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

Justin M. Fine, Maria Fini, Archana S. Mysore, William “Jamie” Tyler, Marco Santello

Address for correspondence:

Justin M. Fine

Psychological and Brain Sciences

1101 E 10th St
Bloomington, IN 47405

Keywords: Transcranial Focused Ultrasound; Response inhibition; Dynamic Causal Modeling; Inferior frontal gyrus; Cognitive control

Abstract

Behavioral inhibition is necessary to suppress prepared actions that become contextually inappropriate. Pars opercularis in right inferior frontal gyrus (rIFG) is consistently implicated in triggering inhibition processes. Competing neural of response inhibition and the role rIFG include either an explicit stopping mechanism invoked by pars opercularis and subcortical connections, or and rIFG-based attentional orienting process that drives processing away from original action signals. Combining a Stop signal task with transcranial focused ultrasound (tFUS) to pars opercularis revealed an improved inhibition performance, driven only by changes in action stopping-related processes. Source-resolved electroencephalographic dynamic causal modeling of a stopping network revealed successful inhibition during baseline (no tFUS) and tFUS was associated with modulated effective connectivity from pars opercularis to subcortical nodes. In contrast, modulation of pathways related to attentional processing were specific to changes in baseline successful and failed inhibiton. These findings provide causal evidence that response inhibition is a specific function of pars opercularis and its subcortical projections in healthy humans.

Introduction

Behavioral inhibition is necessary to suppress impending actions that become contextually inappropriate (Aron, Robbins, and Poldrack, 2004; Baddeley, 1996; Logan, 1985). The control over inhibitory capacities is dramatically reduced in pathologies dominated by impulse control issues (e.g., ADHD; Bari and Robbins, 2013; Dalley, Everitt and Robbins, 2011). A paradigm for probing inhibition is the stop-signal task (Logan and Cowan, 1984). This task involves cueing individuals to execute an action (Go signal) on every trial. On a percentage of trials, they are cued (Stop signal) to attempt inhibiting the response after some delay following a Go signal. This task allows deriving the stop signal reaction time (SSRT), a latent measure of stopping speed.

A predominant framework implicates a right-lateralized prefrontal stopping circuit driving inhibition (Aron, Robbins, and Poldrack, 2014; Chambers et al., 2006), with a locus in a posterior right inferior frontal gyrus (rIFG), *pars opercularis* (Aron and Poldrack, 2006). This area is assumed to either directly implement motor braking (Aron et al., 2014; 2016) or indirectly via projections to subcortical nodes. This network also includes pre-supplementary motor area (pre-SMA; Chambers et al., 2009), subthalamic nucleus (STN) and striatum (Aron et al., 2007; Mallet et al., 2016). These control allocations have been supported by increased rIFG and pre-SMA neural activity during successful compared to failed stopping (Aron et al., 2007; Boehler et al., 2010; Cai and Leung, 2011; Li et al., 2006). Assuming the potency of inhibitory processes relates to between-subjects SSRT, rIFG BOLD and STN activity consistently correlates negatively with SSRT (Aron and Poldrack, 2006; Li et al., 2006). Additionally, inhibition and SSRTs are altered in those with rIFG lesions (Aron et al., 2003) and individuals with ADHD (Clark et al., 2007; Morein-Zamir et al., 2014).

Several researchers have proposed that rIFG engages only indirectly with inhibition (Duann et al., 2009; Sharp et al., 2011; Xu et al., 2017). These studies have provided evidence

that rIFG is involved in stop-signal detection, and then signaling pre-SMA to explicitly trigger inhibition (Duann et al., 2009; Sharp et al., 2011). An implication is that rIFG to basal ganglia effects on inhibition are indirect and routed through pre-SMA (Duann et al., 2009; Rae et al., 2015; Watanabe et al., 2015). At a network level, this hypothesis has been partially supported by fMRI stop-signal connectivity studies wherein rIFG bidirectionally interacts with pre-SMA, but only pre-SMA exhibited modulated connectivity with subcortical structures during successful inhibition (Duann et al., 2009; Rae et al., 2015). However, other connectivity studies have implicated both rIFG and pre-SMA connectivity to STN and striatal pathways as predicting inhibition speed (Jahfari et al., 2011; Xu et al., 2017) in support of a dual pathway model of inhibition (Aron et al., 2016). Due to these mixed results, the roles of rIFG and pre-SMA in inhibition, as well as their effective interactions with subcortical nodes, remains an open question.

rIFG's motor inhibition role has also been questioned by findings indicating rIFG is activated by a host of other putatively non-inhibitory tasks (Erika-Florence et al., 2014; Hampshire, 2015; Sharp et al., 2010). rIFG activity scales with stimulus probability (Shulman et al., 2009) and regularly during tasks requiring attentional re-orienting (Vossel et al., 2006; Levy and Wagner, 2011). This has led to an alternative hypothesis wherein there is no inhibitory module in rIFG (Hampshire and Sharp, 2015; Chatham et al., 2012), with inhibition emerging from attentional orienting and local biased competition. These claims are supported primarily by demonstrating rIFG BOLD is equivalently active during stop signal tasks with no inhibitory demands (Sharp et al., 2010). In this framework (reviewed in Hampshire and Sharp, 2015), inhibition could occur through rIFG top-down signals that bias attentional processing by increasing synaptic of sensory receptive cortices (Hampsire, 2015; Feldman and Friston, 2010).

Interpreting these alternative explanations warrants caution. For example, controlling for attentional demands by comparing stop-signal and other tasks may induce unaccounted for

cognitive processes or latent inhibitory demands (Aron et al., 2014). Clearly aligning a pars opercularis sector of rIFG as causally involved in motor braking, while detailing its role in the broader inhibition network, requires exogenously altering neural activity online. To achieve this, several studies have applied transcranial magnetic (TMS; Cai et al., 2012; Obeso et al., 2013; Verbruggen et al., 2010) or current (Jacobson et al., 2011) stimulation online or offline during inhibition tasks. Some studies found offline TMS to rIFG and pre-SMA impairs or improves inhibition performance (Chambers et al., 2007; Verbruggen et al., 2010). However, others have shown pre-SMA TMS can improve inhibition with no rIFG effect (Obeso, 2013). Limitations of previous human neurostimulation response inhibition studies include either a lack of neural measurements or limited spatial accuracy. For example, TMS applied to rIFG during inhibition tasks have most likely engaged ventral premotor areas involved also in action switching (Buch et al., 2010). The implication is that, if rIFG is potentially involved in attentional orienting, inhibitory control, or both, previous rIFG neurostimulation likely introduced broad effects on these processes. Therefore, it remains to be causally established that rIFG engages an explicit motor inhibition mechanism embedded in its connectivity that operates alongside attentional mechanisms (Munkata et al., 2011; Wiecki and Frank, 2013).

Here, we employed transcranial focused ultrasound stimulation (tFUS) applied to pars opercularis while humans performed a stop-signal task. Neural source activity was assessed by combining tFUS with EEG event-related potentials (ERPs) and source analysis. Imperatively, we found tFUS to pars opercularis had a direct effect of improving response inhibition by altering stopping related processes. To determine how tFUS causally altered the inhibition network, we employed dynamic causal modeling. This combined tFUS-EEG-DCM approach allowed comparison of network changes in non-stimulated failed- and successful inhibition trials, as well as to tFUS-altered successful inhibition. The latter contrast is central to providing direct insights into neural activity in which inhibition performance may change due to tFUS, but

outcomes were categorically the same, i.e., successful stopping. Our main network hypothesis was that an explicit rIFG inhibition mechanism would be embodied in direct rIFG to subcortical connectivity, and the connection weighting would reflect tFUS induced changes in stopping efficiency (SSRT). We confirmed evidence for a pars opercularis role in motor inhibition by demonstrating DCM connectivity of rIFG with a hidden subcortical layer was directly altered during tFUS. Our DCM also indicated only No-tFUS successful versus failed stopping, rather than successful stopping in general, was differentiated by attentional mechanisms linked to changes in the recurrent synaptic gain of sensory cortices (i.e., visual cortex). Importantly, these results support the proposal that rIFG is directly involved in both an explicit response inhibition function and attentional orienting.

Methods

Participants. The experiment consisted of three experimental groups. The main experimental group was the right inferior frontal gyrus (rIFG) transcranial focused ultrasound (tFUS) stimulation group (n = 25; 19 males, mean age 24.1 yrs SD 3.2 yrs). This group received tFUS stimulation. A second group was used as cortical site control group, whom received stimulation over ipsilateral somatosensory cortex (S1) (n = 23; 15 males, mean age 22.4 yrs. SD = 3.3 yrs). A third group received a sham stimulation (n = 15; 8 male, mean age 24.2 yrs SD = 2.8 yrs) and was used as control for possible auditory effects of tFUS over rIFG (sham rIFG). All individuals were right-handed (self-reported) volunteers and received financial compensation for participation in the study. Before being enrolled in the study, each subject was screened for neurological disorders and previous history of epilepsy, stroke, or brain injury. Furthermore, a neurologist from the Barrow Neurological Institute (Phoenix, AZ) screened all subjects' T1 scans after structural MRI acquisition before participation in the study.

Stop Signal Task and tFUS design. The current study used the conventional Stop Signal Task (SST) that involved both 'Go' and 'Stop' trials (Verbruggen and Logan, 2008; Fig.

1A). We performed the experiment using PsychoPy software (Pierce, 2007). Each trial started with a centrally-located fixation cross on the monitor. In both trial types, the fixation cue was replaced by a green ‘Go’ circle (3° x 3° visual angle), with an exponentially-distributed time interval (mean: 500 ms; standard deviation: 50 ms). Subjects were instructed “to press the up key as soon as they detected the Go circle” (top panel, Fig. 1A). In ‘Go’ trials, the circle vanished when the button was pressed or after 800 ms had passed from the fixation cross stimulus. In ‘Stop’ trials, the Stop cue consisted of the Go circle being replaced by a red square (bottom panel, Fig. 1A). If the subject successfully inhibited his/her 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). The 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 designated by having each subject first perform a practice block of 50 Go trials only. After this block, the 4 SSD levels were set to 25, 35, 75 and 95% of the mean Go reaction time (RT) distribution computed across the 50 Go trials. These SSDs were fixed throughout the experimental session. This choice was determined by two factors. First, 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). Second, we sought to determine the effects of online tFUS at different SSDs to estimate the effects of stimulation on neural and behavioral responses at different stages of a Go process predicted by response inhibition models (Verbruggen and Logan, 2008; Schall, 2017). All trials were separated by an inter-trial interval of 2000 ms (±300 ms randomly drawn jitter).

We delivered tFUS either simultaneously with (1) the Go signal in both Go and Stop trials, or (2) the Stop signal (Fig. 1A). The purpose of delivering tFUS during Go trials was to determine the neural and behavioral effects of tFUS to rIFG independent of a stopping signal.

Specifically, this allowed us to assess whether any effects of tFUS on stopping behavior are related merely to alteration of the timing of an underlying Go process. Therefore, we used 5 types of trials. The first two consisted of Go trials with No-tFUS or with tFUS locked to the Go cue. The other three trials consisted of Stop trials: trials with No-tFUS, with tFUS locked to the Go cue, and with tFUS locked to the Stop signal. These three types of Stop trials were examined across the four SSDs.

