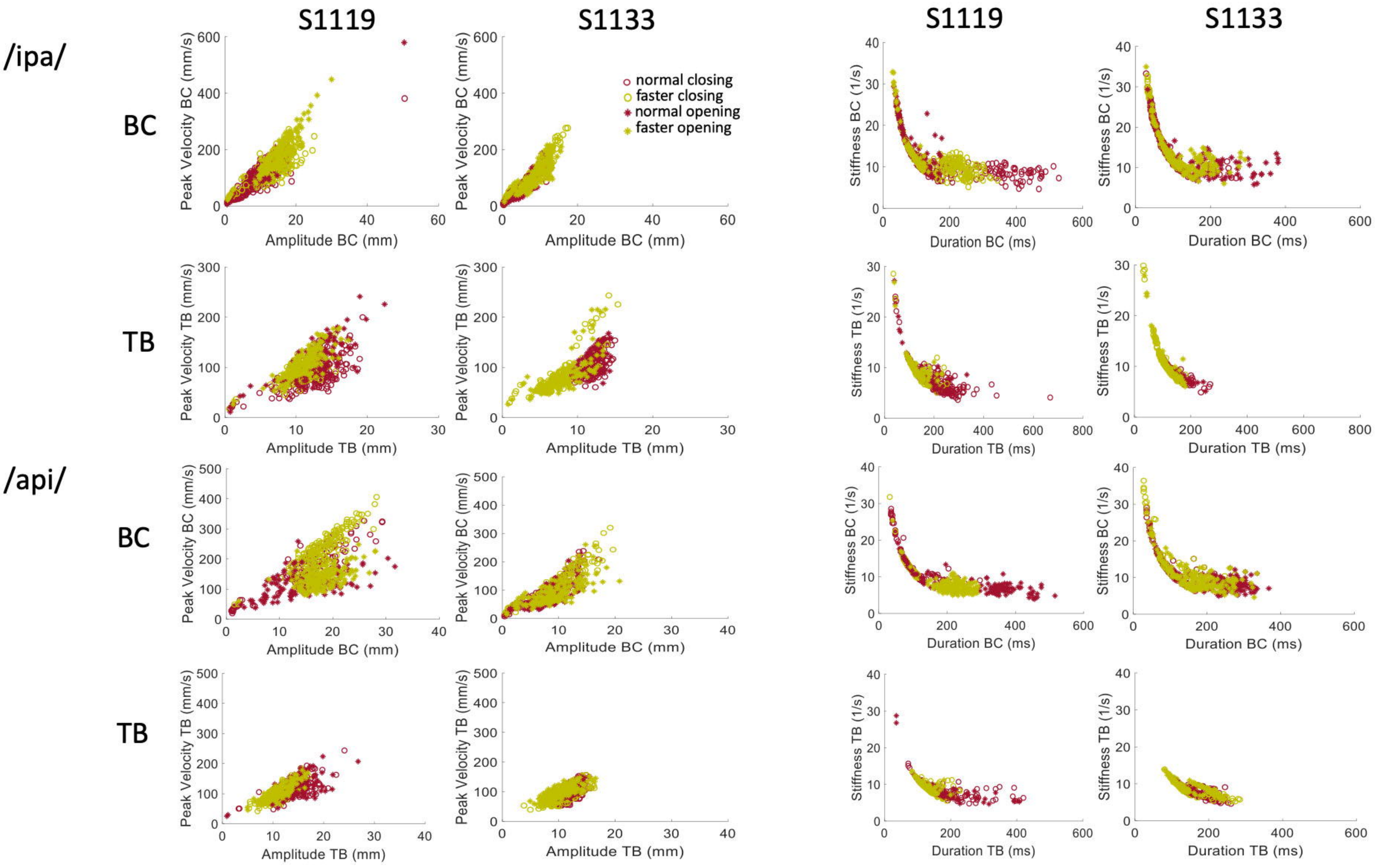


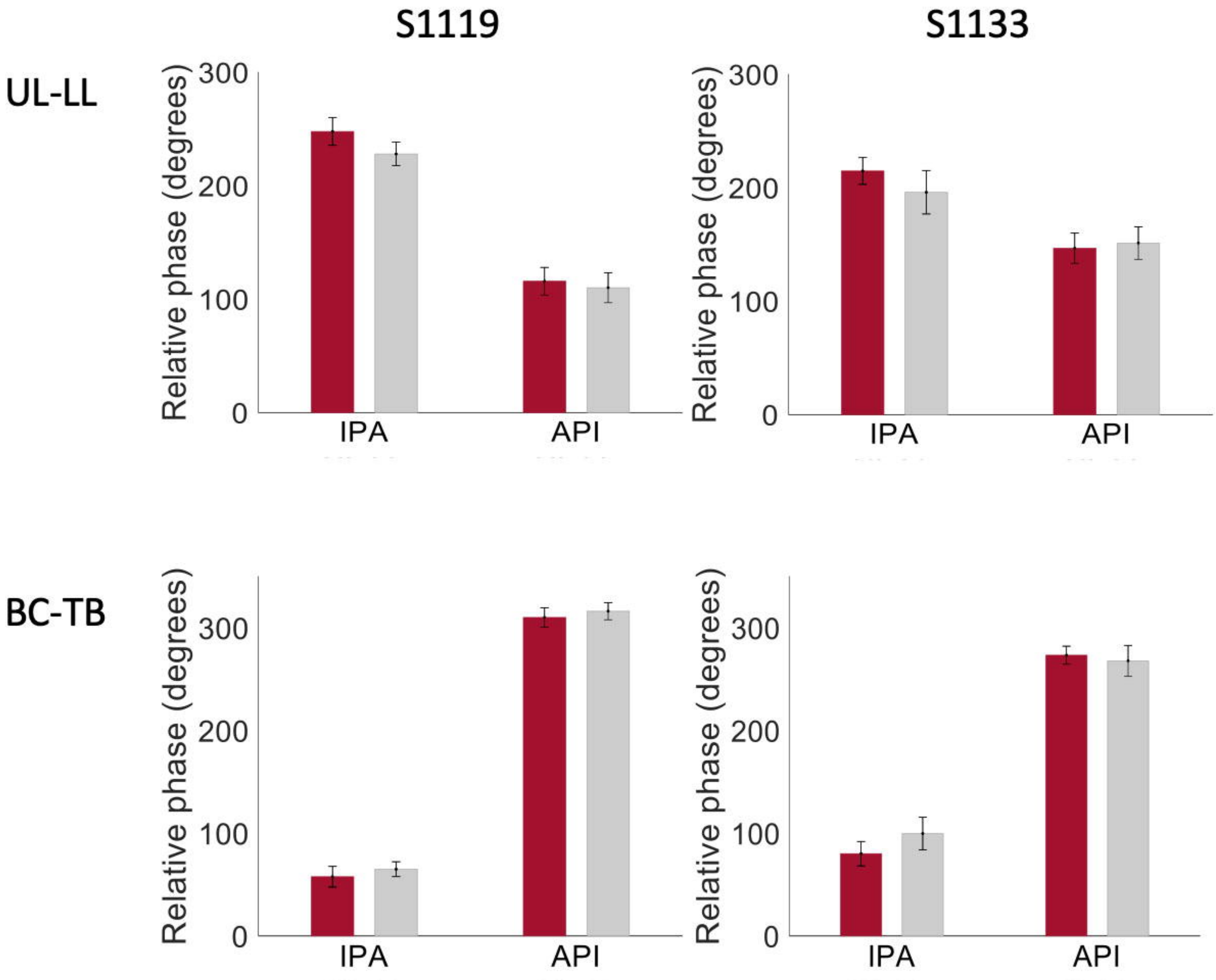
S1133 /ipa/ faster rate

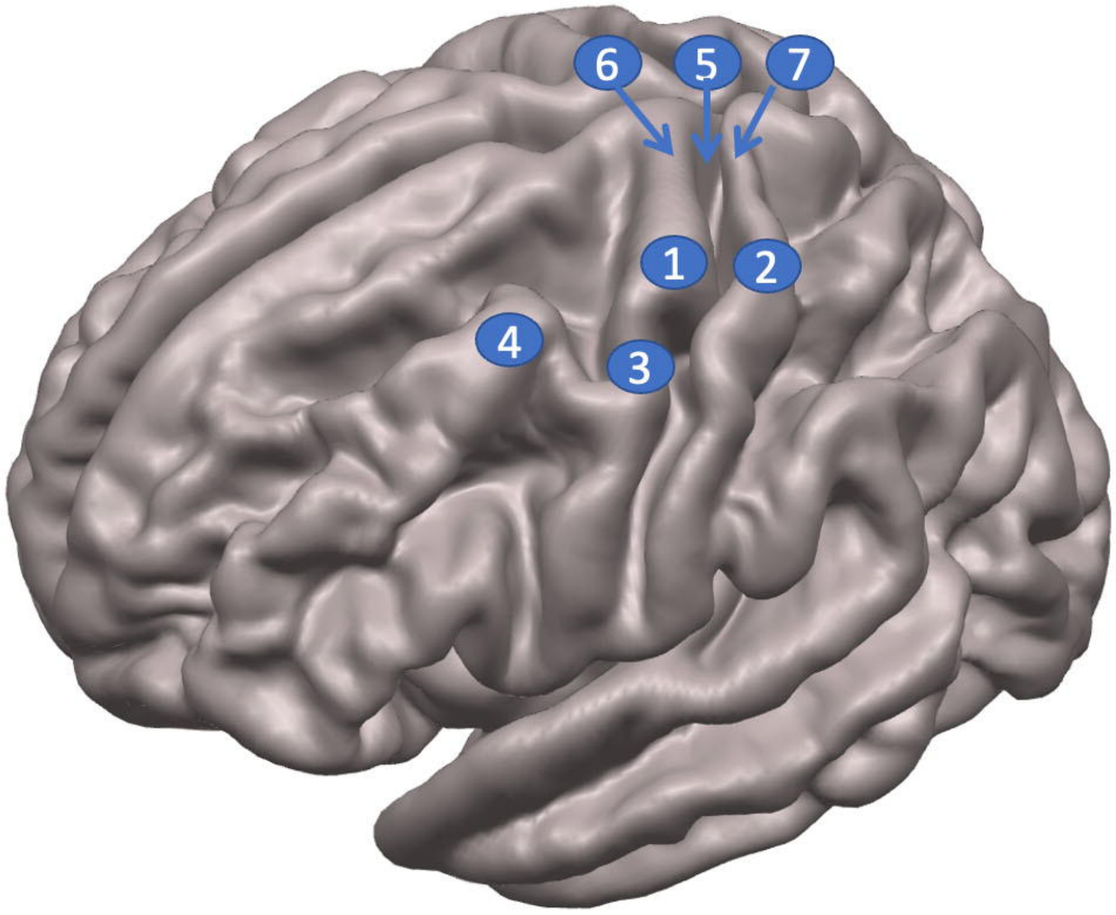


