Transcranial focused ultrasound enhances behavioral and network mechanisms underlying response inhibition in humans

Justin M. Fine,[1] Maria E. Fini, Archana S. Mysore,

William J. Tyler, Marco Santello

School of Biological and Health Systems Engineering,
Arizona State University, Tempe, AZ

[1]Current address: Psychological and Brain Sciences,
Indiana University, Bloomington, IN 47405

Correspondence: Justin M. Fine
Psychological and Brain Sciences
Indiana University
1101 E 10th Street
Bloomington, IN 47405
Email: justfine@iu.edu

Keywords: Transcranial focused ultrasound; response inhibition; dynamic causal modeling; inferior frontal gyrus; cognitive control;

Abstract

Changing environmental conditions necessitates a control system that can suppress unwanted actions. The cortical network typically recruited during inhibition is centered around right inferior frontal gyrus (rIFG) and pre-supplementary motor areas (pre-SMA). Although it has been proposed that inhibition is directly triggered and controlled by rIFG activity projected to pre-SMA, an alternative framework posits that rIFG as mediating attentional reorienting or context monitoring. Therefore, the causal role of rIFG in response inhibition remains unknown. Here, using a stop signal task in humans concurrent with EEG and high-resolution transcranial focused ultrasound stimulation (TFUS), we found stimulation of the pars opercularis sector in rIFG increased inhibitory performance through effects isolated to stopping speed. Behavioral outcomes across stimulation conditions were only tracked by timing of P300 event-potentials, providing the first causal link between rIFG, inhibitory control and P300 timing. For rIFG to drive inhibitory control, we hypothesized it should exert a top-down, neural inhibitory connectivity with pre-SMA. Using dynamic causal modeling (DCM) of EEG evoked responses, we found rIFG top-down activity drives local inhibition within pre-SMA through a top-down gain modulation of inhibitory populations. This mechanism alone predicted the neural responses and tracked inhibitory outcomes and speed during both TFUS conditions. Connectivity along the fronto-parietal pathway corresponded to inhibitory success but not stopping speed across TFUS contexts. Together, our observations indicate inhibiting movements involves both a direct neural inhibitory function through rIFG and secondary attentional pathway modulation.

Introduction

Response inhibition is necessary to stop unwanted actions (Logan & Cowan, 1984). Prominent hypotheses indicate right inferior frontal gyrus (rIFG) drives inhibition along with pre-supplementary motor area (pre-SMA) (Aron, 2011). Neuroimaging indicating differential activation in the pars opercularis sector of rIFG during inhibition (Aron et al., 2006; Boehler et al., 2010) and lesion studies (Aron et al., 2003) support a central role for rIFG. Pre-SMA activation during action-preparation (Chikazoe et al., 2009; Rushworth et al., 2004) further indicates pre-SMA generally represents actions that rIFG inhibits. However, alternative positions suggest that rIFG mediates context-monitoring and attention instead of inhibition (Hampshire et al., 2010; Munakata et al., 2011). Supporting evidence includes fMRI indicating rIFG activates similarly across tasks involving inhibition and context-monitoring, while pre-SMA activation scales with inhibition (Erika-Florence et al., 2014; Sharp et al., 2010).

Given rIFG's known connectivity to pre-SMA and attentional processing pathways (i.e., fronto-parietal), rIFG activity during inhibition could reflect both functions through top-down coupling (Wiecki & Frank, 2013). The key to determine whether rIFG drives inhibitory control is linking temporally specific neural activity to neural inhibitory mechanisms that are necessary for direct response suppression. Insights from neural network modeling (Lo et al., 2009) suggests response inhibition can emerge through top-down gain modulation of populations in action-coding areas such as pre-SMA (Chikazoe et al., 2009). This is consistent with inhibitory effects of feedback (i.e., top-down) connections through projections to inhibitory interneurons (Bastos et al., 2012; Roux & Buzsáki, 2015). While neuroimaging indicates rIFG to pre-SMA connectivity

scales with inhibitory success (Duann et al., 2009; Rae et al., 2015), fMRI temporally filters fast activity underlying response inhibition. Additionally, fMRI connectivity methods cannot discern among top-down and bottom-up interactions or synaptic gain mechanisms. Therefore, it remains unknown whether rIFG top-down interactions with pre-SMA directly regulates inhibition (Aron et al., 2014) or attention and detection of context-changes (Munakata et al., 2011; Sharp et al., 2010; Xu et al., 2017).

To address these gaps, we combined neurostimulation of rIFG using transcranial focused ultrasound (TFUS) with electroencephalography (EEG) to record event-related potentials (ERPs) in humans performing a stop signal task. The N200/P300 ERP complex has been shown to be a strong indicator of inhibition. N200 amplitude typically diminishes during successful stopping, coinciding with P300 onset latencies. Evidence points to P300 onset as reflecting a top-down instantiation of the stopping process because it reliably tracks inhibition success and stop signal reaction times (SSRT; Wessel & Aron, 2015). As we found TFUS improved stopping through shortening the SSRT, we predicted P300 onsets would track the TFUS induced change in success and SSRT.

To determine whether rIFG deploys response inhibition through top-down inhibitory mechanisms that generate the link between P300 onsets and inhibition speed, we utilized dynamic causal modeling (DCM). We formally compared models with top-down control mechanisms predicted from direct inhibitory models versus attentional orienting. Network models of inhibition posit top-down modulation of inhibitory interneuron gain over action coding areas (Lo et al., 2009). In contrast, attentional orienting is linked to top-down gain modulation (Desimone & Duncan, 1995) of units

driving bottom-up responses in a fronto-parietal network (Corbetta et al., 2008). Bottom-up activity is instantiated by the superficial pyramidal cells within DCM microcircuit models (Bastos et al. 2012). Indeed, predictive coding models of attention are linked to the recurrent-gain of superficial pyramidal populations within the microcircuits (Feldman & Friston, 2010). These predictions have received empirical support from DCM studies (Auksztulewicz et al., 2018; Brown & Friston, 2013) indicating attentional processing engenders frontal modulation of superficial gain in fronto-parietal network areas. Here we aimed to dissociate between inhibitory and attentional frameworks by using TFUS with formal comparison of DCMs with top-down rIFG gain modulation over inhibitory interneurons, superficial pyramidal units, or both. We hypothesized that if rIFG connectivity directly triggers inhibition and scales stopping speed, rIFG should modulate inhibitory interneuron gain in pre-SMA.

Methods and Materials

Participants

Healthy adults were randomly assigned to one of three experimental groups. The main experimental group received transcranial focused ultrasound (TFUS) stimulation to right inferior frontal gyrus (rIFG) (n = 25; 19 males, mean age = 24.1 yrs., SD = 3.2 yrs.). A second group received stimulation to the ipsilateral somatosensory cortex (n = 23; 15 males, mean age = 22.4 yrs., SD = 3.3 yrs.) and was used as the cortical site active control group (S1). A third group received a sham stimulation near the right temple (n = 15; 8 males, mean age = 24.2 yrs., SD = 2.8 yrs.) and was used as control for possible auditory effects of TFUS (sham rIFG). All individuals were right-handed (self reported) and received financial compensation for participation in the study. Before enrollment,

each subject was screened for neurological disorders and a history of epilepsy, stroke, or brain injury. A neurologist from Barrow Neurological Institute (Phoenix, AZ) screened all subjects' T1 MRI and cleared them before study participation.

Behavioral Task

Response inhibition was assessed using the Stop Signal Task involving both 'Go' and 'Stop' trials (Fig. 1A) programmed in Opensesame (Mathôt et al., 2012). Each trial started with a central fixation cross. In every trial, fixations were replaced by a green 'Go' circle (3° x 3° visual angle), separated by an exponentially-distributed time interval (mean: 500 ms; standard deviation: 50 ms). Subjects were instructed "to press the up key when detecting the Go circle" (top panel, Fig. 1A). In 'Go' trials, the circle vanished either after responses or 800 ms elapsed. In 'Stop' trials, the stop cue was a red square which appeared around the green circle (middle and bottom panel, Fig. 1A). If the subject successfully inhibited their response with respect to the Stop cue within 800 ms, the red square was extinguished, and the trial was considered a successful inhibition. The time required to inhibit a response following the Stop signal is defined as stop signal reaction time (SSRT) (see below). Timing of the Stop cue relative to Go cue, i.e., the stop signal delay (SSD), was presented at one of four fixed, but subject-specific SSDs. The SSDs were chosen by having each subject perform a practice block of 50 Go trials to determine their baseline Go reaction time (RT). After this block, the 4 SSD levels were set to 25, 35, 75 and 95% of the mean Go RT. These SSDs were fixed throughout the experimental session and were presented in a random order across Stop trials. Using a set of fixed SSDs allowed us to calculate the SSRT using routines that

are less susceptible to low trial numbers (Matzke et al., 2013; see Data processing). All trials were separated by a 2s inter-trial interval (±300 ms random jitter).

Transcranial Focused Ultrasound design.

Transcranial Focused Ultrasound (TFUS) was delivered either simultaneously with (1) the Go signal in both Go and Stop trials, or (2) the Stop signal (Fig. 1A). We delivered TFUS during Go trials to determine the effects of TFUS on neural responses (see below) to rIFG independent of a stopping signal, and assess stopping behavioral effects are related merely to alteration of a Go related process. We designed five types of trials. The first two consisted of Go trials with no TFUS or with TFUS locked to the Go cue (No-TFUS and Go-TFUS trials, respectively). The other three trials consisted of Stop trials: No-TFUS trials, Go-TFUS, and Stop-TFUS trials where TFUS was locked to the Stop signal. TFUS delivery for Stop trials was evenly distributed across the 4 SSD levels.

The overall probability of a stop trial was set to 35% of all trials. This proportion of trials accommodates the need for enough Stop trials to examine TFUS effects on Stop trials across all SSD levels, while still making the occurrence of Go trials more frequent. Than Stop trials Each experimental session consisted of 1200 trials distributed across 12 blocks. Blocks were segmented into stimulation and no-stimulation blocks; the former containing trials with and without stimulation, and the latter containing no stimulation. Trial types (Go and Stop trials) were randomly and evenly distributed throughout the experiment. We chose a blocked design to mitigate possible carry-over effects of the stimulation across trials. By using two control groups (S1 and sham rIFG),

we could determine the extent to which behavioral and/or neural responses associated with TFUS were specific to the target site (rIFG).