After a practice block (n = 50 trials) involving simple Go RTs, subjects performed 12 blocks of trials that were randomly intermixed with Go and Stop trials with and without tFUS. The practice block was used to determine a distribution of RTs for setting the levels of the SSD in the main experiment. The overall probability of a stop trial was set to 35% of all trials (Fig. 1B). We chose this level to accommodate the need for large amounts of trials required to examine the effects of tFUS on Go and Stop trials across all SSD levels. Each experimental session consisted of 1200 trials. We delivered tFUS on 50% of the Go and Stop trials (Fig. 1B). tFUS delivery for Stop trials was evenly distributed across the 4 SSD levels. These trials were split into tFUS Stop trials wherein stimulation was initiated at Go versus Stop (see Figure 1A-B). The trial numbers were chosen to enable the comparison between tFUS and non-stimulation trials across successful and failed inhibition trials, while allowing a reasonable number of trials to be performed without inducing significant fatigue to the participants. tFUS trials were randomly and evenly distributed throughout the experiment. Subjects were given 2 minute-breaks every 300 trials.

EEG acquisition. EEG was recorded using a 64-channel ActiCap system (BrainVision, Morrisville, NC), with a standard 10–20 layout. Data was recorded at a sampling rate of 5 kHz, with resolution 0.1 μV and bandpass filter of 0.1–100 Hz. Impedances were always 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 a

CapTrak camera system (BrainVision, Morrisville, NC) with 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 Data Processing).

Structural MRI acquisition (T1) and processing. For purposes of tFUS neuronavigation and co-registering EEG electrode placement, 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 surface volume labeling procedure with the Destrieux atlas (Destrieux et al., 2010). 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 was used for identifying tFUS targets.

tFUS targeting, setup and parameters. The ultrasound transducer was coupled to the scalp by first coating the surface with acoustic coupling gel, and then placing it in a fixed mold with an overlaid acoustic gel mold contained in a 3D printed housing. Acoustic coupling gel was then placed on the scalp, and the ultrasound housing affixed to this point of the head with an elastic suspension system that did not interfere with EEG electrode placement. This procedure was used for active tFUS in the main rIFG and S1 groups. Figure 1C shows the rendering from one of the individual's T1 and scalp in the rIFG group, along with the 3D rendering of the transducer housing (green object) and the pars opercularis mask (pink anatomical structure). In the auditory rIFG control group, we employed a sham tFUS (similar to Legon et al., 2018) by placing the gel coated transducer perpendicular to the rIFG target. This sham procedure was done to ensure there was still an auditory effect of the ultrasound transducer without active stimulation. In the current work, we used a Brainsight neuronavigation system (Rogue industries) to guide placement of the transducer beam profile with respect to each individual's

brain anatomy. This was done by importing each subject's preprocessed T1 and mask into Brainsight and registering the US transducer housing into individual's anatomical space. The alignment and cortical registration were accomplished by registering the individual's T1 derived anatomy using the left and right periauricular, and nasion. To visualize the US targeting into the cortex, we created a custom design in Solidworks that rendered the transducer housing and ellipsoidal beam profile (focal length of 30 mm) projection into the registered cortex (Fig. 1C). During targeting (pre-experiment), the experimenter adjusted the alignment of the transducer on the scalp until the beam was projected into the centroid of the desired target. For the rIFG group, we delivered tFUS to the center of the pars opercularis mask (Fig. 1C). tFUS for the S1 cortical site control group was delivered to the centroid of the ipsilateral post-central gyrus mask. In the sham rIFG group, we also used a pars opercularis mask to ensure comparable US transducer placement between this group and the rIFG group. The tFUS transducer and housing were fixed in place using a custom-made netting cap and tension cord system that went over the EEG cap without disrupting the position of the electrodes. To ensure accurate tFUS placement throughout the experimental session, the rotational and cartesian displacement of the beam profile from the cortical target was tracked. The overall accuracy was measured as deviation from the original alignment of the beam with the anatomical target. During the experimental session, we sampled the position of the tFUS transducer during each break. Accuracy was very high, with an average deviation of ±1.5 mm displacement across all subjects and sessions.

The setup and parameters used for tFUS in this experiment were nearly identical to those used by Legon et al. (2014). Briefly, we used a single-element focused US transducer with a center frequency of 0.5 MHz. This center frequency was chosen because optimal frequencies for tFUS transmission while minimizing cranial attenuation are frequencies less than 0.65 MHz (Hayner and Hynynen, 2001; White et al., 2006). Previous work has also indicated

that 0.5 MHz can alter both neural and behavioral responses. The transducer was 30 mm in diameter with a 30-mm focal length. Legon et al. (2014) previously performed an acoustic water tank profiling of this transducer to examine the acoustic pressure fields and transmission properties through a piece of human cranium. In terms of spatial resolution of the acoustic intensity, their testing indicated that measurements at ~30 mm yielded a lateral beam resolution with a full-width half maximum (FWHM) of 4.56 mm. These tank measurements also revealed that transcranial transmission yielded an approximate spatial-peak pulse average intensity (I_{sppa}) of 5.8 W/cm^2 at the transducer focal length (~30 mm) when transmitted through cranium. While Legon et al. (2014) showed transcranial attenuation of intensity compared to free space (I_{sppa} ~= 23 W/cm^2), their work and others have shown that the above transducer, center frequency, and pulse parameters described below are capable of eliciting both neural (e.g., EEG) and behavioral changes across a range of neural structures and tasks.

The tFUS waveforms were generated using a two-channel, 2 MHz function generator (BK Precision). Briefly, channel 1 was programmed to deliver ultrasound at a pulse repetition frequency (PRF) of 1.0 kHz while channel 2 drove the ultrasounds transducer at a 0.5 MHz acoustic frequency in a bursting mode. Channel 1's receival of a trigger from the PC was used as an external trigger for channel 2. The waveform pulse duration was set to 0.24 ms (24% duty cycle) and a total stimulation duration of 0.5 s. The pulse duration was set on channel 2 by changing the cycles per pulse to 120. The total stimulation duration was controlled by setting channel 1's number of pulses to 500. The transducer power was driven by sending channel 2's output to a 40-W linear RF amplifier (E&I 240L; Electronics and Innovation). The amplifier's output was sent to the above-described focused US transducer (Blatek, Inc., State College, PA). The waveforms were triggered in alignment with experimentally-relevant temporal events (see description below and Fig. 1A). It has been previously verified that the resulting waveform does not incur any 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 (Mueller et al., 2017; Treeby and Cox, 2010). Reference peak pressure planes for the simulations were derived from a water tank test and 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 the water tank test. The max pressure plane at the 30 mm focus in the water tank 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, and a Courant–Friedrichs–Lewy = 0.1, and simulation time was set to 6 pulses (duration = 250 μs) to ensure simulation stability.

Water tank data indicated a max pressure of 1.05 Mpa and spatial peak pulse average intensity (Isppa) of 22.4 W/cm^2 at the focus. This output aligns with previous studies and simulations (Legon et al., 2014) of the same transducer. Comparison of simulations and water 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. The lateral full-width at half maximum (FWHM) of the max pressure at the beam was 4.39 mm in simulation (Fig. 1C).

For simulating transcranial US, we extracted 3-dimensional maps of the skull from a CT (1 mm resolution) and brain from T1 MRI scan (1 mm resolution) from 3 preoperative patients at Barrow Neurological institute. The MRI and CT were both co-registered and normalized to the MNI space in SPM12. To mimic our approach of targeting in the actual experiments, we surface registered the gray matter volume to the AAL atlas (Mazoyer et al., 2001) and identified the centroid of *pars opercularis*. This allowed us to map from world coordinates of the scan to MNI

coordinates of the target. These 3 subjects had an average stimulation location of x=48, y=18, z=6. Conversion between Hounsfield units in the CT to sound speed and density were done using the relations 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). Results for the rIFG targeted through skull simulation are presented in the Results section.

Data processing and experimental variables

Behavioral variables. The main variables under consideration were the 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 allows estimation of the distribution of SSRTs while assuming an ex-gaussian distribution of SSRTs. Importantly, 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 and S1) 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 stimulation because stimulation did not alter the Go RT.

EEG artifact removal. Continuous EEG data were first down-sampled to 250 Hz, then high-pass filtered (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 respective groups and were always interpolated after artifact removal. The remaining data processing involved creating epochs from Stop trials locked to stop signal onset (-100 to 500 ms peristimulus) because our analysis focused on this epoch. Individual epochs were then rejected from further analysis if they contained large scalp EMG or rare events (<8% of all trials). The ERPs baseline corrected by subtracting the activity from -100 ms to the stop signal. Out of the 53 participants, 3 subjects were excluded from analyses due to EEG recording issues (impedance >25 kΩ across channels). The remaining data were bandpass filtered from 1-25 Hz. EOG artifacts related to eye movements and blinks were removed using Independent Components Analysis using eeglab (ICA; Delorme and Makeig, 2004). On average, 3.35 components were removed per participant. Because we later applied a Hanning taper to the edges of the event-related potentials (ERP) before dynamic causal modeling, we applied the same procedure to the ERPs after cleaning using the standard Hanning taper in the FieldTrip toolbox.

Statistical analysis

Behavior analysis. To quantify how the probability of response inhibition changed over levels of SSD, P(respond|signal), and across tfUS conditions within and across groups we fit a 2-parameter logistic model. This was done to analyze P(respond|signal) as a curve. This was achieved using by fitting the 2-parameter linear 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).

Separating response and response inhibition from neural data. Our analysis of ERPs was based on the premises of the Independent Horse Race model (Logan and Cowan, 1984) – which posits independent accumulation of Go and Stop activity till one of them reaches threshold. Due to the nature of measuring these processes through EEG and the inhibitory processes, there is overlap of neural processes related to stopping and going during stop trials. Therefore, we removed Go-related activity from both successfully (SS) and unsuccessfully (US) inhibited Stop trials through subtracting the Go ERPs from the Stop ERPs. This approach has been used in previous EEG/ECoG (Mattia et al., 2012) and Dynamic Causal Modeling-EEG studies (Kibleur et al., 2016). We used 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 that were 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 US trials, we found latency matched Go trials with RTs with a different procedure. We first calculated 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.