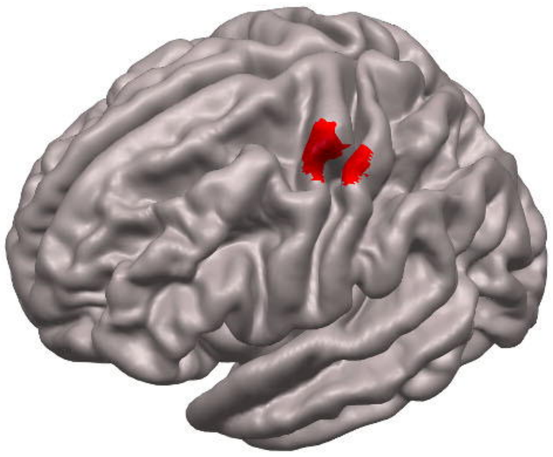
S1133 /api/ faster rate



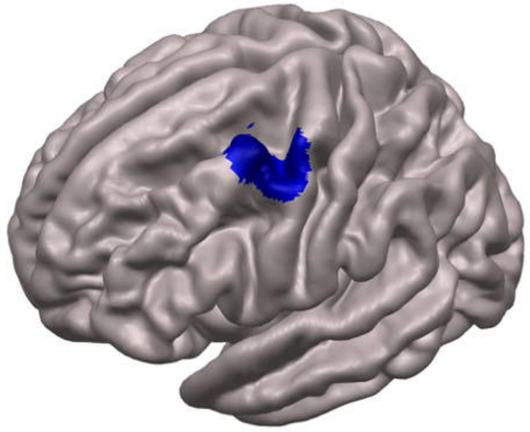

/ipa/
/api/
BC
TB
S1119
S1133
normal closing
faster closing
normal opening
faster opening
Peak Velocity BC (mm/s)
Peak Velocity TB (mm/s)
Amplitude BC (mm)
Amplitude TB (mm)
Stiffness BC (1/s)
Stiffness TB (1/s)
Duration BC (ms)
Duration TB (ms)