EEG and structural imaging acquisition

EEG recording. EEG was recorded using a 64-channel ActiCap system (BrainVision, Morrisville, NC), with 10–20 layout. Data was recorded at a sampling rate of 5 kHz, with 0.1 μV resolution and bandpass filter of 0.1–100 Hz. Impedances were kept < 5 kΩ. Online recordings utilized a ground at AFz and left mastoid reference. At the beginning of each session, electrode layouts with respect to each individual's head shape were registered using the left and right preauricular, and nasion as fiducial landmarks. This allowed for later co-registration with each individuals T1 structural MRI scan and for source-localized analysis (see below).

Structural MRI (T1). For guiding TFUS neuronavigation and co-registering EEG electrode placement for source analysis and modeling, we obtained a structural T1 MRI scan for each participant. T1 volumes were collected using an 3D MPRAGE sequence (TR = 2300 ms, TE = 4.5 ms, 1 x 1 x 1.1 mm^3 voxels, field of view 240 x 256 mm^2, 180 sagittal slices) in a Philips Ingenia 3T scanner with a 32-channel head coil. Brainsuite was used to process T1s, which included cortical extraction sequence and a surface label-registration procedure with the BCI-DNI atlas. After labeling, we checked the locations and created a mask of either pars opercularis (rIFG group) or the centroid of ipsilateral S1 (control group). This volume labeling and mask creation procedure were used for guiding TFUS target identification.

TFUS targeting, setup and parameters

A BrainSight neuronavigation system (Rogue industries) along with subjects' T1 scans were used to guide placement of the focused ultrasound transducer beam profile for stimulation. This was done separately with respect to each individual's neuroanatomy and mask created from T1s. The first step involved creating a subject-specific mask from cortical atlas registration and projecting into the Montreal Neurologic Institute (MNI) coordinate system. When planning the TFUS target, we considered both MNI coordinates and individual anatomy. For example, metanalysis (Chikazoe et al., 2009; Levy & Wagner, n.d.) have shown specific activation of the pars opercularis (around x=48, y=16, x=18) for contrasts of successful inhibition versus Go trials and successful versus failed inhibition trials. In the case of the rIFG group, we first identified the pars opercularis MNI coordinates. During target planning, we confirmed the coordinates were inside the anatomical region of pars opercularis. We visually confirmed each subject's pars opercularis TFUS target was rostral to the inferior precentral sulcus and dorsal to the sylvian fissure, and ventral to the inferior frontal sulcus. For the S1 group, stimulation was targeted near MNI coordinates of x=-43, y=-29, z=54 and within the left post-central gyrus.

Before TFUS transducer setup, neuronavigation registered subject's T1 scans in virtual space, with their head and the ultrasound transducer in real space. Alignment and cortical registration were performed using nasion, tip of the nose, philtrum, and left and right periauricular notch and tragus as fiducial landmarks. A 3D printed housing held the TFUS transducer and optical trackers, and silicon spacers (ss-6060 Silicon Solutions, Cuyahoga Falls, OH). Acoustic gel was applied to both transducer and scalp. We recorded stimulation target coordinates after placing the transducer in target

alignment. In the sham rIFG control group, we employed a sham TFUS (Legon et al., 2018) by placing the transducer perpendicular to the rIFG target. The accuracy within non-sham groups was measured by tracking the deviation of the TFUS beam profile from the cortical target throughout the experiment. During the experimental session, we sampled TFUS transducer spatial target deviation during each break. Accuracy was very high, with an average deviation of ±1.5 mm displacement across all subjects and sessions.

TFUS setup and parameters in this experiment were nearly identical to those used in Legon et al (2014). Briefly, we used a single-element TFUS transducer with a center frequency of 0.5 MHz, focal depth of 30 mm, a lateral spatial resolution of 4.5 mm^2, and axial spatial resolution of 18 mm^2 (Blatek, Inc., State College, PA). TFUS waveforms were generated using a two-channel, 2 MHz function generator (BK Precision). The system operated by channel 1 triggering the computer which produced a pulse repetition frequency (PRF) of 1.0 kHz. Channel 1 triggered channel 2, which produced short bursts at the 0.5 MHz acoustic frequency. This produced an ultrasound waveform with a carrier frequency of 0.5 MHz, PRF of 1.0 kHz, and duty cycle of 24%. Each stimulation duration was 0.5 s. Transducer power was driven by output from a 40-W linear RF amplifier (E&I 240L; Electronics and Innovation). It has been previously verified that the resulting waveform does not cause heating of skin or skull bone (Legon et al., 2014).

Computational simulation of TFUS propagation

We quantified peak pressure amplitude, peak intensity and accuracy of the beam distribution with TFUS target to rIFG using the pseudospectral simulation method in K-

wave (Treeby & Cox, 2010). Reference peak pressure planes for the simulations were derived from previous data (Legon et al., 2014). Simulation parameters were first validated by simulating the transducer in water to compare the simulation results with those from previous water tank tests. The max pressure plane at the 30-mm focus was used as a source input pressure for the transducer during the simulation. The transducer was modeled to have a 30-mm radius of curvature. Water simulations used a homogenously medium of water density (1000 kg/m^3) and speed of sound (1482 m/s). We created a computational grid over a 256 x 256 x 256 with 1-mm spacing. The points per wavelength were 6, Courant–Friedrichs–Lewy = 0.1, and simulation time was set to 6 pulses (duration = 250 μs) to ensure simulation stability.

For simulating transcranial ultrasound, we extracted 3-dimensional maps of the skull from a CT (1-mm resolution) and brain from T1 MRI scan (1-mm resolution) from three preoperative patients at Barrow Neurological institute. The MRI and CT were both co-registered and normalized to the MNI space in SPM12. Figure 1B shows one subject's T1 and scalp in the rIFG group, renderings of the transducer housing and the pars opercularis mask. To mimic our approach of targeting used in the experiments, we surface registered the gray matter volume to the BCI-DNI atlas and identified the centroid of pars opercularis. This allowed us to map from world coordinates of the scan to MNI coordinates of the target (Fig. 1D). The average stimulation location for these three subjects was x = 48, y = 18, and z = 6. Conversion from Hounsfield units in the CT to sound speed and density were done using the relations described in Aubry et al (2003). All skull materials were set using these parameters, while other tissues were treated as homogeneous with parameters set to that of water. Attenuation was modeled

as a power law with a $\beta = 0.5$ and absorption was also modeled with a $b = 1.08$ (Treeby and Cox, 2010).

Numerical simulations for predicting TFUS accuracy were based on CT and MRI data from 3 preoperative patients. A visualization of the rIFG group normalized T1 and transducer placement is shown in Figure 1A, along with non-normalized T1s and transducer placements for four subjects (Figure 1B). To assess transcranial stimulation accuracy, the simulated transcranial transmission was compared against simulations of TFUS transmission through water. Differences between these simulations shows the estimated effect of any power absorption and change in acoustic profile after skull transmission. Numerical simulation parameters (see above) were derived to ensure the water simulation here matched the water tank results from a previous study using the exact same transducer and TFUS experimental parameters (Legon et al., 2014). Simulation of ultrasound through water predicted a max pressure of 1.05 Mpa and spatial peak pulse average intensity (Isppa) of 22.4 W/cm^2 at the focus (Figure 1C). The lateral full-width at half maximum of the max pressure in water simulations was 4.39 mm (Figure 1C). These model simulations closely align with previous studies using the same transducer and experimental TFUS parameters.

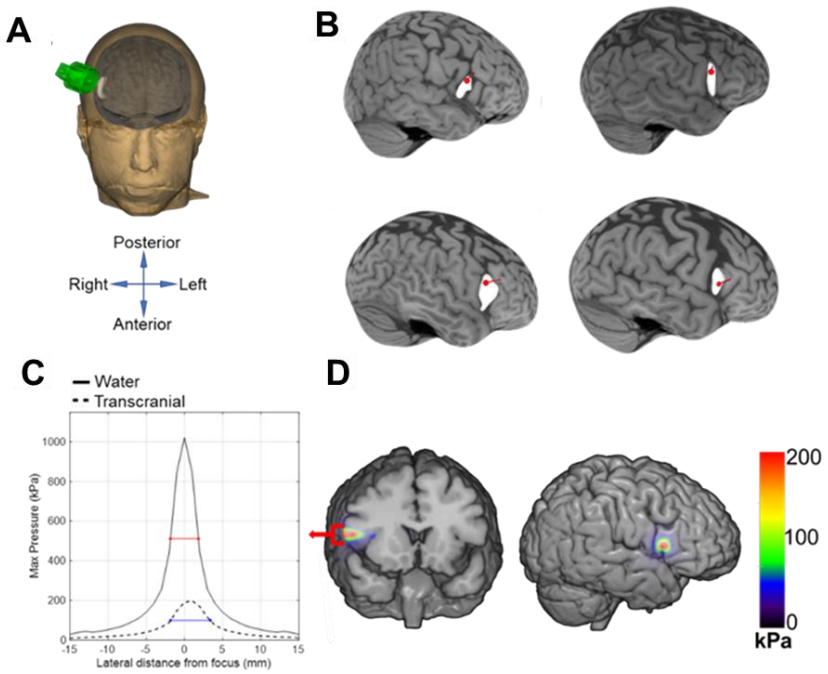

Figure 1 **A**. The left plot shows the average neuronavigation location of TFUS applied to all MRIs used in the rIFG group. **B**. The structural scans on the right show renderings of the targeted TFUS point in pars opercularis in 4 subjects. **C**. Lateral maximum pressure profile obtained at 30-mm depth focus in both water and transcranial ultrasound simulation on a CT scan from one patient (solid and dotted lines, respectively). Horizontal red and blue lines denote full-width half maximum of the spatial profile of lateral pressure. **D**. Simulated transcranial pressure profile onto T1 MRI plot shown as a color overlay.

Comparison of simulations and the previous water tank data indicated a 97% match of pressure/intensity at the focus taken over a 5 mm^3 voxel section in all 3 planes at the focus. Next, modeling of transcranial transmission predicted an average maximum intensity of 2.8 W/cm^2, which is the intensity range of non-thermal neuromodulation and exhibits nearly instantaneous measurable effects on EEG (Legon et al., 2014). Comparing the water and transcranial simulation, accuracy was assessed by comparing shifts in peak pressure (Figure 1C). Skull transmission compared to water was shifted 1.25 mm laterally and had a lateral beam profile full-width half maximum of 5.1 mm (Figure 1C). These transcranial simulations indicate high spatial precision, with

>95% of pressure or energy (kPa) being constrained to pars opercularis sector in the rIFG group (Figure 1D).