Analysis of inhibition-related ERPs. Our analysis focused on event-related potentials (ERPs) from source-localized analysis and Dynamic Causal Modeling (DCM; David et al., 2006). A primary drive for these analyses is that previous work has revealed a set of potentials that often accompany response inhibition following a Stop signal (reviewed in Huster et al., 2013

and Kenemans, 2015). The main ERPs found in inhibitory tasks include a N2/P3 complex that has a fronto-central and radial topography. This component has been hypothesized to be generated mainly by the pre-SMA/SMA in the medial frontal cortex (Kenemans and Kähkönen, 2011) and has been considered to reflect a critical signature of reactive stopping elicited by stop signal tasks (Kenemans, 2015; Wessel and Aron, 2017). Furthermore, the P300 of this complex has been proposed to be a relevant marker of stopping efficiency, given that it predicts the SSRT (Wessel and Aron, 2016) and successful versus failed inhibition (Kok et al., 2003). The rIFG is also considered a critical node for stopping based primarily on lesion (Aron and Fletcher, 2003) and offline non-invasive stimulation studies (Aron, 2011). The ERP associated with stopping in this area is typically associated with a negative amplitude difference comparing successful and failed stopping that emerges around 200 ms (N200; Schmajuk et al., 2006). We examined these ERPs and P/N100 responses which are sometimes elicited over sensory areas, depending on whether the stimuli used for Go and Stop signals are in the same sensory modality (Kenemans, 2015). By combining tFUS and EEG, we can examine some of the issues that are addressed by ongoing debate, i.e., which of these potentials are stopping-relevant, and how their activity is generated by a network model through DCM – all without signal interference induced by stimulation.

Sensor-level analysis. To examine the standard ERP effects typically found in the SSTs, we first examined activity at the sensor level. This was done using permutation based dependent samples t-tests. Spatiotemporal activity was examined and multiple-comparison corrected for using a cluster-based p-value correction of $p < 0.01$ and 5000 permutations for each contrast considered. The contrasts included comparison of (1) successful – unsuccessful trials over both tFUS conditions, a (2) successful (No-tFUS) – successful (Stop-tFUS) contrast, and (3) and interaction contrast comparing SS – US difference between the No-tFUS and Stop-tFUS conditions. The first contrast is typically used to determine which areas exhibit ERPs (or

brain areas) that differentiate successful inhibition (Swaan et al., 2012). The interaction was used to determine how the SS-US contrast differed between No-tFUS and Stop-tFUS conditions. The second contrast was the main focus, which was used to determine which scalp ERPs differentiated successful stopping in the No-tFUS and Stop-tFUS conditions. We anticipated that this latter contrast would reflect processes that mainly include those responsible for both the success and efficiency of inhibitory processes, e.g., SSRT.

P300 onset and SSRT tFUS effects. Because recent work has indicated that the frontocentral P300 onset latency is related to the SSRT (Wessel and Aron, 2015), we hypothesized that tFUS altered inhibition through the SSRT. Therefore, we would see a shift in this latency as well. We calculated the shift in P300 onset crossings between 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.

Source localization. To estimate the activity in source space from the sensor recordings, we used a group inversion with the multiple sparse-priors approach as implemented in SPM12. The individual subject data passed to the inversion routine were mean sensor ERPs per condition. We performed the group inversion only for the rIFG groups and analysis because this was the only group to exhibit behavioral effects from tFUS. In this procedure, the individual's recorded electrode locations and individual T1 were warped to the default MNI anatomical brain and cortical mesh provided in SPM. These meshes were used to calculate the forward solution using a boundary element head model. All conditions were used in the group inversion routine. Because the multiple sparse-priors approach attempts to fit the sensor data with respect to the lead-field matrix, we performed the inversion over a window starting from the stop signal up to

500 ms. While narrow windows are considered better for time-resolved estimation, we wanted to estimate overall changes in activity using the same window we would employ later for dynamic causal modeling.

Source analysis: whole-brain contrasts and regression. Based on examination of the sensor level data, we first identified time windows surrounding the ERPs discussed above (N100, N200, and P300). To balance the number of points contributing to the source estimate used for analysis, we used the same window size for estimating the source activity of each ERP. The time windows were centered around the across-subject mean peak activity of each ERP with a window of ±40 ms. These time windows were used to create 3D source image activity interpolated into MNI voxel space for each subject and condition. The resulting images were spatially smoothed (6 mm full-width half maximum) evoked power from 1-25 Hz. We used evoked power because we were concerned with 'activation' and time window, but not the direction of voltage deflection. This choice was driven by the fact that we already established the canonical inhibition related ERPs at the scalp level. These evoked images were analyzed used a flexible factorial design to implement a repeated-measures ANOVA, including all main effects and interactions. The factors included (1) inhibition success (SS or US trial), and (2) stimulation condition (No-tFUS or Stop tFUS). The resulting statistical parametric maps were analyzed with a threshold set at $p < 0.005$ (peak-level, uncorrected) and cluster-wise FWE $p < 0.05$. For expositional brevity, these SPMs are presented in the supplementary material except for the conjunction F-contrast showing the overlap of the SS-US and tFUS F-contrasts. This conjunction shown in the main text both confirms the differential SS-US effect in rIFG and the spatially precise impact of tFUS.

We also conducted a whole-brain SPM linear regression. This was done using the same time-windows identified above. We regressed the difference in evoked activity between no-tFUS and Stop-tFUS SS trials against the difference in each subject's SSRT for these conditions.

Using this approach, we examined both positive and negative contrasts (t-tests) in each window. This analysis was done to both (1) determine prior source locations for the DCM analysis (see below) and (2) identify which areas predicted the change in SSRT due to tFUS. Because a primary goal was determining DCM priors, we used a lenient uncorrected cluster threshold of $P < 0.01$ and a minimum cluster-extant threshold of 20 voxels.

Dynamic Causal Modeling (DCM). To model the connectivity of the network involved in successful stopping, we used the canonical microcircuit model (Bastos et al., 2012) for all areas except the contralateral motor cortex. For left motor cortex (M1), we used a recently developed version built-off the canonical microcircuit that more closely resembles the agranular structure of M1 (Bhatt et al., 2016). Our sequential model building and analysis focused both on a priori areas of interest that have been well described in response inhibition literature (e.g., Wessel and Aron, 2017) – including rIFG, rpre-SMA, rDLPFC – and those areas that we identified in comparing the effects of Stop tFUS and No-tFUS changes in mean SSRT on evoked power. Given the typically right lateralized areas found to be modulated by inhibition, we focused mainly on pathways on the right hemisphere. We used source locations identified in both (1) the regression of change in SSRT and (2) the whole-brain SPM interaction F-contrasts examining locations for which Stop-tFUS and no-tFUS were different for the SS-US comparison. The significant source cluster peaks ($P < 0.01$ threshold) revealed by these analysis were labeled using the AAL atlas labeling toolbox, and used to identify source coordinates in MNI space ([x,y,x]: rIFG [48,28,4]; pre-SMA [6, 24, 54]; rDLPFC [30, 28, 40]; rParietal [10,-74,56]; rTemporal [52,-18,-12]; LM1 [-37,-25,-62]; right inferior occipital gyrus (rIOG) [46, -76,10]). These source locations are in general agreement with previous literature in both fMRI, EEG, and MEG studies of SSTs (Aron et al., 2007; Rae et al., 2014) and meta-analyses of cortical locations and boundaries (Mayka et al., 2006). In building the model, we also included a hidden deep source to model the potential connectivity effects to and from cortical sources. Because

the main output of the basal ganglia mediating inhibition and responding is the thalamus, we used a source location [4, -16,-2] identified during a previous fMRI stop signal study (Boehler et al., 2010). This approach of using a deep source node in DCM for EEG has been previously employed by David et al. (2011) as well as Kiblur et al. (2016), wherein the latter case they also used a stop signal task. In the DCM, we used an equivalent current dipole model and allowed the inversion process to optimize the source locations.

To determine the appropriate model connectivity structure, we first examined the grand mean ERP data in SS No-tFUS trials from a window spanning 0 – 500 ms. Recent work has shown that estimating the model structure from grand means (across subjects) is sufficient and provides a close approximation to fixed-effects selection when using ERP data (Litvak et al., 2015) rather than fitting DCMs across all subjects for each variation in connection structures.

To optimize the model structure, we performed Bayesian model selection using family-wise fixed effects in several iterations. In the first set of iterations, we inverted 48 different models from the grand mean data, and then partitioned it into several families (Penny et al., 2011). These families were based on (1) the structure of the pre-frontal hierarchy (Figure 2), (2) structure of the lower hierarchy (Figure S1), (3) whether exogenous inputs were supplied to rIOG, rIFG, or both, and (4) whether the cortical nodes projecting to the hidden deep source were of the forward or backward type. Inputs were modeled as a Gaussian bump with a prior onset of 60 ms. With respect to comparison (4), pulling evidence from previous inhibitory control, primate tracing, and tractography studies and reviews, the areas we examined for different connections projecting to the deep node included rIFG, pre-SMA, right dorsolateral prefrontal cortex (rDLPFC) and rIOG. A backward connection from M1 to the deep node was also included based on putative M1 to basal ganglia connections (Nambu et al., 2002) and previous DCM modeling (van Wijk et al, 2018). Because the efferent connections from

Thalamus output is excitatory and has been found to target deep pyramidal layers (Yamawaki et al., 2014), the above listed areas all received forward connections from the deep source.

In this modeling, we assumed hierarchically connected areas always entailed a backward and forward connection between nodes. Nodes that were lateral in a hierarchy were supplied with both forward and backward nodes (Figure S1). A final note is on the connections between lower areas and prefrontal areas. Rather than optimize all possible connection permutations, we chose to instantiate connections that reflect the cognitive and attentional control-related differences typically postulated to operate in dorsal and ventral pathways (Corbetta and Shulman, 2001). Both streams received forward input from rIOG. The ventral stream included a forward connection from rTemporal cortex to rIFG. The dorsal stream included forward projections from rParietal to both rDLPFC and pre-SMA.

After inversion of each model, we used a fixed-effects Bayesian model selection to perform family inference and calculate the model posterior probability to determine the winning model in each family. The family model probability results from these 4 different family comparisons are shown in Figure S2. This analysis indicated that, across families, the winning model had (1) a prefrontal hierarchy with laterally connected rDLPFC and rIFG above pre-SMA, (2) a parallel structure of rParietal and rTemporal cortices without a lateral connection, (3) exogenous inputs to both rIOG and rIFG, and (4) backward cortical to deep connections. A diagram of the final model structure is shown in the DCM results section.

The next step in model building involved determining which connections were modulated by conditions specific changes between No-tFUS SS and US trials and, primarily, between No-tFUS SS and Stop-tFUS SS trials. For our purposes, this included extrinsic connections between areas and intrinsic gain connections within an area. Instead of testing an expansive set of permutations of condition-specific modulations of extrinsic connections (B matrix in DCM), we a priori opted to use the recently developed Parametric Empirical Bayesian modeling framework

for DCM (Friston et al., 2016). As will be explained below, this involves doing Bayesian statistics on full DCMs that will have all condition modulatory parameters of interest entered a hierarchal-general linear model from which hypotheses can be tested. This obviates the need for conventional statistics and model reduction over all extrinsic connections. If a DCM model is referred to as full, this includes condition-driven modulations of all extrinsic connections. Otherwise it includes an explicitly stated set of connections.