S1119
S1133
UL-LL
BC-TB
Relative phase (degrees)
300
200
100
0
IPA
API

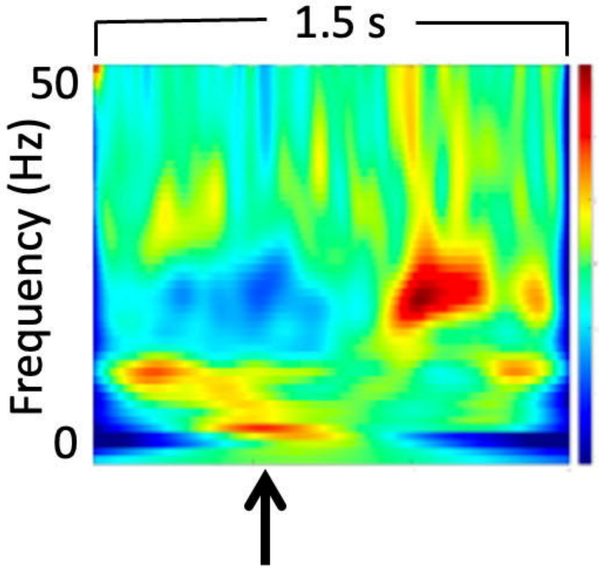


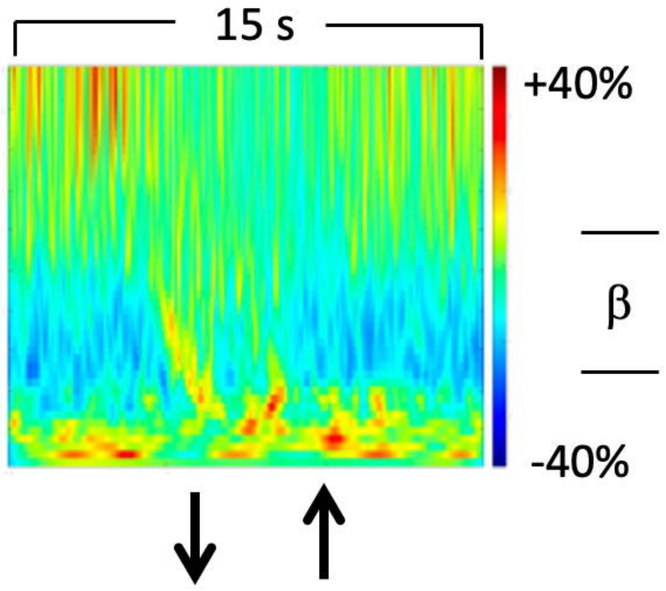


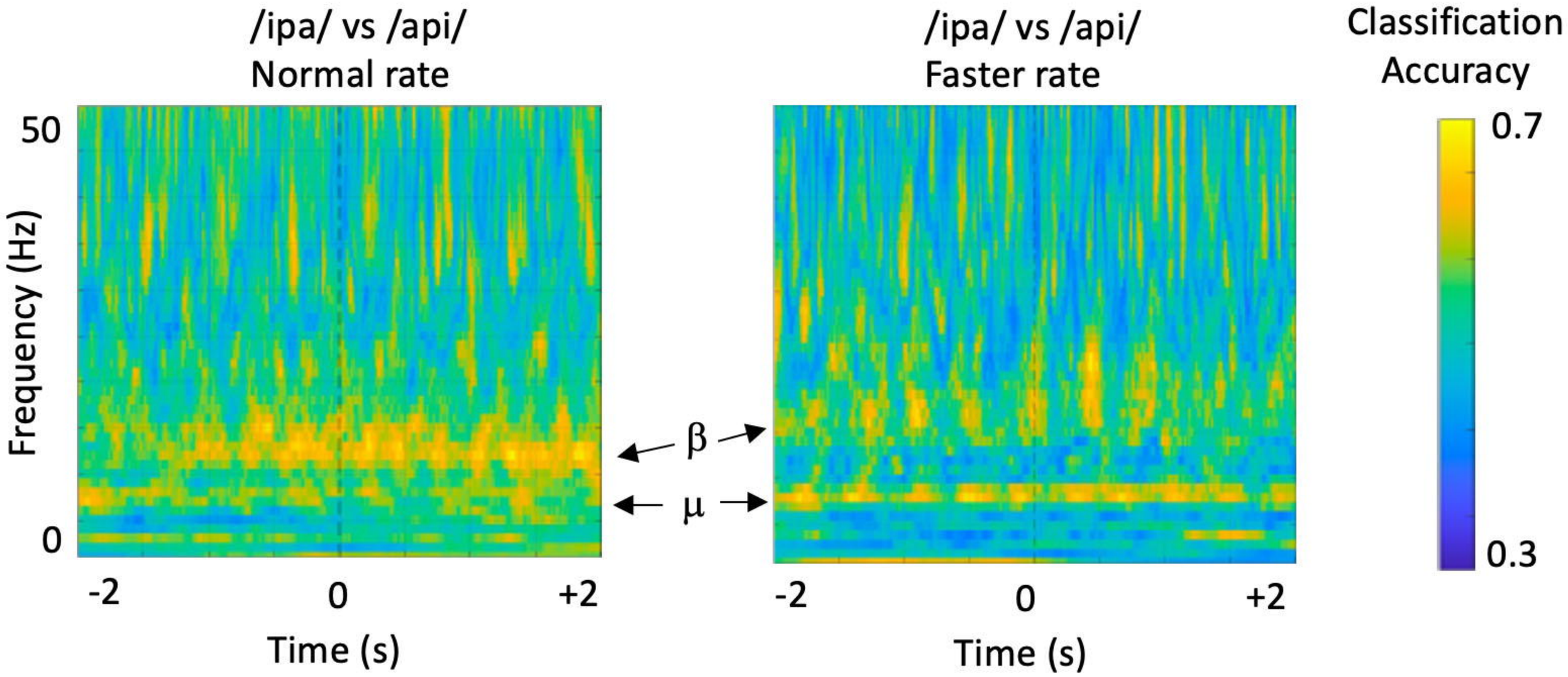