Statistical analysis

Behavioral variables*.* Our behavioral analyses focused on the following variables: Go trial reaction time (Go RT), percentage of successfully inhibited responses on Stop trials (successful stopping) per SSD, failed inhibition reaction time, and SSRT. The SSRT was estimated using a hierarchical Bayesian parametric approach (Matzke et al., 2013a) that estimates distribution of SSRTs while assuming an ex-gaussian parametric form. We chose this approach as Matzke et al. (2013a) showed that it performs well even when there are only a few trials available per SSD level. This SSRT estimation procedure was run separately per group (rIFG,S1, and Sham) and trial types (No-TFUS Stop trials, Go-TFUS Stop trials, and Stop-TFUS Stop trials). As we report in the Results section, we used combined RTs from Go trials with and without TFUS because stimulation did not alter the Go RT.

The probability of response inhibition changes over levels of SSD, P(respond|signal), and across TFUS conditions were assessed within and across groups by fitting a 2-parameter logistic mixed effects model with random intercepts and slopes to obtain subject- and condition-specific model parameters. P(respond|signal), denoted as p, were converted to a negative logit ($\log((1-p)/p)$) before fitting. As our main goal was to estimate the logistic curve slope (**β)**, we ran the mixed-effects model (using LME4 in R) with the full interaction of SSD and stimulation condition (no-TFUS, Go-TFUS, Stop-TFUS). Logistic slopes per subject were estimated by combining fixed and

random coefficients. **β** parameters were analyzed using a mixed-design ANOVA on **β** with factors of Group (3 levels) and TFUS (3 levels: No, Go, Stop).

EEG pre-processing

Continuous EEG data were first down-sampled to 250 Hz, then high-pass filtered (0.1 Hz) and re-referenced to the scalp average. Any channels that displayed artifacts for more than 25% of the total session were removed before further processing, but were later interpolated after artifact rejection. We removed channels that were designated unsuitable for analysis by visual inspection and absolute temporal standard deviation (> 5) across channels. It is important to note that, in each of the stimulation groups, the cortical sites of rIFG and S1 were close to the F8 and CP4 electrodes. Therefore, these electrodes could not be used for EEG recording in their groups and were interpolated after artifact removal. Remaining data were epoched from Stop trials locked to stop signal onset (-200 to 600 ms peristimulus). Individual epochs were rejected from analysis if they contained large scalp EMG or rare events (< 8% of all trials). Out of the 63 participants, 3 subjects were excluded from analyses due to EEG recording issues (impedance >25 kΩ across channels). The remaining data were bandpass filtered from 0.5-40 Hz. EOG artifacts were removed using Independent Components Analysis using eeglab (ICA; Delorme & Makeig, 2004). On average, 3.35 components were removed per participant. Remaining ICA components were back-projected to electrodes and then averaged to create event-related potentials (ERPs) per condition, e.g., different SSD levels for No-tFUS Stop trials). The ERPs baseline were normalized by subtracting the activity from -100 ms to the stop signal onset. Because we later applied a Hanning taper to the edges of the ERPs before dynamic causal

modeling, we applied the same procedure to ERPs after cleaning using a Hanning taper.

Isolating inhibitory response ERPs

As noted in the Introduction, our analysis focused on sensor, source, and source network models of ERPs because several types of ERPs have been linked to the success and speed of stopping. Due to the macroscopic nature of EEG signals, there is likely a temporal overlap of neural processes related to stopping and going during Stop trials. We aimed to separate Go-related activity from stopping-related activity separately for both successfully (SS) and unsuccessfully (US) inhibited Stop trials. This was done by subtracting the Go trial ERPs from the Stop trial ERPs, and using the approach employed by Mattia et al (2012). On a per-subject basis, we found Go-trial ERPs (No-TFUS and TFUS) that had RTs latency-matched to SS trials based on each subject's SSRT. These Go trials had to have RTs either equal to or greater than the SSRT. For unsuccessful stopping trials, we found latency-matched Go trials with RTs by first calculating each subject's mean signal-respond RT for each of the two highest SSDs. We then calculated the difference in SSD (ms) and searched for Go RT trials for each SSD that fell within the mean signal-respond RT ± half the difference of the SSD (ms). This was done to prevent overlap of activity from both faster and slower Go RTs and signal-respond RTs. These steps were performed separately for the highest and second highest SSD. This procedure was done separately for SS and US trials for both TFUS conditions. After correcting the SS and US stop trial, the corrected ERPs were averaged across the two highest SSDs per subject (corresponding to the 85% and 105% mean Go RT of each subject). These ERPs were used for the remaining analysis.

ERP analysis

Sensor-space analysis. ERPs were examined at the sensors level using permutation-based dependent samples t-tests that account for spatiotemporal clustering. We performed tests utilizing multiple-comparisons with cluster-based p-value corrected at $p < 0.01$, and used 5000 permutations for each contrast. Two different contrasts were used to analyze the differences for stopping success overall and TFUS effects on ERPs. The contrasts included comparison of (1) successful stop (SS) – unsuccessful stop (US) trials collapsing over No and Stop-TFUS conditions, (2) a TFUS effect comparison of SS (No-TFUS) – SS (Stop-TFUS) contrast, and (3) the interaction contrast of SS – US trials with No-TFUS and Stop-TFUS. The first contrast is typically used to determine which areas exhibit ERPs (or brain areas) that differentiate successful inhibition. The second contrast was used to determine which scalp ERPs differentiated successful stopping in the No-TFUS and Stop-TFUS conditions. The third contrast (interaction) was used to determine how the SS-US contrast differed between No-TFUS and Stop-TFUS conditions. The second and third contrasts were the primary focus of our analyses and expected to provide the most direct indication of inhibitory specific ERP effects. Our reasoning was as follows: if the first SS-US contrast reveals ERP activity and latencies that differentiate successful and failed inhibition, and TFUS alters the potency of inhibitory processes and ERPs that differentiate the level of successful inhibition (second contrast) , the interaction contrast should isolate changes in neural response strength to stop signals and behavioral inhibition outcomes that are jointly modulated by TFUS.

Because recent work has indicated the frontocentral (ERP) P300 onset latency, rather than peak amplitude, is related to the stopping speed (SSRT) across-subjects (Wessel and Aron, 2015), we regressed the between-subject changes in SSRT as a function of the P300 latency change between No-TFUS and Stop-TFUS. To achieve this, we computed the shift in P300 onset crossings between No-TFUS and Stop-TFUS conditions in two steps. First, we took the across-subject mean frontocentral ERP waveform in a time-window of ±50 ms around the zero crossing. To calculate each subject's zero-crossing time, we calculated the dynamic time warping distance from the template mean ERP to the subject's ERP. Second, this distance was added to the median zero-crossing time to obtain an individual subject crossing for both the No-TFUS- and Stop TFUS-locked conditions. These changes in P300 onsets were then regressed against individual subject differences in SSRT between conditions.

Source-localization and source statistics. To localize source generators of ERPs differentiating inhibitory outcomes and provide initial locations for our network connectivity, we estimated the ERP activity in source space. This step employed a multiple sparse-priors approach as implemented in SPM12. The approach estimates the cortical source activity of sensor ERPs from the SS and US trials across TFUS conditions. We chose this Bayesian estimation method because it has been shown to have good accuracy during simulation recovery studies (Belardinelli et al., 2012), and this method focuses on finding the simple, sparse solutions given the ill-posed nature of source localization. Cortical meshes were used to calculate the EEG forward solution using a boundary element head model based on each subject's T1. Source inversion was performed over a window starting from the stop signal up to 450 ms to capture

activity occurring over the whole stopping period. This window of source data was converted to a 3D image, spatially smoothed (8 mm full-width half maximum) and interpolated into MNI voxel space for each subject and condition. This was converted into a statistical parametric map (SPM) and analyzed used a flexible factorial design to implement a repeated-measures ANOVA to examine the main effect of inhibition success (SS or US trial). The resulting SPM was analyzed with a threshold set at $p < 0.005$ (peak-level, uncorrected) and cluster-wise family-wise error rate $p < 0.05$.

Dynamic Causal Modeling

Dynamic causal modeling (DCM) was used to analyze different neural models of effective connectivity. This approach was applied to determine the network changes underlying successful inhibition and how they unfolded during successful stopping in both No-TFUS) and Stop-TFUS conditions. Furthermore, we used the DCM to examine how different sources and their connectivity under response inhibition and TFUS conditions lead to the generation of ERPs linked to successful stopping with a focus on the P300.

DCM is a source analysis that estimates how each neural source – modeled using biophysically based neural mass models – and their connectivity relates to measured ERP sensor data. DCM has now been applied in countless studies to examine effective source connectivity in EEG/MEG (for overview see Moran, 2015). The approach uses Bayesian source model evidence to map the estimated source data to the sensor data by using the lead field generated based on each subject's T1 MRI. Because DCM attempts to explain the measured data, we can compare and optimize different source activity networks and connectivity structures to infer the most plausible

explanation for the data as is routine in modeling procedures. Additionally, because DCM for ERPs derives from physiological models driving source activation and connectivity, while using a forward model mapping source to sensor space, it is valuable for differentiating how connectivity changes drive ERPs in experimental conditions.

Sources were based on previous neuroimaging studies of inhibition tasks, and the MNI locations used in the DCM were based on those found in our source analysis (see Results). We used source analysis areas that have been previous linked to inhibitory and attentional control (Corbetta and Shulman, 2001; Swaan et al., 2009; Wessel and Aron, 2017). This analysis revealed evoked power differences during inhibition in areas conforming to the right lateralized network typically found during inhibition (Boehler et al.,, 2010; Levy & Wagner, 2011b; Sharp et al., 2010). Locations included rIOG, rIFG, rTemporal, rParietal, and pre-SMA, with the source analysis results used as source locations in the DCM. The MNI coordinates employed in the DCM for each source included rIFG ([48,28,4]), pre-SMA ([0, 24, 54]), rParietal ([40,-54,50]), rTemporal ([52,-18,-12]), and right inferior occipital gyrus (rIOG) ([46, -76,10]). These locations are in strong agreement with those found in meta-analyses of inhibitory control tasks (Rae et al., 2014).