After determining a model structure, we inverted this model for each subject's ERPs. This inversion was performed twice, using different combinations of trials to assess different hypotheses. The DCM was first inverted using the No-tFUS US and SS trial ERPs. The US trial was set as the baseline. This inversion allowed us to first compare how the network connections were modulated between US and SS trials in a baseline network without the effects of tFUS. Specifically, we wanted to examine how changes in connectivity (DCM B matrix) distinguished between unsuccessful and successful inhibition. Analysis of this general linear model (described below) focused on the B matrix which describes how connections changed between conditions. This comparison was done to mirror the standard comparison of SS and US trials that is typically employed to reveal what inhibitory mechanisms were more potently activated during SS trials (Aron et al, 2014). Furthermore, using this contrast of conditions provides a baseline to ask whether the same connectivity mechanisms were altered when comparing No-tFUS and Stop-tFUS SS trials. Therefore, the second DCM model inversion was applied to the No-tFUS and Stop-tFUS SS trials. The No-tFUS trials were used as a baseline for the fit. Again, this allowed a focus on the connectivity modulation between the two conditions and treated the non-stimulation condition as the baseline network.

To analyze the resultant DCM condition-specific changes and the modulation of connectivity parameters by experiment relevant variables (e.g., SSRT), we tested group-level effects using the Parametric Empirical Bayes (PEB) framework. We give a brief overview of

PEB (for in-depth discussion, see Friston et al., 2016) for hypothesis testing and connectivity parameter extraction using a PEB. Building a PEB statistical model involves creating a hierarchical model with, in our case, two levels. The lower level is the subject level, which is the results of DCM fits to individual's ERP data. This first level includes the posterior means and uncertainties for each subject's DCM connectivity parameters. The PEB framework statistically models these parameters using a Bayesian general linear model (GLM) at the group level. As is the case with GLMs (and mixed models), the model attempts to explains the connectivity parameters as between-subject and within-subject variability, while allowing for between-subject differences in connectivity parameters to be treated as random effects. The PEB Bayesian GLM allows using subject-based DCM parameters to be examined using a linear model with respect to explanatory variables at the between-subject (group) level. Because the PEB framework yields group-level estimates as empirical priors, this approach also allows changing parameters that were estimated at the subject level by allowing them to be estimates distributed around a group mean effect. This process stands in contrast to the typically used "summary statistic" approach, which often involves applying several t-tests or correlations per connection.

For clarity, we built two separate PEB models for the DCMs fit to (1) No-tFUS US and SS trials, and (2) DCMs fit to No-tFUS and Stop-tFUS SS trials. The former PEB model was used to test hypotheses regarding mean changes in connectivity between unsuccessful and successful inhibition at baseline (No-tFUS). The latter PEB model was used to address (1) how the inhibition network connectivity changed on average between tFUS conditions (intercept in GLM), and (2) how individual differences in the change in SSRT between tFUS conditions was embedded in changes in connectivity parameters across subjects (covariate). Therefore, in the first model we only used the mean intercept as the explanatory variable to examine mean changes in connectivity between No-tFUS US and SS trials. The mean change parameters represent the gain change in connectivity going from US to SS trials. In the second model,

which examined No-tFUS SS and Stop-tFUS SS trials, we used a design matrix of explanatory variables that included an (1) intercept representing mean changes in connectivity (mean tFUS effect), (2) the change in individual subject's SSRT between tFUS conditions, , and (3) the tFUS change in individual subject SSRT variability. Before entering the SSRT covariates, they were transformed to a gain change by taking the log-ratio of Stop-tFUS SSRT (mean or variability) over the No-tFUS SSRT (mean or variability). The covariates were then z-scored to have a zero mean and yield standardized PEB model parameters.

When initially estimating each of the two PEBs, we always entered in the full model including all extrinsic and intrinsic connection modulatory parameters. Because our hypotheses centered around changes along prefrontal and deep areas, and ventral pathway interactions, we first compared the PEB model with and without dorsal pathway connection parameters. We show in the results that, across both PEBs, the model without dorsal pathway connection changes yielded a better model with respect to the explanatory variables. This allowed a substantial reduction in connectivity parameters needing to be tested.

Once the group-level GLM parameters were estimated with respect to modulations of extrinsic and intrinsic connectivity, we used this framework to test several hypotheses regarding mean and SSRT driven changes in connectivity parameters. Hypothesis testing proceeds by Bayesian model reduction of the GLM. This involves turning off/on different connectivity parameters and comparing the free energy of reduced models. Comparing models in this manner is similar to performing classical hypothesis testing via model reduction in mixed models by employing likelihood ratio or F-tests.

Hypotheses were tested by designing models with different combinations of parameters on/off. The model space of these hypotheses was defined in a factorial space that focused on 4 factors that could be modulated by mean connectivity changes (PEBs 1 and 2) or connectivity modulation via tFUS induced changes in SSRT (PEB 2). These factor/hypothesis spaces were

driven by previous work. The first set of models considered how inhibition is related to pathways the backward connections from rIFG and pre-SMA to the deep (i.e., basal ganglia) nodes. The second factor, driven by work indicating that pre-SMA interacts with rIFG before deep projections (Rae et al., 2009), tested for modulation of their backwards and forwards connections. The third factor was motivated by proposals that differences in inhibition (SS vs. US) might be mediated changes in attentional orienting, which predict changes in intrinsic self-inhibitory gain in either rIFG, rIOG, or both. Therefore, this factor examined for modulation of the superficial pyramidal cell gain across the nodes. Finally, to examine how SS versus US and different tFUS effects on inhibition depend on top-down vs bottom-up processing, our final comparison tested for the inclusion of either backward, forward or both sets of connections along the ventral pathway. Given that each of 4 factors had 3 levels each, plus a null (all zero) model, the first PEB surmounted to testing 3^4+1 (82) models, and PEB model 2 included $2*3^4+1$ (163) models.

Rather than summarize these effects as the free energy for each model, which would surmount to a severe reduction in discerning the probability of a winning model, the hypotheses were grouped into families, and model hypotheses were tested at the family level. After using family model comparison on reduced GLMs, we used Bayesian model averaging (BMA) to obtain GLM model estimates of connectivity parameters having a posterior probability >95%. BMA was performed on all families within a factor, weighting the summarized parameters by the probability of the family. For example, BMA parameters for the second PEB mode that includes the mean SSRT as a predictor, for example, should be interpreted as would a linear regression coefficient; similarly, the mean term would represent the mean change in connectivity. These BMA parameters were reported as the final changes in connectivity and should be considered to have a 95% probability of being non-zero.

Before conducting our main analysis presented, and to reduce the model space, we first considered whether the best PEB model would include ventral and dorsal pathway projections after fitting to individual subjects. Specifically, we asked if describing the baseline difference of no-tFUS SS and US inhibition involved modulation of either or both the dorsal and ventral pathways. Using the family-based hypothesis comparisons test described above for PEBs, we created factor spaces including the dorsal, ventral, or both pathways. Both sets of pathways included top-down and bottom-up connections that were grouped together. The dorsal pathway included the rParietal and rDLPFC nodes and the ventral pathway included rIFG and rTemporal. Bayesian model comparison (probability = 1) revealed strong evidence in favor of the model with just ventral pathway connection modulations from US to SS trials. Based on this result, we excluded Dorsal pathway connections from the rest of our DCM analysis to reduce the space of parameters.

Results

Three different groups of human participants participated in a Stop-Signal task during online tFUS. On random trials, MRI-guided tFUS was applied in different locations according to the group assignment. The main experimental group involved tFUS applied to the pars opecularis section of rIFG, while others served as a control site (ipsilateral S1) and auditory sham control that did not receive tFUS. We applied tFUS to the rIFG group at different time points during Go and Stop trials. We hypothesized that if rIFG implemented motor inhibition, then tFUS effects would be limited to alteration of stopping but not going. Neural impact of tFUS on putative inhibition related ERPs were considered at sensor and MRI-based source localized, and DCM network levels.

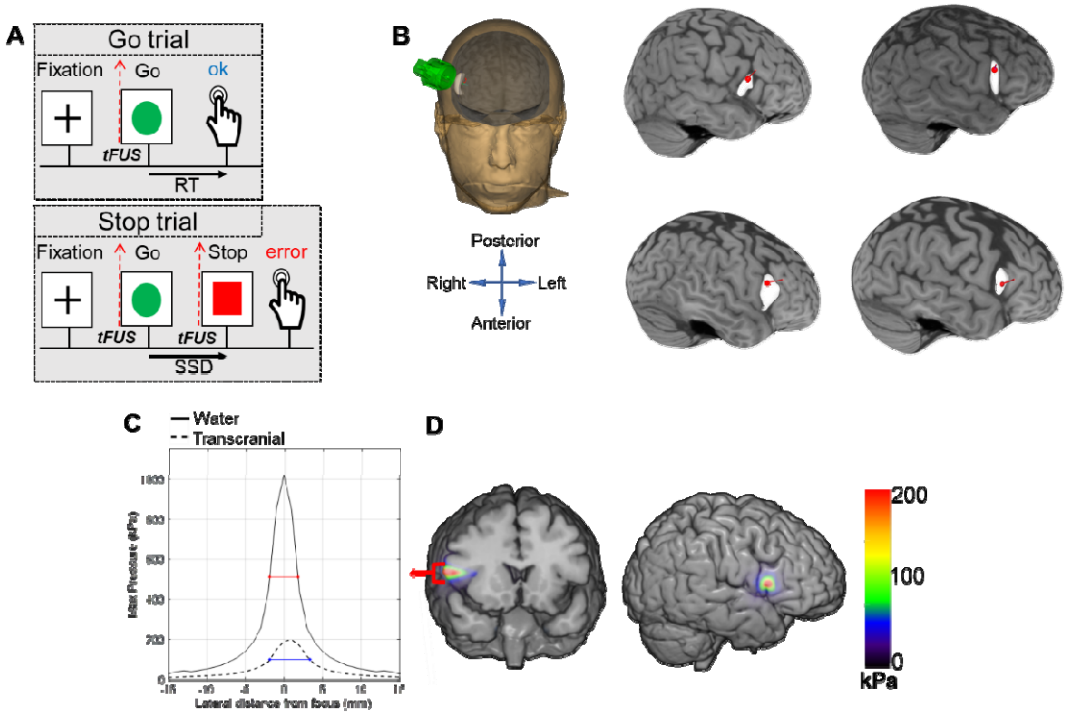


Figure 1 **A.** visual layout of Stop-Signal task and corresponding times when transcranial focused ultrasound (***tFUS***) was delivered. **B.** The left plot shows the average neuronavigation location of tFUS applied to all MRIs used in the rIFG group. The structural scans on the right show renderings of the targeted tFUS point in pars opercularis in 4 subjects. **C**. Ultrasound simulations on a CT scan from one patient. Left plot shows lateral maximum pressure profile obtained at 30-mm depth focus in both water and transcranial simulation. Horizontal red and blue lines show 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.

Numerical simulation of tFUS to rIFG

To determine intensity and accuracy of tFUS after skull transmission in the rIFG group, we used numerical simulations based on CT and MRI data from 3 preoperative patients. We predicted an average maximum intensity at the focus was 2.8 W/cm^2. This is within the range of previous work indicating non-thermal neuromodulatory effects (Legon et al., 2014). At the measured focus, the shift in peak pressure due to skull transmission was 1.25 mm laterally and had a lateral full-width half maximum of 5.1 mm (Fig. 1C). These simulations indicate a high spatial precision with >95% of energy limited to *pars opercularis* (Fig. 1D).