Button Press

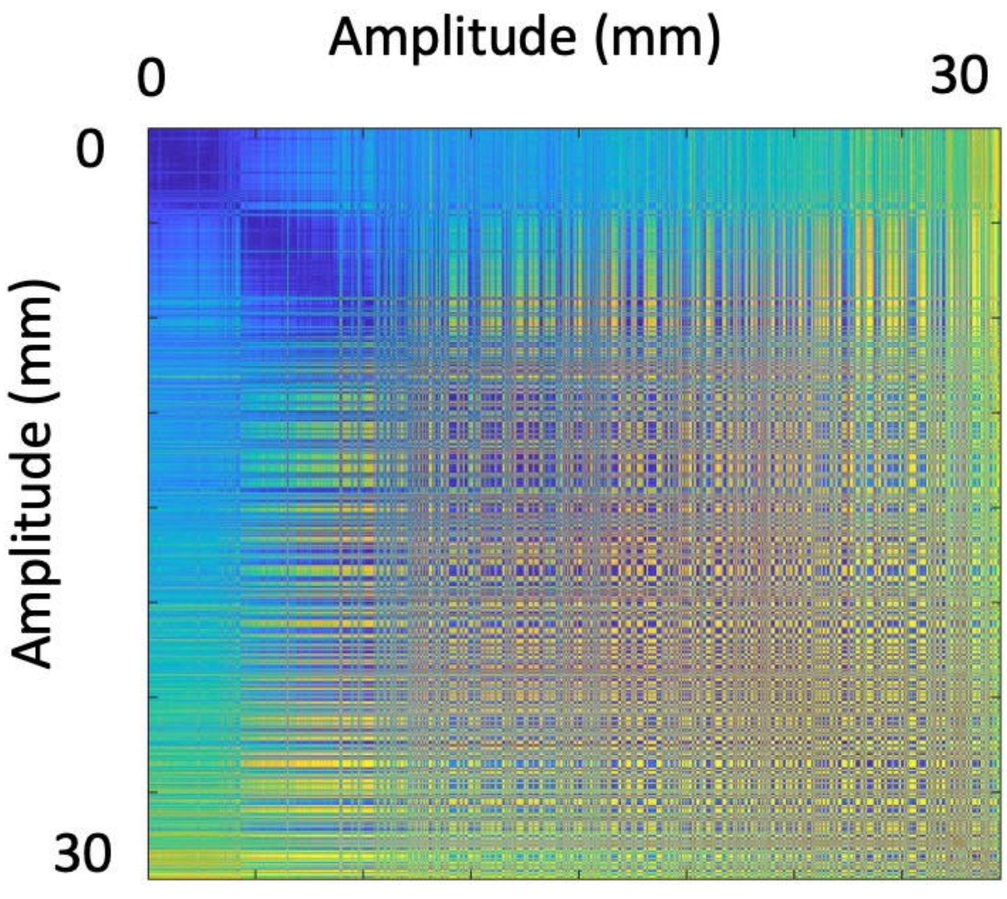


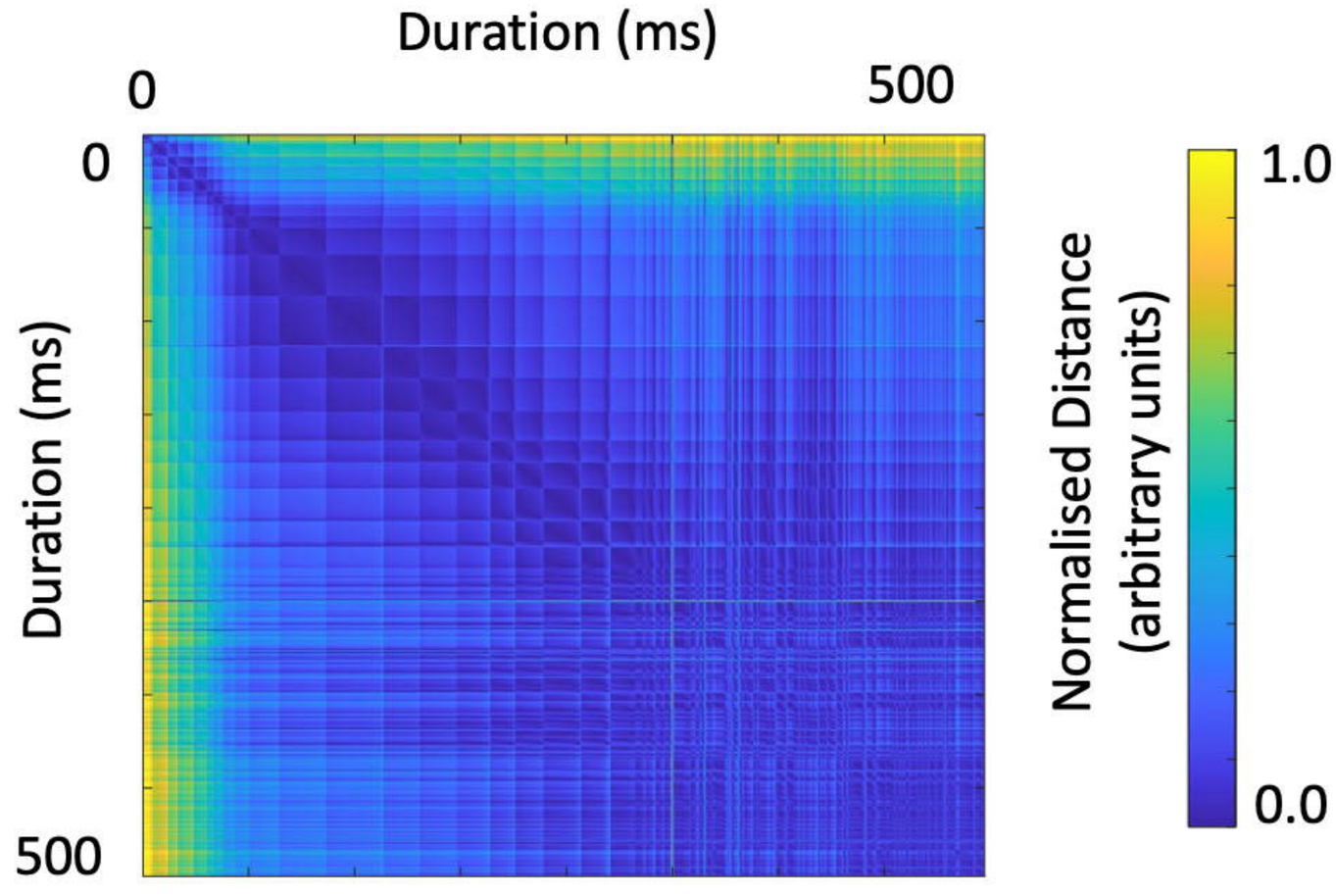
Speech






/ipa/ vs /api/
Normal rate
/ipa/ vs /api/
Faster rate
Classification
Accuracy
0.7
0.3
50
0
Frequency (Hz)
β
μ
-2
0
+2
Time (s)
-2
0
+2
Time (s)


Amplitude (mm)
0
30
0
30
Amplitude (mm)


Duration (ms)
0
500
0
500
Duration (ms)
Normalised Distance
(arbitrary units)
1.0
0.0