Sources were modeled with the canonical microcircuit neural mass model that has been shown to embody hierarchical Bayesian predictive coding theories of brain function (Bastos et al., 2012). This model has now been used several times in the DCM framework to model EEG/MEG connectivity during studies of attention and stimulus prediction (Auksztulewicz & Friston, 2015; Brown et al., 2013). The model implements source data generation as four distinct neuronal populations that include the superficial

and deep pyramidal cells, spiny stellate cells, and inhibitory interneurons. Each of these layers also has recurrent inhibitory gain connections that model layer-specific inhibitory neurons. Connectivity between each population is derived from laminar-specific connectivity patterns in the cortex (Douglas & Martin, 2004; Felleman and Van Essen, 1991; Shipp, 2007). Because connections between sources are implemented as either forward (i.e., bottom-up) or backward/feedback (i.e., top-down) connections, this allows us to distinguish how specific types of connectivity are altered during both successful inhibition and TFUS. Central to using this DCM modeling was arbitrating between neural mechanisms required for a direct inhibitory control model versus an attentional control model. Specifically, we were concerned with the top-down interaction between rIFG to pre-SMA and temporal cortex. Both frameworks and physiology suggest a rIFG driven top-down, activity dependent modulation of the recurrent gain in pre-SMA. Direct inhibition models predict rIFG descending activity should alter inhibitory interneuron gain in pre-SMA. Attentional models predict rIFG descending activity should directly alter recurrent-gain of superficial pyramidal populations that drive bottom-up excitation. However, previous work proposing rIFG as driving attentional reorientation does not specify what the neural target of rIFG modulation would be. For example, attentional modulation could follow from rIFG modulation of temporal cortex or to pre-SMA and having a chain effect on parietal areas also responsible attentional control. However, such a mechanism likely operates alongside a direct inhibition from rIFG to pre-SMA. To address all of these possibilities, we considered several permutations of models in which the gain was modulated either independently of top-down activity (e.g., pre-SMA isolated gain changes) or was modulated by top-down activity from other sources (e.g.,

from rIFG to pre-SMA). We also considered whether these gain changes occurred in superficial pyramidal or inhibitory interneuron populations estimated within a source and considered several different mixture variants. For example, some models had rIFG modulate inhibitory interneuron gain in pre-SMA and pre-SMA exerted gain modulation over parietal superficial pyramidal units.

Model building first involved estimating the connectivity structure best explaining the ERP data (DCM ‘A’ matrix), determined using a series of Bayesian model selections with family-wise random effects (Penny et al., 2010) by fitting each DCM to each subject’s non-TFUS successful inhibition trials in a time window spanning 0-450 ms (0 = stop signal onset). These trials were chosen as a base condition to determine the optimal model without effects due to TFUS. Model spaces were partitioned into several families based on a (1) pre-frontal hierarchy structure, (2) structure of the lower hierarchy, (3) connections from lower areas (rTemporal and rParietal) to higher areas (rIFG and pre-SMA), (4) whether exogenous gaussian bump inputs were in either rIOG and rIFG or just rIOG.

We further optimized the model again to estimate the condition effect modulation on recurrent gain and extrinsic connections (‘B’ matrix) that differentiated successful and unsuccessful inhibition. The optimized DCM structure were fit to individual subject’s ERPs for both the No-tFUS successful and failed inhibition trials together. The effects of intrinsic gain were examined over 2 families. The first family was modulation in superficial pyramidal versus inhibitory interneurons. The second family assessed whether the post-synaptic gain effects were activity-independent or activity-dependent. They differ with activity-dependent gain meaning an areas recurrent gain is modulated

by the level of top-down descending activity (i.e., voltage dependent) from another. All of these effects were tested combinatorically over all areas, such that some sources could have superficial or inhibitory populations modulated either activity-independent or -dependent. The failed inhibition trials were treated as a baseline when modeling these effects. Rather than optimizing the extrinsic connections before statistical analysis, we allowed them all to modulate between conditions (inhibition success) as their effective changes were tested using a hierarchical model across subjects (see below). After determining the DCM gain effects, the finalized DCM were fit to individual subjects' ERPs across successful and unsuccessful inhibition trials, and No-TFUS and Stop-TFUS conditions.

Analysis of DCM results

To analyze the resultant condition-specific connectivity changes, we employed the Parametric Empirical Bayes (PEB) framework (Friston et al., 2016). PEB is a general linear hierarchical modeling approach, which was used to perform hypothesis testing regarding how connectivity parameters both changed on average between No-TFUS and Stop-TFUS conditions and corresponded to scaling of stopping efficiency (SSRT). Rather than considering all possible connections, we constrained the analysis to those top-down effects involving rIFG, and pre-SMA to parietal connections. This subset was chosen because they correspond to the core interactions underlying our hypotheses. Once the group-level GLM parameters were estimated with respect to modulations of extrinsic and intrinsic connectivity, hypothesis testing proceeded using Bayesian model reduction of the GLM. This involves turning off/on different parameters and comparing free energy of reduced models, similar to performing model likelihood

ratio tests commonly used in mixed-modeling. At a design level, we tested for mean connectivity effects differentiating successful and failed inhibition and their interaction with TFUS conditions. The interaction term was also examined with respect to SSRT and mean effects. This addresses the expectation that a connection directly related to triggering inhibition and scaling its potency should be predicted by behavior in non-TFUS trials and behavioral changes during TFUS. After model reduction, the reported PEB model parameters were computed through Bayesian model averaging and cutoff to effects having a > 95% posterior probability.

RESULTS

Human participants performed a Stop-Signal task (Figure 2 top) with online TFUS. Subjects were divided into groups according to TFUS stimulation type: (1) active stimulation to right pars opercularis (rIFG), (2) active stimulation control site (ipsilateral primary somatosensory cortex, S1) to account for non-site specific TFUS effects, and (3) sham stimulation to account for TFUS auditory artifacts (sham rIFG). Since TFUS has been demonstrated to elicit nearly immediate effects on ERPs (Lee et al., 2016), stimulation was applied online (see Methods) at the onset of “go” or “stop” signal cues, separately for both Go and Stop trials (Figure 2). We hypothesized that if rIFG drives motor inhibition, TFUS behavioral effects would be limited to alteration of stopping during TFUS applied simultaneous with the stop signal (Figure 2 bottom row).

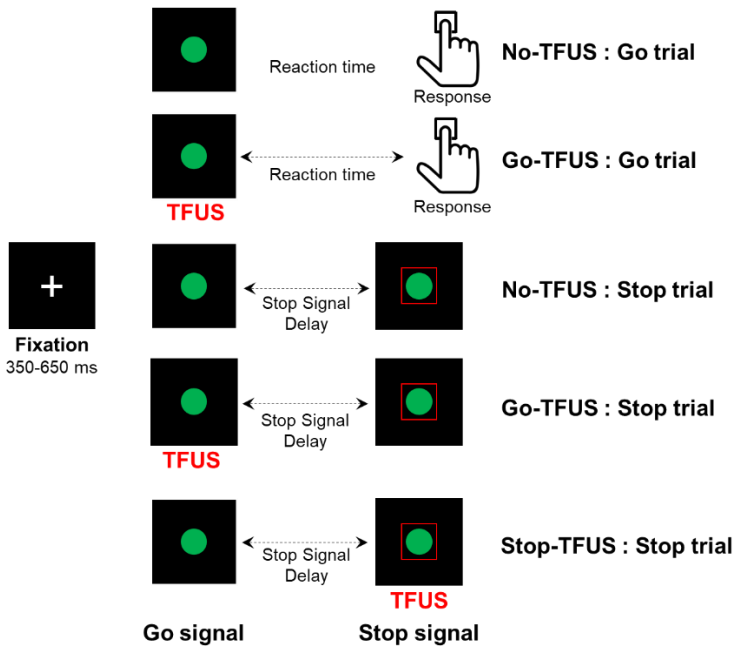

Figure 2. Timeline (left to right) of experimental trials for Go and Stop trials, and alignment with TFUS onsets. TFUS lasted 500 ms. Each trial had a fixation with a randomly chosen delay of 350-650 ms. In Go trials with TFUS (2nd row), TFUS onset simultaneous with Go signal. In Stop trials with TFUS, TFUS started either with Go signal (4th row) or Stop signal (5th row).

TFUS to rIFG improves stopping behavior

We first addressed how probability of failing to inhibit responses, P(respond|signal), changed across TFUS conditions and groups, by fitting a 2-parameter logistic mixed-model to obtain response inhibition curve slopes (**β**) across subjects and TFUS conditions (Fig. 3A). Analysis of **β** revealed only the rIFG group exhibited a TFUS-altered P(respond|signal). Importantly, behavioral effects of TFUS were only found for the rIFG group (Fig. 3A). ANOVA results indicated a significant Group x TFUS interaction ($F(2,50) = 3.8$, $p = 0.034$, $\eta_p^2 = 0.17$), and an overall effect of TFUS condition ($F(2,50) = 11.74$, $p = 0.002$, $\eta_p^2 = 0.29$). Follow-up one-way ANOVAs across TFUS onsets but within-groups, showed only the rIFG group exhibited differences across onsets. Follow-up t-tests in this group showed **β** for Stop-TFUS was lower than No-

TFUS and Go-TFUS conditions (both $p < 0.01$: mean **β**'s indicating change in probability for approximately 25% change in normalized SSD: No-TFUS = 0.35 (0.12), Stop-TFUS=0.27 (0.08), Go-TFUS=0.35 (0.11)) (Fig. 3A). Therefore, effects of stimulation were limited to the rIFG group and Stop-TFUS trials.