Only tFUS to rIFG alters response inhibition

We first addressed how probability of failing to inhibit a response, P(respond|signal) (Figure 2A), changed across tfUS conditions within and across groups. This was done by fitting a 2-parameter logistic linear mixed-model to obtain a slope (**β)** measure of the response inhibition curve across all subjects and tFUS conditions. The model provided a good fit to the dataset with a mean R^2 = 84%. Analysis of **β** indicated only rIFG group exhibited a tFUS altered P(respond|signal) (Figure 3). 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 paired one-way ANOVAs across tFUS conditions, but within-groups, showed only the rIFG group exhibited differences across tFUS levels. Follow-up t-tests in this group showed **β** for Stop-tFUS was lower than both the No-tFUS and Go-tFUS conditions (both $p < 0.01$: mean **β**'s indicating change in probability for an approximately 25% change in normalized SSD: No-tFUS = 0.35 (0.12), Stop-tFUS=0.27 (0.08), Go-tFUS=0.35 (0.11)). Results indicate only the rIFG group was affected by tFUS, as expected, and only during Stop-tFUS.

Figure 2A shows improved inhibition performance during Stop-tFUS for the rIFG group occurred at longer SSDs (65% and 95% SSD). A repeated-measures ANOVA on P(respond|signal) for the rIFG group across all SSD levels and tFUS conditions revealed a significant interaction ($F(6,102) = 8.21$, $p < 0.0001$, $\eta_p^2 = 0.33$). Contrast 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 the Stop-tFUS in the highest two SSDs (all $p < 0.01$; Bonferroni $\alpha = 0.0125$). Effects of Stop-tFUS yielded both improved inhibition performance during stop-signal presentation and was pronounced at later SSDs.

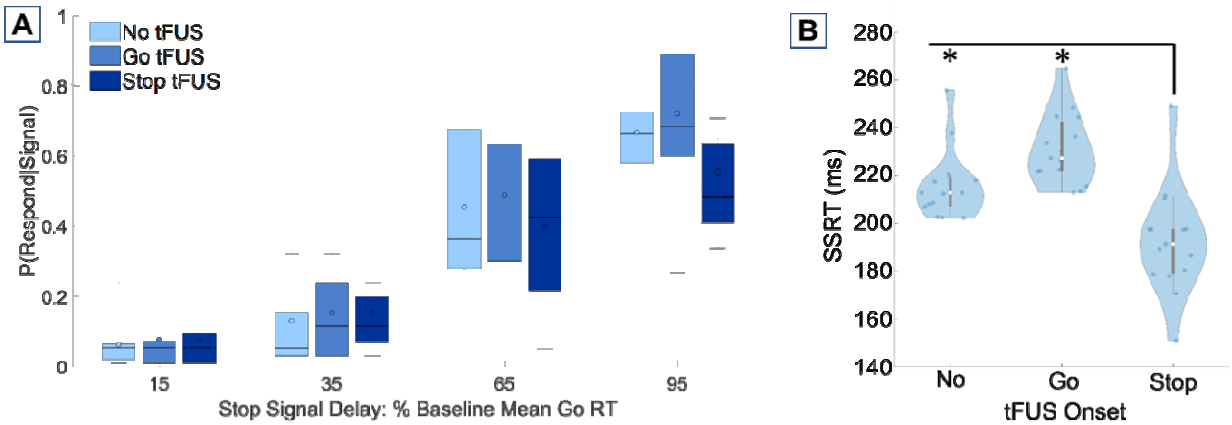


Figure 2. **A.** The probability of responding across stop signal delays in the rIFG tFUS group for all tFUS conditions. **B.** Violin plots showing across-subject distribution of stop signal reaction times (SSRT) calculated across subjects for the rIFG group.

We anticipated rIFG tFUS based changes to P(respond|signal) resulted from shortening of the stopping process speed (i.e., SSRT) because it did not affect other behavioral variables (e.g., Go RTs, see Supplementary Material). SSRT analysis in a mixed-design ANOVA indicated a significant Group x tFUS interaction ($F(4,100) = 10.2$, $p < 0.001$, = 0.21). Follow-up one-way ANOVAs within-groups indicated only rIFG group SSRTs (Figure 2B) differed between tFUS conditions ($p < 0.05$). A t-test confirmed their SSRTs were shortest during Stop-tFUS (Figure 2B). This result, along with the above P(respond|signal) outcome, indicate Stop-tFUS to rIFG altered inhibition by a shortening of the stop process (SSRT).

Neural Components Underlying Inhibition

In the rest of the results we focus our analysis on the rIFG group because this was the only group exhibiting behavioral effects of tFUS. Furthermore, we only analyze No-tFUS and Stop-tFUS conditions because Go activity was subtracted from neural data at Stop trials (see Methods).

Scalp ERPs are altered by tFUS. Our first analysis was sensor-level ERPs. We examined three contrasts using the cluster-based permutation t-tests: (1) successful –

unsuccessful contrast over both tFUS conditions, (2) successful (No-tFUS) – successful (Stop-tFUS) contrast, and (3) interaction comparing successful – unsuccessful between the No-tFUS and Stop-tFUS conditions. The cluster-based scalp maps show the progression of these clusters time-locked to the stop-signal onset (Figure 3A). All ERPs in Figure 3 are locked to stop-signal onset (0 ms).

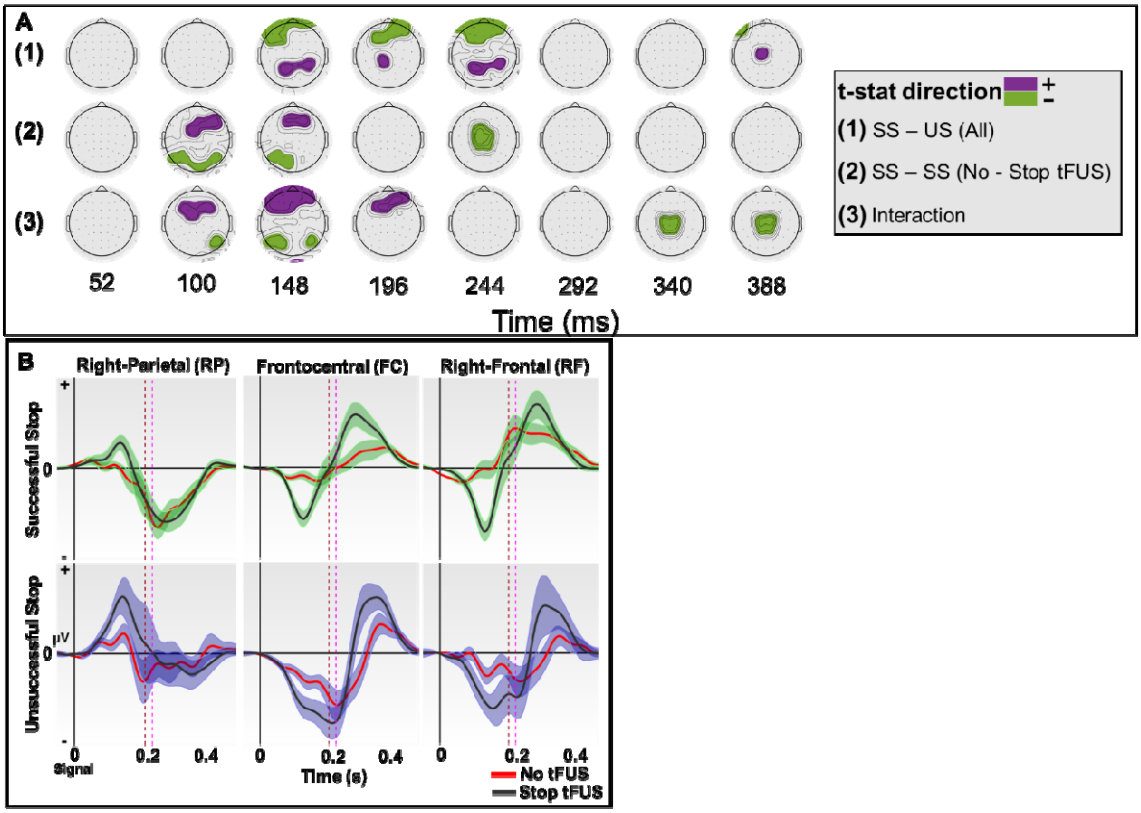


Figure 3. (A) scalp plots showing time-dependent cluster-corrected permutation paired samples t-tests ($p < 0.01$). Colored contours represent significant clusters. (1) Significant difference of successful stopping (SS) minus unsuccessful stopping (US). The SS and US are the average of both No-tFUS and Stop-tFUS trials. (2). Significant cluster differences of SS No- tFUS and SS Stop-tFUS trials. (3). The interaction term calculated as SS-US of No-tFUS trials minus the SS-US of Stop-tFUS trials. (B). Average ERP time courses of three clusters identified by the permutation testing that were differentiated by statistical contrasts. From left to right, the first cluster (left column) is right-parietal electrodes (CP6, CP4, P6, P8), the second cluster (middle column) is fronto-central electrodes (C1, Cz, C2), and the third cluster (right column) is right-frontal electrodes (F8, F6, F4). The vertical dashed lines represent the stop-signal reaction times for the No-tFUS (magenta) and Stop-tFUS (maroon). The time-axis onset at 0 ms corresponds to the stop signal in both plots (A) and (B).

We identified several ERPs differentiated SS and US trials. The first ERP contrast of SS-US (Figure 3A, row 1) indicated SS trials exhibited a larger ERP centralized over a right-parietal cluster around the time-range typically found for the P100 peak (100-148 ms). The interaction on SS-US and No-tFUS and Stop-tFUS showed this effect occurred earlier and during the peak onset (80-100 ms) during Stop-tFUS. This P100 interaction occurring during the ERP onset further indicated the difference increased during SS tFUS trials. Consideration of the SS-SS contrast (tFUS effect) supports the conjecture that the success and stopping speed is directly related to larger P100 sensory/parietal responses. This supports other studies reporting an enhanced sensory ERP during successful stopping (Boehler et al., 2009; Bekker et al., 2005).

We also examined the N100 in the right-frontal cluster in the 80-120 ms window (Figure 3B). We did not find a difference in the frontal N100 with respect to the overall SS – US contrast (Figure 3A). This result is in line with several other studies noting a non-significant effect of this contrast (Kenemans, 2015). However, we did find the ERP was substantially larger in SS tFUS trials compared to SS No-tFUS trials indicating a direct contribution to stopping efficacy.

Another ERP that mainly predicts failed inhibition is a right-frontal N200 (Liotti et al., 2010; Wessel et al., 2015). The ERP peaking around 200 ms appeared mostly, if not only, in failed stop trials (Figure 3B). This N200 was apparent in both the No- and Stop-tFUS conditions (Figure 3, bottom right). If N200 scaled with the proximity of a stopping process to successful inhibition, and tFUS merely reduced the overall threshold for stopping, one would expect a smaller N200 during US Stop-tFUS trials. However, Fig. 3B shows that N200 was, in fact, larger for US Stop-tFUS trials.

Examining the frontocentral cluster revealed the standard N200/P300 complex most commonly associated with response inhibition (Huster et al., 2013; Kenemans, 2015). SS – US contrast indicated the N200 was larger in US trials. This aligns with previous studies indicating the N200 is often only apparent in failed stop trials (Greenhouse and Wessel, 2013; Ramautar

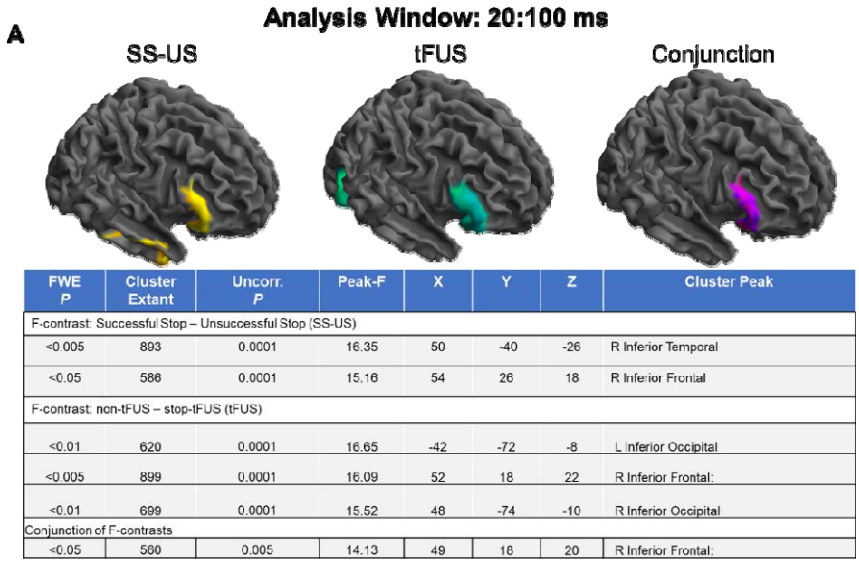
et al., 2004). An interaction also resulted from the N200 differing between tFUS conditions only for US trials. The frontocentral N200 amplitude was notably larger in the Stop-tFUS trials (Figure 3B, bottom-middle). The frontocentral P300 also differed between SS and US trials. This analysis indicated that the US trials exhibited a larger amplitude around the peak time (290-320 ms) (Figs. 3A-B). However, because the P300 peak occurs after SSRT, this neural marker is presumably too late to reflect inhibition (Huster et al., 2013). Contrasting waveforms of SS trials for No-tFUS and Stop-tFUS (Figure 3B, upper-middle) indicate the P300 component related to SS is the onset latency (Wessel and Aron, 2015). Imperatively, this observation indicates P300 onset and tFUS related SSRT changes were causally related. We confirmed this by analyzing shifts in P300 onset for all subjects. The correlation between P300 onset and SSRT (0.61, $p < 0.05$) indicated a larger shift between the P300 onsets was predictive of the tFUS related SSRT change. This result supports previous claims (Wessel et al, 2015) that P300 is related to inhibitory control and SSRT timing through the onset latency, rather than amplitude.

Whole-brain source SPM and source-based regression analysis

We also examined source-based activity to localize tFUS-induced changes in evoked activity and generating source location priors for DCM. We hypothesized that, if differential rIFG activation indexed SS versus US, then conjunction analysis should reveal an overlap between SS – US and No-tFUS – Stop-tFUS conditions if rIFG activity is related to successful stopping (Aron et al., 2014). As expected, whole-brain SPMs (Figure 4A) revealed the only area exhibiting an overlap was a pars opercularis-centered rIFG peak during the 20-100 ms time window (other windows in Supplementary Material).

To understand how tFUS altered stopping efficacy, we compared changes in SSRT between Stop-tFUS and No-tFUS SS trials using whole-brain SPM linear regression. In the regression, positive contrasts indicate ΔSSRT (No-tFUS – tFUS) is associated with larger activity in No-tFUS trials. Negative contrasts indicate tFUS SS trials exhibit larger evoked

activity predicted by larger changes in SSRT. The negative contrast in Figure 4B indicates that the first site to exhibit increased tFUS-related predictions is rIFG. This occurred both in the -40:20 and 20:100 time-windows. The positive contrast found in early time windows indicates Stop-tFUS exerted effects on stopping by also decreasing early activity in both bilateral parietal and right temporal sites. We also found pre-SMA was only predictive of SSRT after rIFG, with the pre-SMA modulation peaking at 100:180 ms. These results show that areas typically associated with successful inhibition were predictive of tFUS-induced changes in behavior while occurring before the SSRT itself.



FWE *P*	Cluster Extant	Uncorr. *P*	Peak-F	X	Y	Z	Cluster Peak
F-contrast: Successful Stop – Unsuccessful Stop (SS-US)							
<0.005	893	0.0001	16.35	50	-40	-26	R Inferior Temporal
<0.05	586	0.0001	15.16	54	26	18	R Inferior Frontal
F-contrast: non-tFUS – stop-tFUS (tFUS)							
<0.01	620	0.0001	16.65	-42	-72	-8	L Inferior Occipital
<0.005	899	0.0001	16.09	52	18	22	R Inferior Frontal:
<0.01	699	0.0001	15.52	48	-74	-10	R Inferior Occipital
Conjunction of F-contrasts							
<0.05	560	0.005	14.13	49	16	20	R Inferior Frontal:

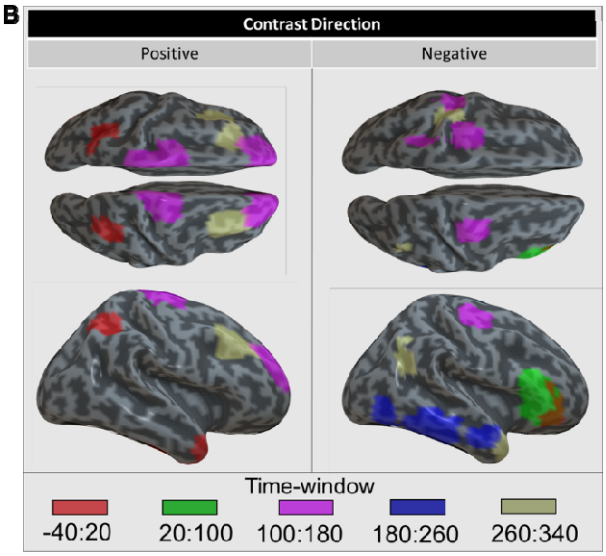


Figure 4. **A.** Whole-brain SPM F-contrasts of evoked activity in the 20:100 ms window. Figure shows the surface mesh projections of the overall SS – US (left) tFUS contrast (middle), and conjunction (right). Below the surface meshes is a table listing the statistics. **B.** Whole-brain SPM linear regression of activity No-tFUS and Stop-tFUS SS trials against SSRT changes in No-tFUS and Stop-tFUS SS trials.

tFUS effects in an inhibition network: Dynamic Causal Modeling (DCM)

The above tFUS results provide evidence that rIFG activity is causally related to inhibition. Principally, differences in local activation can result from both local and inter-areal connectivity (David et al., 2006). We used DCM to quantify network effects. Bayesian model selection established a winning model as a hierarchical network in which rDLPFC and rIFG sat at the top of the hierarchy and pre-SMA being below these areas. The model also included nodes for right temporal (rTemporal), right inferior occipital gyrus (rIOG), right parietal (rParietal), a hidden subcortical node (Deep) and left motor cortex (M1). These results accord with networks proposed by previous studies implicating both motor inhibition and attentional orienting (Weicki and Frank, 2013; Munkata et al., 2011). The DCM (Figure 5) was fit to individual subject's data to determine how connectivity parameters differentiated (1) No-tFUS baseline SS and US inhibition differences, and (2) between No-tFUS and Stop-tFUS SS trials and accompanying changes in SSRT. The above approach revealed the models agreeably captured the spatiotemporal properties of the ERP scalp data across both sets of model fits (Figure 5).

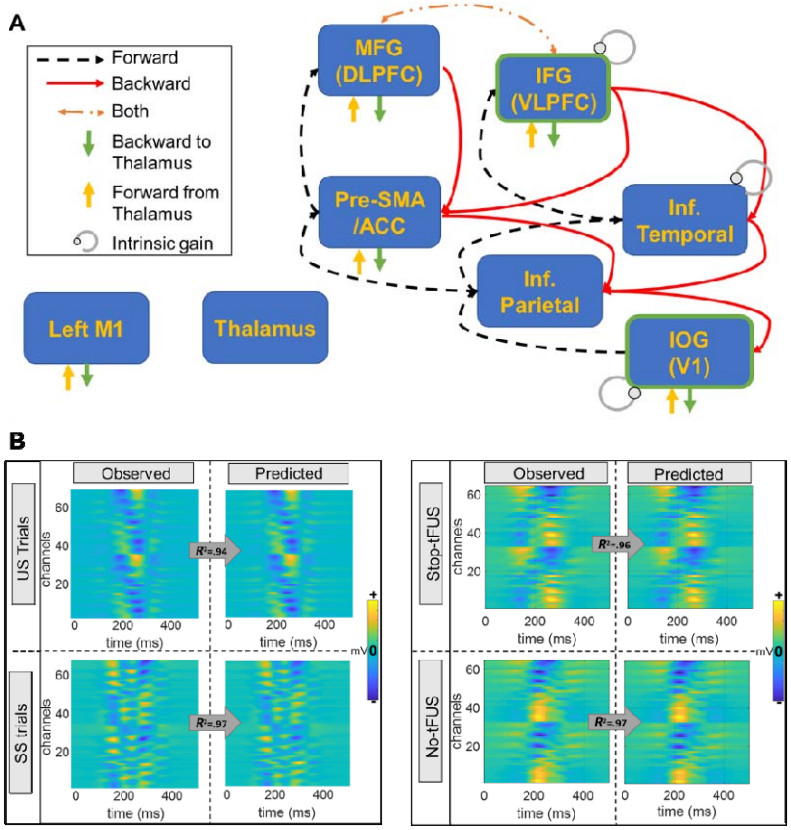


Figure 5. **A**. Connections used to implement the final dynamic causal model structure. Exogenous inputs entered through rIFG and rIOG represented as a green box. All nodes except Left M1 and Thalamus were all located in the right hemisphere. **B**. Left panel shows the mean observed and predicted scalp ERP data derived from the dynamic causal model fit to the No-tFUS US and SS trials. Right panel shows the mean observed and predicted scalp ERP data derived from the dynamic causal model fit to the No-tFUS SS and Stop-tFUS SS trials. These results show that the final model provided a good fit to the data. Note that "0" on the x-axis denotes the time at which the Stop signal was delivered.