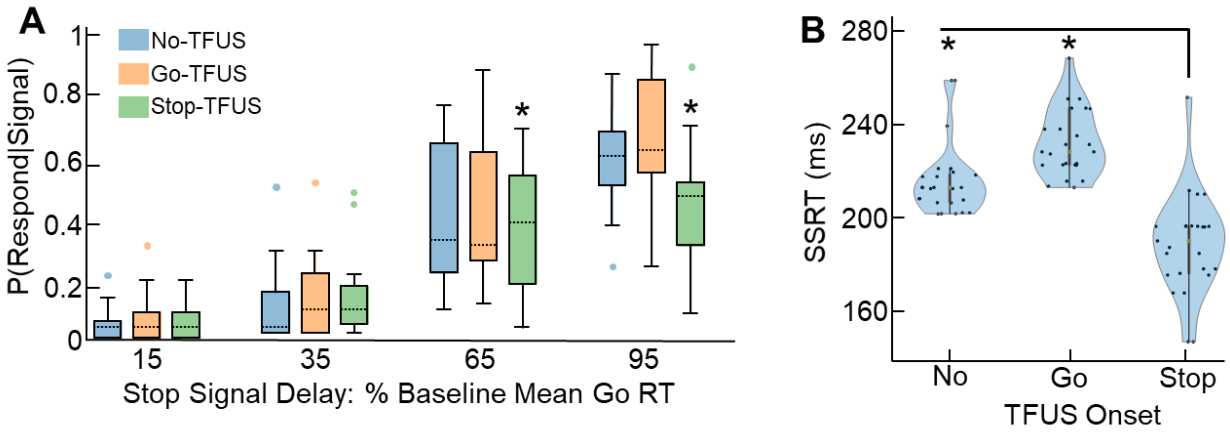

Figure 3. **A.** Probability of responding across stop signal delays in the rIFG group for all TFUS onsets. **B.** Violin plots of rIFG group across-subject distribution of stop signal reaction times (SSRT).

The effects of rIFG TFUS improvements to inhibition performance seemed predominant at longer SSDs (65% and 95% SSD; Fig. 3A). A repeated-measures ANOVA on P(respond|signal) for rIFG group across SSD levels and TFUS onsets supported this with a significant interaction ($F(6,102) = 8.21$, $p < 0.0001$, $\eta_p^2 = 0.33$). T-tests between Stop-TFUS and the average of No- and Go-TFUS across all SSDs indicated the interaction resulted from a reduction in P(respond|signal) for Stop-TFUS in the highest two SSDs (all $p < 0.01$; Bonferroni $\alpha = 0.0125$). These results indicate Stop-TFUS induced improvements of inhibition were more pronounced at longer SSDs (Fig. 3A).

Based on the prediction that rIFG drives an inhibitory process and our finding that TFUS improved response inhibition only in this group, we hypothesized rIFG TFUS changes to P(respond|signal) should result from a shortening of the stopping speed, i.e., SSRT; notably, TFUS did not affect other behavioral variables, e.g., Go RTs. SSRT analysis in a mixed-design ANOVA indicated a significant Group x TFUS interaction ($F(4,100) = 10.2$, $p < 0.001$, $\eta_p^2 = 0.21$). Follow-up one-way ANOVAs within groups indicated only rIFG group SSRTs (Fig. 3B) significantly differed between TFUS onsets ($p < 0.05$), with t-tests confirming that SSRTs were indeed shortest and only altered for Stop-TFUS trials (Fig. 3B). This indicates rIFG Stop-TFUS altered inhibition specifically by shortening the SSRT.

Separate analyses were used to determine whether P(respond|signal) changes from TFUS were also affected by non-inhibitory behaviors. We first addressed whether TFUS (rIFG, S1, or rIFG Sham) exerted any effects on simply responding to the Go signal by extracting mean RT from ex-gaussian distributions fit using maximum likelihood (Lacoutoure and Cousineau, 2008). Means for each subject were analyzed using a 3×2 mixed-design ANOVA with factors of Groups (3) and TFUS condition (2: No-TFUS and Go-TFUS trials). This analysis allowed us to assess if “going”, independent of “stopping”, was altered by potential TFUS auditory artifacts (sham group), stimulation to unrelated areas (S1 group), or whether TFUS to rIFG also influenced Go RT processes independent of a stopping context (rIFG group). We found no significant effect of TFUS onset, group, or their interaction (all $p > 0.05$). These results suggest that neither TFUS (rIFG and S1 groups) nor auditory factors alone (sham group) altered Go RTs independent of a stop signal.

Improvements in inhibition from TFUS could also have emerged from longer signal-respond RTs or reduced SSRT variability. Two mixed-design ANOVAs were used to examine the subject-level signal-respond RT means and SSRT variability with Group (3 levels) and TFUS (3 levels: no-TFUS, Go-TFUS, Stop-TFUS). We found no significant interactions or effects of TFUS on either the signal-respond RT or SSRT variability. Together, these behavioral results indicate that only TFUS to rIFG improved response inhibition by shortening a process related to the stopping speed.

Neural responses underlying inhibition

In the following results, we report analysis that focused on the rIFG group because this was the only group exhibiting behavioral TFUS effects. Furthermore, we only analyzed No-TFUS and Stop-TFUS conditions because Go activity was subtracted from neural data from Stop trials (see Methods). Sensor-level ERPs were examined in this group across three contrasts using cluster-based permutation t-tests: (1) successful – unsuccessful stopping contrast over each TFUS conditions, (2) successful (No-TFUS) – successful stopping (Stop-TFUS) contrast, and (3) interaction comparing successful – unsuccessful stopping between the No-TFUS and Stop-TFUS conditions. The peak amplitude effects are shown in Figure 4C and correspond to ERP time-windows and voltage deflections of an N100/P100, N200 and P300 that are typically found during inhibition.

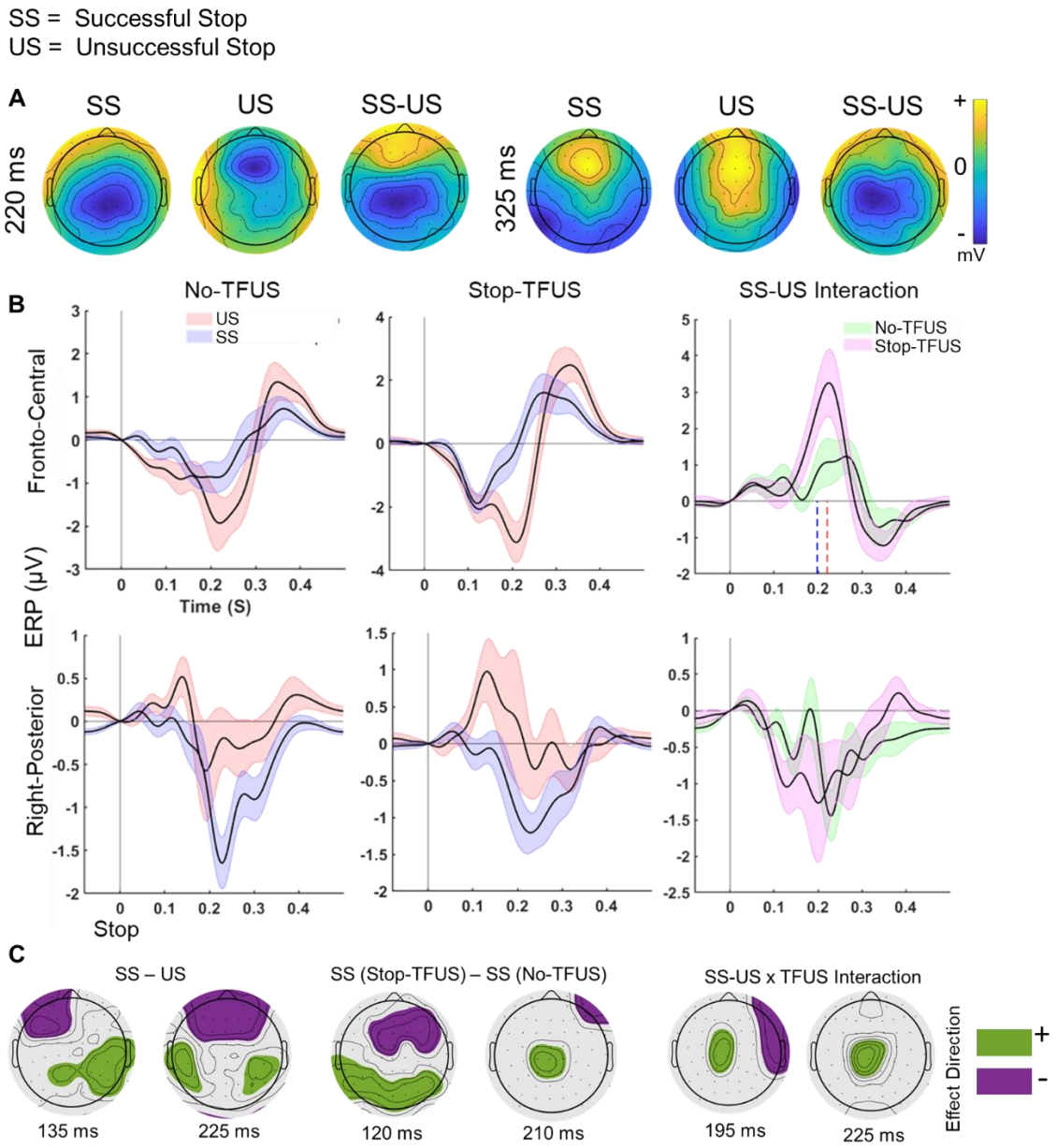

Figure 4. **A.** average scalp map of SS, US, and difference maps showing ERP amplitude at time points that aligned with canonical ERP effects in response inhibition. **B.** ERP plots across subjects for two different a priori chosen regions of electrodes to capture the typical N200/P300 in Fronto-Central electrodes (Cz and Fcz; top row), and N100 sensory responses in Right-Posterior (CP4 and P4; bottom row). Plots show mean ERP for SS and US trials across TFUS conditions (columns) and their interaction (right column). The dashed lines show the median SSRT across subjects for the No-TFUS (red line) and Stop-TFUS (blue line) under the interaction in Fronto-Central electrodes (right column; top row). **C.** Permutation sensor t-tests showing the maximal effects for three different comparisons: (1) SS-US, (2) SS trials across TFUS conditions, and (3) the interaction effect of SS-US and TFUS.

We observed an early (100-150 ms) right-posterior P100 response peaking around 135 ms that differentiated successful from unsuccessful stopping (SS and US, respectively; Figure 4C left). The location and timing of the peak and are indicative of a visual P100, visible in average time-course and t-stat map (Figure 4B and 4C). These peak amplitudes were larger in US trials. Contrasts of SS trials across TFUS conditions additionally revealed the P100 was smaller on Stop-TFUS SS trials (Figure 4C, middle). A difference in SS trials was also found for an N100 response in both frontal and right-frontal electrodes (Figure 4C middle). N100s were generally larger for the Stop-TFUS trials. Neither of these ERPs exhibited an interaction of SS-US and TFUS trials. These results indicate smaller amplitudes of early sensory responses predict successful stopping, but the lack of interaction indicates they were not directly related to inhibition.