The first set of analyses examined what connectivity parameters were altered during No-tFUS successful inhibition. The first comparison examined whether differential rIFG and pre-SMA interactions were related to baseline successful stopping, as suggested by several functional and anatomical studies (Obeso et al., 2013; Rae et al., 2015; Swaan et al., 2012; Figure 6A). Model comparisons revealed a winning family included interactions between both areas, but with a moderate posterior probability (0.78). BMA across families revealed both the connection from rIFG to pre-SMA (136%) and pre-SMA to rIFG (193%) were altered during SS trials. Changes in this connection suggest that SS trials were supported by prefrontal

interactions. Considering theoretical models proposing that increases in pre-SMA projections (Forstmann et al., 2008) to striatum render increased RTs, this result might reflect an effect of blocking the impetus pre-SMA provides towards responding to the Go signal (Logan et al., 2015).

The next comparison tested the hypothesis that both pre-SMA and rIFG projections to deep areas are necessary for successful inhibition (Figure 6B). Only a family including changes from rIFG to deep was predictive of successful inhibition. BMA indicated rIFG exhibited a mean reduction in the backwards connection of 65% during successful stopping. This result agrees with previous fMRI studies (Aron et al., 2006; Jahfari et al., 2011; Li et al., 2006) indicating inhibitory processes by driven by cortical to basal ganglia interactions (Aron et al., 2016; Jahafari et al., 2011).

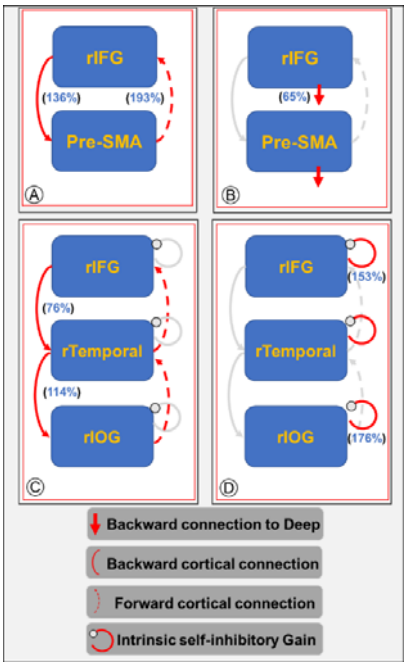


Figure 6. Family-based model comparison for different hypothesized interactions. Plots contain modulatory parameter of connections with strong positive evidence as being different between US and SS trials. Parameter estimates (>95% posterior probability) are in parentheses next to modulated connection in exponential percentage change. Anything above 100% reflects an increase in SS trials compared to US (opposite for below 100%). A. Test of rIFG and pre-Sma interactions. B. Tests of rIFG and pre-SMA backward projections to deep node. C. Tests for the ventral pathway connections. D. Comparison of intrinsic superficial pyramidal cell gains.

Another hypothesis that has been put forth regarding successful inhibition is that it is mainly mediated by attentional orienting in ventral pathways (Hampshire and Sharp, 2015). DCM implementations of attentional process can be cast in terms of hierarchical predictive coding models (Feldman and Friston, 2010; Friston, 2008). Previous work suggests increased attention in sensory areas emerges as increased recurrent intrinsic gain (increased disinhibition) of superficial pyramidal cells thought to report prediction errors through forward connections (Auksztulewicz et al., 2015). In DCM, recurrent gains would weight prediction error signals driven changes in top-down (backward) connections (Friston, 2008). We designed the next two comparisons to examine whether SS trials exhibited these network changes. Extrinsic connection analysis in No-tFUS SS indicated a winning family including only changes in top-down backward connections predicted successful inhibition (Pp =1). BMA indicated, however, the inhibitory connection from rIFG to rTemporal was reduced while backward connections from rTemporal to rIOG increased in connectivity (Figure 6C).

The final SS and US modulation comparison analyzed changes in recurrent gains of rIFG, rTemporal and rIOG (Fig. 6D). The winning model (Pp =0.99) included an increase in gain for both rIOG (176%) and rIFG (153%). Importantly, the increased rIOG gain is predicted by attentional orienting models of response inhibition (Hampshire and Sharp, 2015) and accords with previous DCM studies that have manipulated attentional cueing (Auksztulewicz et al., 2015). Mechanistically, the increased rIOG gain in SS trials results in ascending signal that has a larger effect on decreasing rIFG top-down expectation signals in backward connections.

Stopping efficiency (SSRT) is driven by lower and prefrontal interactions

Another primary goal of understanding inhibitory control is quantifying how the efficiency of stopping, evaluated through the SSRT, is implemented via network pathways. Previous work has employed between-subjects' correlations of SSRT and connectivity parameters to isolate

the pathways involved in this process (e.g., Jahfari et al., 2011). Because previous studies using fMRI and a variety of connectivity methods revealed pre-SMA and rIFG to basal ganglia connections are correlated in different directions with SSRT (Jahfari et al., 2011), we adopted this approach when analyzing the baseline No-tFUS SS versus US contrast. This correlational analysis indicated both connections were correlated in opposite directions with SSRT. Figure 7 shows that increasing connectivity from rIFG to deep targets predicted shorter SSRTs. The opposite pattern was found for pre-SMA, wherein increasing pre-SMA backward activity predicted longer SSRTs.

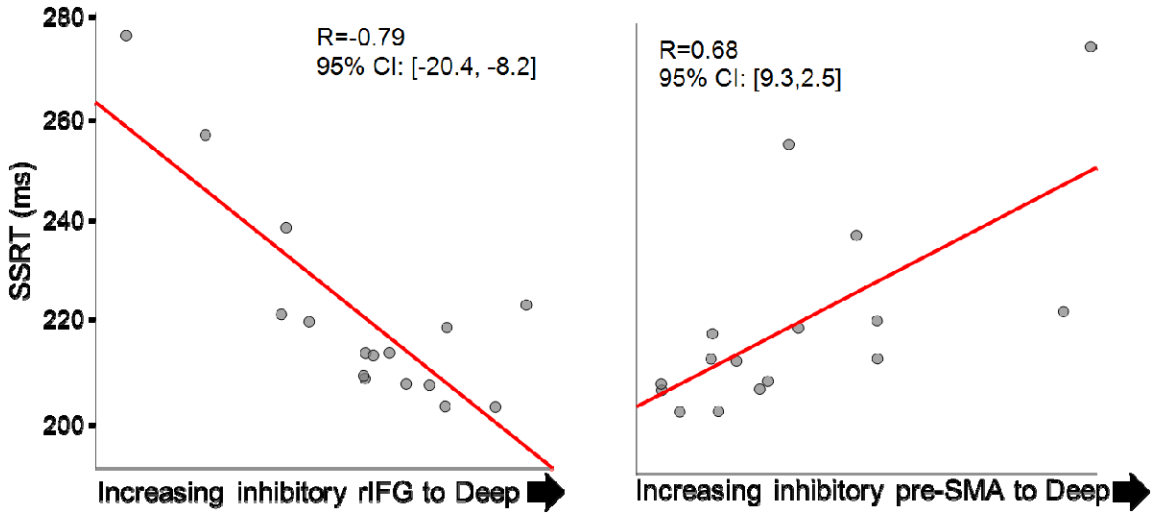


Figure 7. Left and right plots show linear regression fits predicting individual subjects' No-tFUS SSRT from the backward connection to the deep area projecting from rIFG and pre-SMA, respectively. Larger values on the x-axis denote decreasing backward inhibitory connectivity.

tFUS to rIFG causally dissociates inhibitory mechanisms

Building on tFUS's effect of increasing inhibition performance, a primary question was whether the changes in connectivity strength between No-tFUS SS and Stop-tFUS SS trials would reflect stepwise changes in the connections modulated in the SS and US No-tFUS comparison. This comparison is in line with the idea of failed stop trials resulting partially from

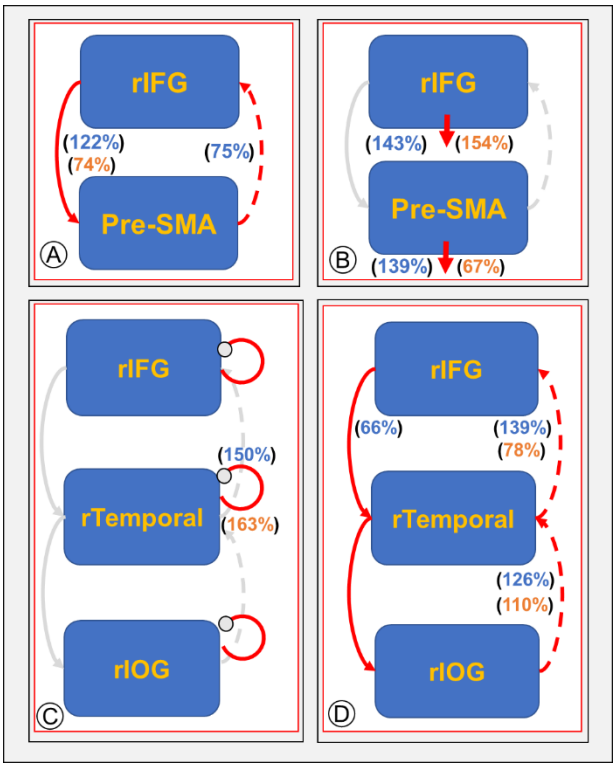
less active inhibition mechanisms. By applying tFUS to rIFG, we were able to causally dissociate rIFG's role in implementing inhibitory mechanisms. In the following results, mean changes represent the average modulation of connectivity going from No-tFUS SS to Stop-tFUS SS trials. SSRT based effects represent how changes in SSRT between tFUS conditions predict the change in connectivity between tFUS conditions.

In first analyzing the family of models examining the rIFG and pre-SMA interaction (Figure 8A), there was no clear winning family of models for both the mean and SSRT effect. For the mean, the model with projections from rIFG to pre-SMA (122%) and pre-SMA to rIFG (75%) was highest in probability (Pp=0.79). Of these connections, SSRT change was only negatively related to the rIFG to pre-SMA pathway (r^2 = 74%). indicating that an increase in inhibitory connectivity from rIFG to pre-SMA during Stop-tFUS trials predicted a larger change in SSRT. Therefore, changes in SSRT were not related to forward connections between these nodes, but instead were modulated by a top-down inhibitory modulation from rIFG to pre-SMA.

When comparing models testing the role of rIFG or pre-SMA to deep backward connections, family analysis revealed a strong effect of the mean change and SSRT change for both backward connections (Pp>0.95 both effects). BMA in the winning family indicated both effects were above threshold for both connections. Both rIFG (143%) and pre-SMA (139%) exhibited increased connectivity during tFUS. Concerning changes in tFUS related changes in SSRT, the direction of effects was opposite for rIFG (154%) and pre-SMA (67%): for rIFG, decreased inhibition yielded increased SSRT change, and the opposite for pre-SMA. Together, these results point to differential interactions of both pre-frontal areas with a deep node in responding to tFUS. These results are further in line with the directionally opposite linear correlations of the No-tFUS SSRT to each of these backward connections (Figure 8B).