The ERPs typically associated with indexing inhibition is the N200/P300 complex. Notably, this complex often appears in a fronto-central cluster and the N200 peak is most prominent during failed stopping. Examination of the scalp maps for SS and US trials in this time window shows the expected topography (Figure 4A). The fronto-central ERP plot (Figure 4B, top row) shows the N200 peaks around 200 ms, but does so most clearly in US trials. Comparisons of SS and US trials did support this effect of a larger N200 during US trials (Figure 4C, left). Additionally, comparing SS trials across TFUS conditions indicated this ERP was larger for the No-TFUS condition. This result is in line with the finding that smaller N200 responses are associated with inhibitory success. The interaction effect (Figure 4C, right) indicated the maximal difference corresponded to N200 timing, exhibiting a larger difference between SS-US difference during Stop-TFUS trials. N2 amplitude decreased with increasing inhibitory performance, with the smallest

peak in Stop-TFUS SS trials (Figure 4B, top row). The fronto-central P300 also differentiated both SS and US trials, with a lower amplitude for US trials. However, there was no interaction indicating it was linked to inhibitory performance.

Our main ERP hypothesis was that P300 onset timing reflects an inhibitory process rather than its peak. This hypothesis stems from several observations. The first is the N200 and P300 likely reflect a mixture of underlying components with the N200 amplitude being reflected in the P300 onset. In this case, a shift in P300 onset would drive our observed interaction contrast of the fronto-central N200 by altering the time at which the N200 peak amplitude is reached. Examining time-course of frontro-central ERPs (Fig. 4B top row) indeed indicated that reduction in N200 amplitude is likely driven by the P300 onset occurring earlier. The second motivation is that an inhibitory marker should track inhibition timing (SSRT), which has been found for P300 onsets (Wessel & Aron, 2015). Therefore, we predicted that if P300 onset timing was a direct indicator of inhibition timing, the P300 onset latency should correlate with the SSRT difference across TFUS conditions. Visually, contrasting waveforms of SS-US difference waveforms across TFUS conditions (Fig. 4B, upper right) supports our prediction that P300 onset shifted earlier in alignment with Stop-TFUS induced SSRT shifts. We found TFUS-induced changes in P300 onset significantly correlated with SSRT (0.61, $p < 0.05$), thus providing direct support that P300 latencies track inhibition speed.

Dynamic Causal Modeling of ERPs in the inhibition network

The combined effects of TFUS on behavioral and evoked responses indicated inhibition performance increased with a shortening in SSRT which was matched by a shifted P300 onset latency. Given that TFUS to rIFG altered inhibition, a core question

is to what extent rIFG cortical network interactions drive these behavioral changes and expression of evoked responses. To address this, we modeled evoked responses using Dynamic Causal Modeling (DCM) to determine which connections differentiated successful and failed stopping trials, and those that scaled parametrically with SSRT across TFUS conditions. Our focus was on a restricted network of areas that has consistently been found to be involved during inhibition. These areas include a fronto-parietal and cingulo-opercular network, centered around rIFG and pre-SMA, right temporal (rTemporal), right inferior occipital gyrus (rIOG), and right parietal (rParietal) areas (Levy & Wagner, 2011). To derive DCM locations for these sources, a source analysis on the mean difference between SS and US ERPs was performed in the 0-450 ms window. Source maps (Figure. 5A) revealed areas typically associated with successful inhibition. Maximal amplitudes of these source locations were used in the subsequent DCMs (see Methods).

A series of Bayesian model selections with family-wise random effects were used to determine the optimal structure by fitting different models to sensor responses during non-TFUS successful and unsuccessful inhibition trials to all subjects. The winning family for frontal hierarchies (Figure 5B, top) included recurrent connections between rIFG and pSMA, with rIFG having top-down (backward) connectivity to pre-SMA (exceedance probability: 0.97). The family analysis examining the lower hierarchy (Fig. 4B, bottom) revealed rIOG was, as expected, connected to rTemp and rPC, but these two areas were not laterally connected (exceedance probability: 0.89). Finally, the family examining lower to higher areas connections recapitulate known ventral and dorsal pathways (exceedance probability: 0.92). rIFG had top-down and bottom-up

connectivity with rTemp, and pSMA had top-down projections to rPC and forward connections to pSMA. Finally, we found the best model to have inputs at both rIOG and rIFG (exceedance probability: 0.88).

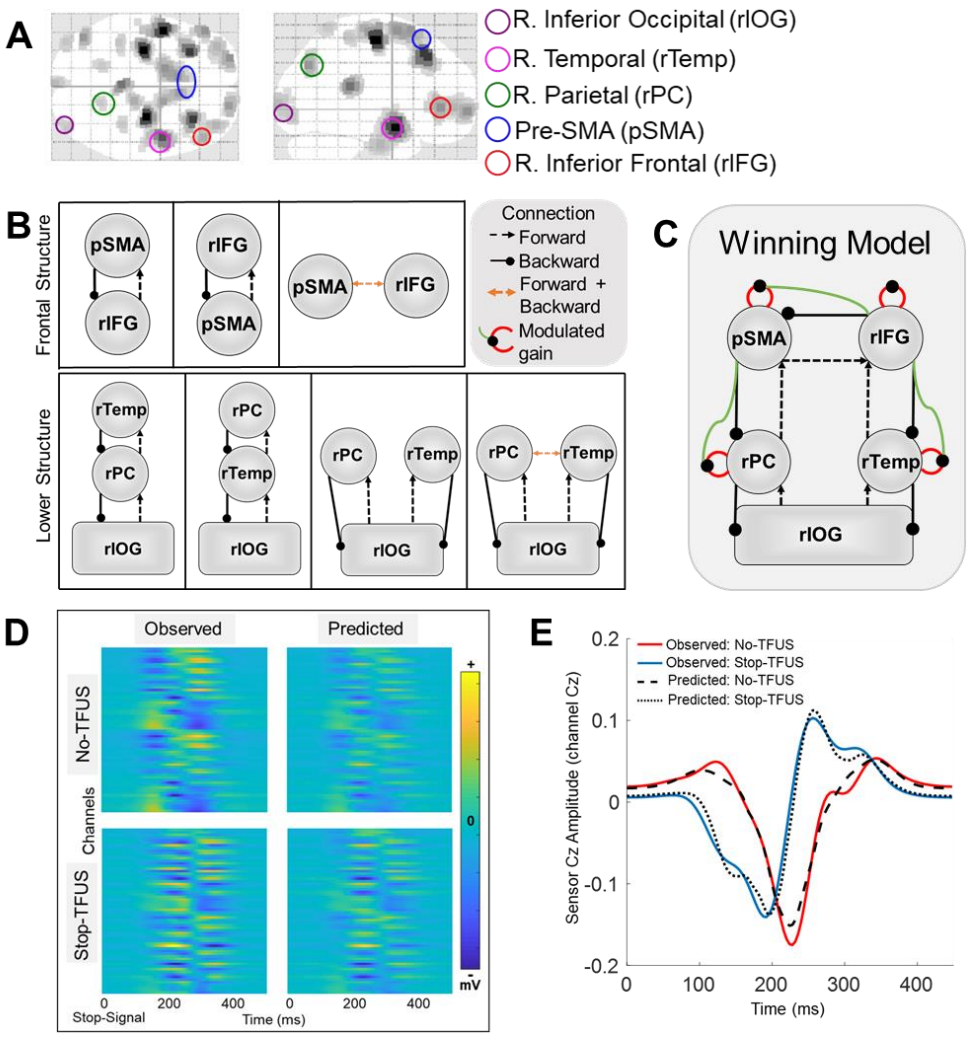

Figure 5. **Dynamic Causal modeling building: sources, structure, and fit**. **A.** Source map of Grand Mean ERP differences of SS and US trials from 0-450 ms. Circled nodes colored to show center of node coordinates used for the 5 regions in the DCM. **B.** Different configurational structures used to build base DCM and their connection types. **C.** Winning model showing the final configuration fit to individual subject data. **D.** Map showing the observed and predicted ERPs for all channels for the average of No-TFUS and Stop-TFUS trials. **E.** The subject mean Predicted and Observed ERP from channel Cz. Plotted for the successful stop trials across both

TFUS conditions. This shows how the model captured the differences in N200, and P300 onset timing and amplitudes.

Cortical control of stopping involves top-down rIFG gain control

A final set of model comparisons examined permutations of the above model with modulation of postsynaptic (recurrent) gain. We considered models wherein gain changes occurred in either superficial pyramidal or inhibitory interneuron populations, and whether the post-synaptic gain effects were activity-independent or -dependent and modulated by top-down activity. Model selection indicated strong evidence for a model with gain changes in inhibitory interneuron layers (probability: 0.96), all these effects being activity-dependent (probability: 0.93). Therefore, evoked activity during stopping was best explained by a network including gain changes that were dependent on top-down rIFG activity. The winning DCM used for further analysis is shown in Figure 5C. The core result from these model tests is that, although areas such as pSMA, rPC and rTemp are involved in generating inhibition related ERPs, the network involved top-down, activity-dependent gain modulation from rIFG.

The optimized DCM structure (Fig. 5C) was fit to individual's evoked responses as separate models for the TFUS conditions. Fitted models agreeably captured spatiotemporal properties of the ERP data across both sets of model fits (Fig. 5D) with an average fit of all DCM models R^2=92%. The capacity of the model to capture the main characteristics of the N200 and P300 ERPs is shown with the plot overlaying the observed and predicted Cz (Fig. 5E). The models were statistically analyzed in a parametric empirical Bayesian (PEB) model to determine which connections were modulated between SS and US trials and TFUS. As our hypotheses relied on top-down

coupling from rIFG and the effects of neurostimulation, for brevity we only consider these connections.

rIFG connections directly modulate with stopping outcomes

The hypothesis that rIFG regulates stopping in conjunction with pSMA predicts that efferent rIFG connectivity should modulate with the propensity of successful inhibition. Because stopping propensity increased during TFUS, this first hypothesis involved testing interaction of SS-US and TFUS on connectivity. Model averaging and the parameters corroborated this hypothesis (Fig. 6A). Successful inhibitory trial gains were larger within rIFG and rIFG to pSMA, while larger gains were found in US trials for the rIFG to rTemp and pSMA to rPC. The interaction of these parameters with TFUS were found in all of these connections. Intrinsic changes in rIFG gain, rIFG to pSMA, and pSMA to rPC gain modulation all exhibited interactions wherein the difference between US and SS trials was larger for Stop-TFUS trials (Fig. 6B). In contrast, the rIFG top-down modulation to rTemp was larger in the No-TFUS conditions. These effects point to an increased top-down modulation of rIFG connections and cascading onto rPC. In addition, the decreasing effect from rIFG to rTemp indicates stopping propensity was associated with a reduction of top-down coupling in the ventral pathway.

As the top-down effects from rIFG to pSMA resulted in disinhibition of inhibitory interneurons, this effectively increases top-down inhibition over pSMA through rIFG. A similar interpretation holds for the modulation of pSMA to rPC. To elaborate, increases (decreases) in the self-recurrent gain of inhibitory interneurons have the effect of reducing (increasing) overall inhibition in an area (e.g., rIFG). This is because the self-recurrent gain inhibits the interneuron activity, which itself inhibits other layers.

Therefore, we further assessed the predictive validity of these connections as mediating stopping outcomes. This was tested by classifying the two stimulation conditions using a logistic regression, where the gain connections indexing SS-US trial differences were separately used to predict the TFUS condition class with a 10 k-fold cross-validation. Only the rIFG to pSMA (66% classification accuracy, $p<0.05$) and the pSMA to rPC connection (77% classification accuracy, $p<0.01$) were accurate predictors of TFUS conditional classes. The stimulation effects presumably resulted in a chain effect from the first connection modulating the change in the second. Therefore, we reasoned the pSMA to rPC prediction was partially dependent on the driving effect from rIFG to pSMA. This was tested by partialing out effects of the rIFG to pSMA connection on the latter with a linear regression, and then re-running the classification using residuals. The pSMA to rPC classification accuracy was substantially reduced but still significant (69%, $p<0.01$).

rIFG connectivity modulates processing of stopping speed (SSRT)

Our main findings indicated TFUS improved stopping by a mean change in SSRT. We predicted that if rIFG connections also modulated inhibitory processes related to the SSRT, it should predict the change in SSRT across TFUS conditions. Connections were examined using a PEB model comparing changes in rIFG connectivity between SS trials with the SSRT difference as a covariate. The SSRT difference was rescaled to 10 ms units so that connectivity was expressed as a percent

change for these units. Using non-TFUS as a baseline, we found that the intrinsic rIFG

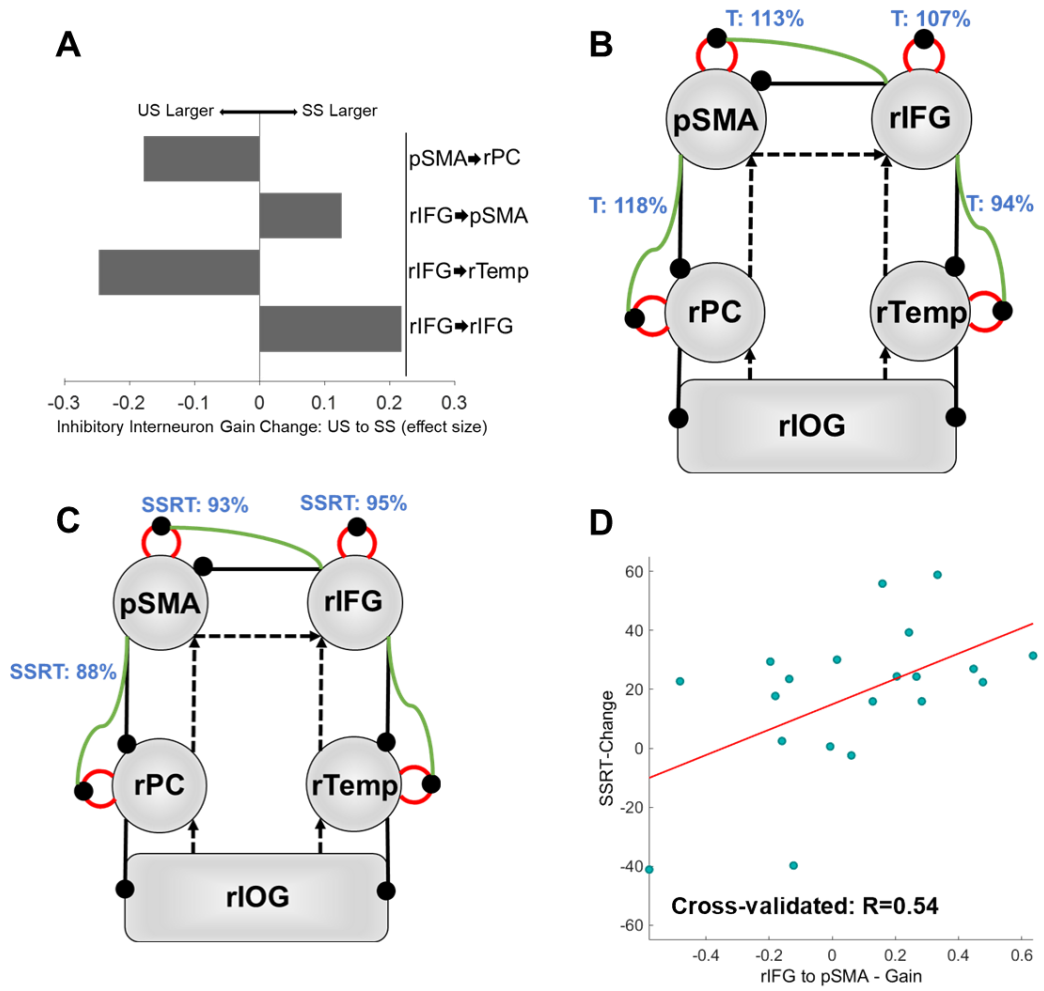

Figure 6. Results of PEB analysis to DCM to distinguish overall success of stopping versus modulation with SSRT. **Top-row**: Results of PEB models for (**A**) mean gain changes between SS and US trials in rIFG and their (**B**) interactions with TFUS. **Bottom-row:** (**C**) PEB result showing parametric scaling of difference in SSRT between TFUS conditions with gain change between TFUS conditions for SS trials. (**D**) Leave-one out cross-validated linear regression using rIFG to pSMA gain between (across TFUS conditions) SS trials as a predictor for the SSRT change.

gain, the rIFG to pSMA and pSMA to rPC coupling change scaled with the difference in SSRT (Fig. 6C). Direction of this effect specifically reflected a decreasing gain from non-TFUS to stop-TFUS, or a larger top-down inhibition scaling with shorter SSRTs. The

core issue was reversing the problem and addressing whether any of these connection changes predict the change in SSRT. Put differently, if a connectivity function is fundamentally related to implementing the cognitive inhibitory process underlying SSRT measures, connectivity changes predicting SSRT changes across a subset of subjects should be generalizable. This predictive strength for this SSRT effect was assessed using a leave-one out cross-validated robust linear regression. rIFG to pSMA top-down gain modulation was the only significant predictor of this effective change in SSRT (R=0.54, p = 0.013; Fig. 5D).

our scalp analysis isolated the P300 onset timing as a key marker of inhibitory outcomes and SSRT. A key question is where do the P300 effects emerge from in terms of the local rIFG gain and its top-down inhibitory modulation over pre-SMA. The connection between scalp and source activity are visible in estimated deep pyramidal source activity for both areas (Figure 7). We focus on deep layer activity because its strong correspondence with scalp ERPs. Timing differences in rIFG activity peaks reflected differences in stopping outcomes and SSRT. These behavioral effects were linked to a larger amplitude and earlier onset in rIFG (Figure 7 top), respectively. As Figure 6 shows, the success and its effective increase during TFUS conditions associated with larger rIFG interneuron self-gain or increased disinhibition. The result is a larger deep pyramidal activity and top-down inhibitory modulation over pre-SMA. Figure 7 (bottom) also shows the N2/P3 ERP complex was localized to pre-SMA. These source results agree with previous localizations of N2/P3 in pre-SMA (Huster et al., 2013a). Comparing activity across areas indicated timing effects corresponding to SSRT were present in rIFG but not pre-SMA, while pre-SMA exhibited a strong

amplitude interaction across TFUS conditions. Together, these observations indicate rIFG activity reflected both inhibitory outcomes and timing. The rIFG activity also points to P300 timing observed at the scalp as resulting from direct rIFG influence rather than pre-SMA.

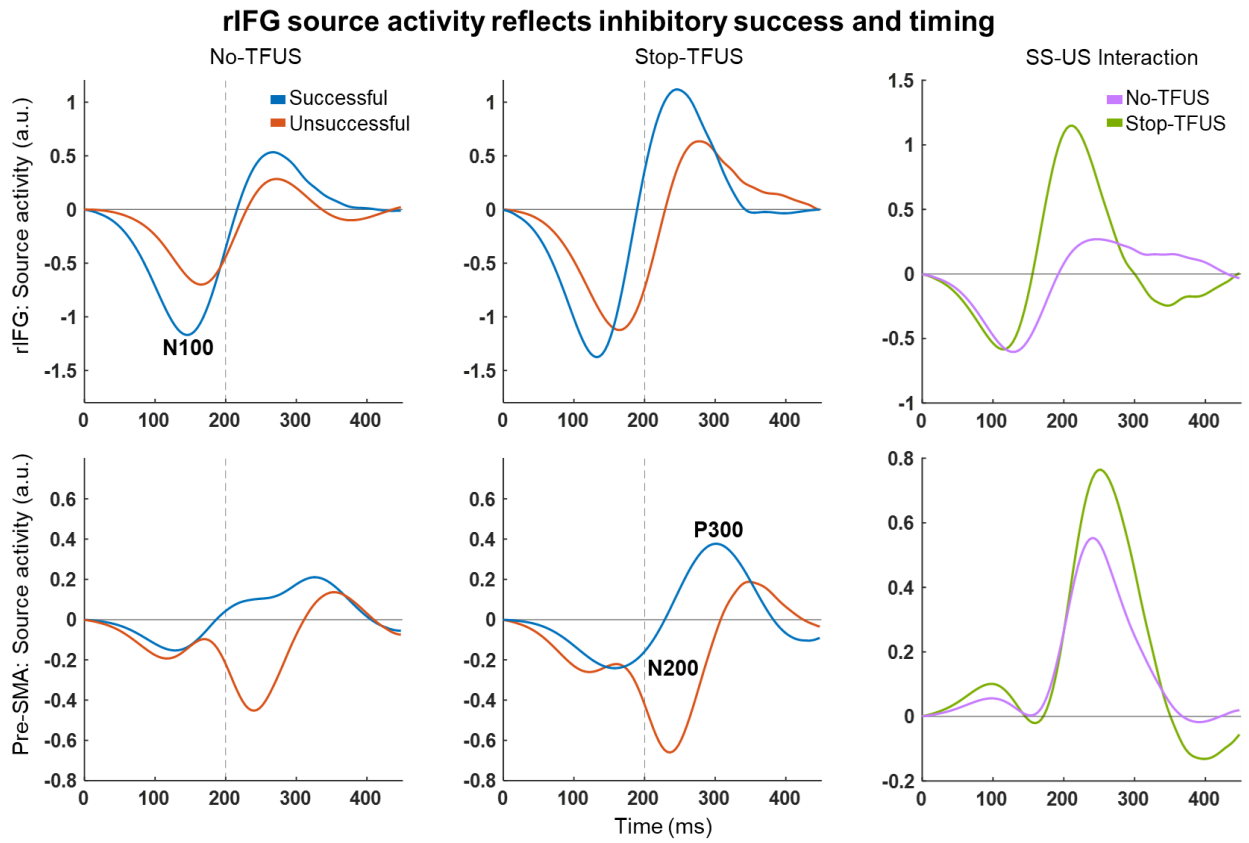

Figure 7. The top and bottom row plots show the estimated deep pyramidal source activity from rIFG and pSMA and how it modulated with these effective changes in coupling across SS trials, its change during TFUS modulation and tracking of median SSRT difference across subjects through onset of ERPs.

Discussion

We provide causal evidence that rIFG directly drives response inhibition in the cortical network including pre-SMA through an inhibitory mechanism linked to ERPs.

We demonstrated this relation by applying online focused ultrasound (TFUS) to human pars opercularis portion of rIFG during a stop-signal task. TFUS improved inhibitory performance with a faster stopping speed (SSRT), without impacting Go responses. Our first neural inquiry aimed to determine if an evoked response represents inhibitory components and could be isolated with TFUS. Inspired by correlative work (Wessel and Aron, 2015), EEG analysis revealed the P300 ERP onset timing tracked the SSRT and inhibitory success across TFUS conditions, supporting a causal connection to behavioral inhibition. Our second goal was determining if the rIFG and pre-SMA network interactions exhibit the top-down inhibitory gain effects predicted by neural networks models, while offering a causal explanation for P300 timing effects. Our approach of using Dynamic Causal Modeling (DCM) to test specific mechanisms driving evoked responses significantly extends previous findings. The modeling pointed to response inhibition being associated with a top-down rIFG modulation of inhibitory interneuron gain within pre-SMA. To our knowledge, these finding provide the first causal link of rIFG to behavioral inhibition outcomes and potency, while simultaneously isolating the core neural markers and their basis in a mechanistically specific neural inhibitory interaction in the rIFG and pre-SMA stopping network.

Our study builds on a robust neuroscience literature that has considered response inhibition from nearly all aspects. By employing online stimulation in parallel with EEG, we isolated P300 onset latencies as the sole predictor of outcomes and SSRT during non-TFUS and stop-TFUS trials. This aligns with recent work suggesting P300 latencies are a viable candidate because they predict outcomes and SSRTs (Huster et al., 2013b; Wessel & Aron, 2015) and latency modulation occurs before but

close to SSRT. This close temporal proximity of neural modulation and SSRT is predicted by stop-signal studies of single units in non-human primates (Hanes et al., 1998) and as well as neural network (Lo et al., 2009) and accumulator models (Boucher et al., 2007; Logan et al., 2015). However, our work is the first to directly stimulate rIFG online and show a common change in P300 onset and behavior. We also found the peak amplitudes of a posterior P1and fronto-central N2 and P3 also differentiated inhibitory outcomes as other studies (Bekker et al., 2005; Kok et al., 2004). These effects were presently limited to non-TFUS trials, suggesting they reflect non-inhibitory processes such as sensory detection (N1; Kenemans, 2015; Lijffijt et al., 2009), surprise or conflict (N2; Alexander & Brown, 2011; Enriquez-Geppert et al., 2010), or outcome-monitoring and attentionally reoriented processing that could occur after the SSRT (P3; Corbetta et al., 2008; Polich, 2007). It is worth noting the N2/P3 complex is regularly source localized to pre-SMA (or ACC; Enriquez-Geppert et al., 2010; Huster et al., 2013a), which might seem at odds with an inhibitory role of rIFG. In contrast, we argue these results indicate rIFG interactions with pre-SMA are responsible for modulation of P300 timing.

The idea that rIFG interactions with pre-SMA govern the P300 timing is consistent with our modeling of ERPs. DCM optimization indicated rIFG governs inhibitory control through a top-down inhibitory coupling to pre-SMA. Model comparison pointed to rIFG connections explicitly related to stopping through a top-down, gain modulation over inhibitory interneurons in pre-SMA. This gain modulation was dependent on descending rIFG activity (i.e., voltage). Additionally, rIFG gain modulation was the only connection predictive of inhibitory success and SSRT across TFUS

conditions. The key implication is that stopping potency was causally related to rIFG activity levels directly increasing local inhibition in pre-SMA. These conclusions diverge from previous fMRI work (Duann et al., 2009; Rae et al., 2015) that found rIFG to pre-SMA coupling predicts inhibitory success and SSRT, but also reported similarities across inhibitory and non-inhibitory tasks (Erika-Florence et al., 2014). Their interpretation was rIFG connectivity was a non-inhibitory process that signaled context-changes to pre-SMA. However, these studies only compared successful stopping against a baseline, without a pure comparison of stopping that we obtained with TFUS.

Although we did not manipulate attention, our modeling also informs debates regarding rIFG connectivity as driving inhibitory versus attentional control (Aron et al., 2016; Hampshire & Sharp, 2015). We found rIFG to temporal and fronto-parietal pre-SMA to parietal connections interacted with inhibitory outcomes and TFUS. These ventral and fronto-parietal pathways, respectively, are typically associated with attentional reorienting (Corbetta et al., 2008; Vossel et al., 2014). Modulation of these pathways aligns with previous DCM studies explicitly manipulating top-down attention and stimulus expectancy (Auksztulewicz & Friston, 2015; Fardo et al., 2017). However, because we found that neither connection predicted SSRT, we expect these putatively attentional pathway effects are secondary to inhibitory effects elaborated between from rIFG to pre-SMA.

Both inhibitory and attentional frameworks predict rIFG exerts top-down gain modulation, which we considered during model building. However, they differ in which neural population gain should be modulated by rIFG. Building upon neural models of response inhibition (Lo et al., 2009; Logan et al., 2015) and circuit mechanisms (Roux &

Buzsáki, 2015), there is agreement that direct inhibition is biologically plausible through tuning of local inhibitory interneuron gain through interareal connections. In contrast, models underlying attention predicts top-down modulation or gain control over bottom-up excitation (Desimone & Duncan, 1995; Feldman & Friston, 2010; Spratling, 2008). DCM embeds excitatory gain through superficial pyramidal cells. Our formal model comparison was consistent with a direct inhibitory control model. The winning model involved rIFG modulating interneuron gain in pre-SMA. This finding has a direct implication for explaining how rIFG could drive inhibition but the P300 onset timing in pre-SMA can reflect the behavioral timing of inhibitory responses. This rests on the effect of neural gain operating as a synaptic time-constant. Gain modulation can directly influence ERP timing of peaks and decay rates. With respect to our results, this suggest the P300 onset tracking of SSRT is directed by rIFG activity levels that modulate inhibitory interneuron gain in pre-SMA. The pre-SMA gain in turn alters the expression of P300 timing. Two caveats are worth noting here. First, our model inferences are based on non-invasive measurements and require verification from cell-specific animal recordings. Second, given rIFG's substantial morphological diversity and connectivity to other regions (Briggs et al., 2019) that is responsive to attentional and perceptual processing factors (Erika-Florence et al., 2014; Levy & Wagner, 2011), we expect our results indicate an explicit inhibitory function operates alongside other rIFG functions.

Most response inhibition analyses of non-invasive measures (EEG or MEG) have focused on sensor level analysis, some considering source activity (Enriquez-Geppert et al., 2010), or invasive electrophysiology with limited cortical coverage (Swann et al., 2012; Wessel et al., 2013). In contrast, a core focus of inhibition studies has been

pinning down exact locations of activated regions and thus employed fMRI. Neuroimaging poses the issue of filtering fast neural responses, with the connectivity methods being agnostic to mechanistic instantiations of top-down and bottom-up interactions. We believe our utilization of ERPs, online TFUS, and DCM for evoked responses is most appropriate for differentiating bottom-up, top-down and gain effects driving evoked responses. This stems from the fact that different theories of inhibition or attention predict different gain mechanisms, and these are instantiated in the microcircuit models within DCM (Bastos et al., 2012). While a clear limitation is indeterminacy with source localization (Grech et al., 2008), high temporal resolution is required to differentiate signals preceding stopping-related events (e.g., SSRT). These methodological concerns are tempered by pairing model-driven hypotheses with TFUS, because connectivity effects hold within the model across conditions despite source ambiguity. Finally, utilizing individual and accurate MRI-based head models increases accuracy of electromagnetic source models (Baillet et al., 2001), and issues with relying on source localization to identify network nodes was circumvented by a priori selecting regions from previous neuroimaging.

In conclusion, competing theories have argued whether response inhibition is a defining process of a cortical network centered on rIFG. Response inhibition likely involves several processes, ranging from sensory cue detection, attention, performance monitoring, and presumably explicit motor inhibition (Munakata et al., 2011; Wessel & Aron, 2017; Wiecki & Frank, 2013). Our behavioral, ERP, and connectivity results favor the framework wherein rIFG is directly involved in explicit motor inhibition coinciding with attentional or sensory processing. The present application of high-resolution TFUS

and ERP modeling provides a neural and methodological framework for testing other aspects of inhibitory and decision-making function. A key question is whether these gain mechanisms are the basis for global versus selective response inhibition. The efficacy of TFUS reported here also indicates a possible neural intervention to pair with treatment for disorders exhibiting inhibitory dysfunction, e.g., impulsivity or ADHD (Bari & Robbins, 2013).