Finally, we examined ventral pathway parameters for tFUS modulation of top-down, bottom-up, and intrinsic gain changes. Model averaging revealed the rIFG to rTemporal top-

down connections (66%) decreased with tFUS, suggesting a direct effect from rIFG tFUS. This decrease in top-down rIFG connection agrees as well with the results found during the SS versus US comparison, suggesting a causal and overlapping pathway driving inhibition performance. All bottom-up connections increased as well (Fig. 8C). With respect to SSRT change (Figure 8D), only the bottom-up connections were related to changes in the SSRT. However, the forward connectivity relations to SSRT were in opposite directions for modulations of the rIOG to rTemporal (110%) and rTemporal to rIFG (78%). Finally, we found rTemporal recurrent gain increased in mean (150%) and was positively related to changes in SSRT (163%). This relation between the SSRT effect and rTemporal to rIFG forward connection was strong enough such that a leave-one-out cross validation prediction of SSRT change using this connection exhibited a large correlation with actual SSRT change (0.84, $p < 0.01$) Figure 8 (bottom panel). These results indicate neural implementation of stopping speed is implemented by a trade-off in favor of relying on bottom-up temporal cortex prediction error signals.



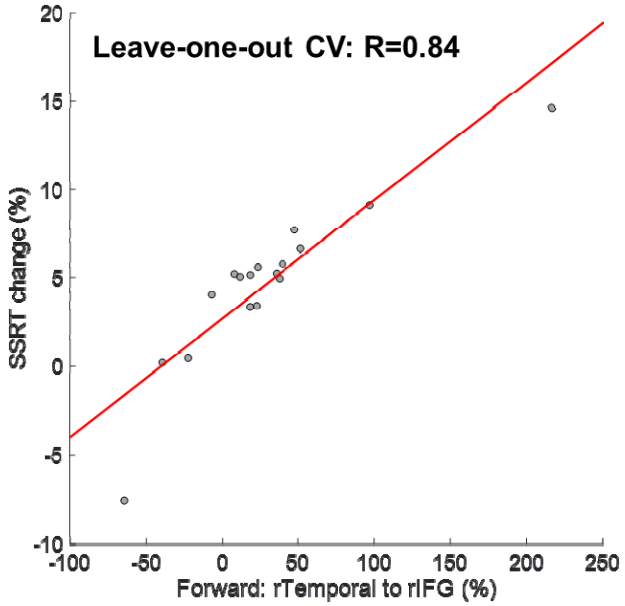


Figure 8. Results of the family-based model comparison for different hypotheses tested when comparing modulation from No-tFUS SS to Stop-tFUS SS trials. The family-based posterior probability (F-Pp) for the winning model is listed below each model. The outcomes of these plots can be interpreted of as revealing the modulatory parameters connections with very strong, positive evidence of US and SS trials being different. Parameters estimates with a greater than 95% posterior probability in these families are presented in parentheses next to the modulated connection. The parameters are presented in exponential form of percentage change. Values above 100% equates to a parameter increase in SS trials compared to US (and the opposite for values below 100%). Parameters in-active in each model are in a gray color. A: hypothesis test of rIFG and pre-Sma interactions. B: Hypothesis tests of rIFG and pre-SMA backward projections to the deep node. Panel (C) is the comparison of intrinsic superficial pyramidal cell gains. D: Hypothesis tests for the ventral pathway connections. The scatter plot (bottom panel) shows the correlation and correlation coefficient of a leave-one out cross validation prediction of SSRT change as predicted using the rTemporal to rIFG forward connection.

Discussion

The current study addressed how pars opercularis, a posterior sector of rIFG, is involved in motor response inhibition by using a stop signal task, tFUS, source-localized EEG and dynamic causal modeling. Behaviorally, rIFG to tFUS concurrent with stop signal onset increased the likelihood of successfully inhibiting a response. tFUS improvements in inhibition resulted from a faster stopping process (i.e., SSRT), without altering Go or failed-stop reaction times in line with the independent race model of inhibition (Logan and Cowan, 1984). The specificity of rIFG tFUS was supported by whole-brain evoked activity indicating only pars-opercularis exhibited a conjunction effect of tFUS with successful compared to failed inhibition (Figure 3). The hypothesis that pars opercularis activity is directly related to stopping success and its efficacy was confirmed by tFUS driven activity differences predicting between-subjects SSRT change (Figure 4B). Recent work has suggested tFUS effects may result from auditory pathway activation (Guo et al., 2018). Auditory effects cannot explain our findings because control groups exhibited no behavioral effects, and no group exhibited evoked auditory cortex activity (Figure S7). In summary, these findings directly link pars opercularis to a motor inhibition function.

Response inhibition involves several processes embedded in connections across a neural network, including stimulus detection and presumably explicit motor inhibition (Munkata et al., 2011; Weicki and Frank, 2013; Wessel and Aron, 2017). The most

debated is whether there is a dedicated neural motor inhibition mechanism attributable to pars opercularis/rIFG. An established perspective is motor inhibition is implemented in rIFG connectivity to basal ganglia (STN and striatum), either in parallel or routed through pre-SMA (Aron et al., 2014), or both. In contrast, others proposed motor inhibition and rIFG is better understood as an emergent outcome of attentional orienting (Hampshire and Sharp, 2015). Extant evidence indicates, however, that motor inhibition and attentional processing are likely all involved in different processing stages of inhibitory control tasks (Wessel and Aron, 2017).

A core component of models positing a direct rIFG motor inhibition process is that its connectivity with subcortical nodes should change with stopping success or efficacy. DCM analysis of No-tFUS successful and failed inhibition support the hypothesis of rIFG and pre-SMA subcortical connections being necessary for motor inhibition. The results of no-tFUS successful stopping also revealed an anti-correlation of SSRT to pre-SMA and rIFG deep connection strength (Figure 7). It is reassuring our findings agree with another fMRI connectivity study (Jahfari et al., 2011) that found successful stopping accompanied increasing pre-SMA to striatum connectivity and predicted longer SSRTs, with the opposite correlation for rIFG to striatum. Pivotal support for rIFG's role is the result that tFUS effects on SSRT were directly related to changes in rIFG and pre-SMA deep connectivity with a similar anti-correlation pattern as the baseline. We take these results to indicate pars opercularis subcortical connectivity is directly involved in driving motor inhibition.

An outstanding issue regarding the inhibition network is how rIFG and pre-SMA interactions areas relate to inhibition processes (Duann et al., 2009; Rae et al., 2015;

Swaan et al., 2011). DCM model selection indicated rIFG projected to pre-SMA through backwards inhibitory connections and pre-SMA to rIFG via excitatory forward connections. This bidirectional connectivity agrees with previous fMRI stop-signal and DTI studies showing pre-SMA and rIFG connectivity (Duann et al., 2009; Rae et al., 2015; Watanabe et al.,2015). Functionally, No-tFUS DCM results revealed successful stopping was accompanied by increases in both connections. However, the baseline and tFUS comparison indicated mean gain changes only in the rIFG to pre-SMA connection maintained the same increase in connectivity (Figure 6A-B). Furthermore, only the rIFG to pre-SMA connection predicted tFUS changes in SSRT. This result points to increasing inhibitory connectivity from rIFG to pre-SMA as being directionally related to inhibition success. Additionally, the significant pre-SMA to rIFG connection was only relevant during baseline stopping invites an alternative conclusion that pre-SMA is not implementing inhibition per se. Instead, human neuroimaging (Wager et al., 2004), ECOG (Swaan et al., 2012), and TMS studies of proactive inhibition and response switching (Rushworth et al., 2002) suggest pre-SMA encodes the set of potential actions. Applied to response inhibition, this predicts successful stopping relies on pre-SMA signaling the expected action(stopping)-rule to rIFG. Therefore, the action rule should not differ for successful No-tFUS and Stop-tFUS trials, which is in agreement with our findings.

The above results provide causal evidence of an rIFG-based motor inhibition functions are reducible purely to attentional orienting. Nonetheless, rIFG-related attentional orienting and top-down expectations are all relevant factors during behavioral response inhibition (Hampshire and Sharp, 2015; Xu et al., 2017; Wessel

and Aron, 2017). Although attention, for example, was not manipulated here, our approach allows examining how putative attentional, ventral pathway mechanisms differentiated inhibition conditions (Corbetta and Shulman, 2002). Predicted mechanisms were derived from predictive coding formulations of attention (Feldman and Friston, 2010) and DCM-EEG studies directly manipulating attention and stimulus expectancy. Previous DCM (Auksztulewicz et al., 2015; Fardo et al., 2017) findings have indeed shown increased attention linked to increased recurrent gain on ascending (forward) sensory (prediction error) signals, and violations of top-down expectations linked with a decrease in top-down and increase in bottom-up connectivity, respectively.

DCM analysis of baseline stopping featured an increase in rIOG (sensory) gain as expected from attentional effects (Figure 7C). A mechanistic explanation is successful stopping involves enhanced attentional weighting on sensory prediction errors. Concerning top-down changes, rIFG to rTemporal decreased and rTemporal to rIOG increased (Figure 7D). Successful stopping at baseline therefore may rely on an increased rIOG gain weighted sensory (prediction error) signal that drives top-down changes in rIFG. However, finding rTemporal to rIOG increased seems at odds with this interpretation because it implies a larger reliance on top-down information during successful stopping. This is explained by the finding that optimized inputs to rIOG were negative and therefore inhibitory (rather than excitatory) across subjects. Therefore, increasing rTemporal backward connectivity rendered an enhanced negative rIOG signals. Although these results are in line with attention-based models of inhibition , this finding does not imply attentional orienting preceded and drove an explicit rIFG motor inhibition process (Aron et al., 2014; Wessel and Aron, 2017).

Examining how tFUS altered successful stopping only indicated a singular overlap between baseline and tFUS condition changes as a decrease in top-down rIFG to rTemporal backwards connectivity. However, this effect was not directly predictive of a change in SSRT. tFUS effects also involved a increase in bottom-up connectivity that were all predictive of SSRT change (Figure 8D). Again, our findings accord with EEG-DCM studies (Auksztulewicz et al., 2015) indicating expectation violations lead to an increased reliance on bottom-up connectivity. The most interesting effect was recurrent gains only increased in rTemporal, indicating SSRT changes were not due to increased sensory weighting. We propose cortically-related SSRT changes were driven by tFUS altering the effects the rIFG to rTemporal connection had on the rTemporal gain. This predicted rTemporal gain correlations with SSRT change should be accounted for by rIFG to rTemporal connectivity. An auxiliary partial correlation supported this hypothesis by showing the correlation betweenchanges in rTemporal gain change and SSRT was rendered null when accounting for the correlation of rIFG to rTemporal and rTemporal gain. An interpretation is tFUS altered how rIFG backward connections modulated the gain in temporal cortex. In line with previous work (Geng and Vossel, 2013; Xu et al., 2017), the implication is temporal cortex encodes the probability of a stimulus context modulation. This interpretation is consistent with fMRI Stop-Signal studies demonstrating temporal cortex encodes a prediction error of stop-signal probability (Hu et al., 2015; Ide et al., 2013).

In summary, our tFUS and DCM findings demonstrate pars opercularis, and rIFG in general, is directly linked to motor inhibition through direct subcortical projections and cortical projections to pre-SMA and temporal cortex. By altering the response inhibition

network, our results indicate an explicit rIFG inhibition mechanism that are supported by attentional orienting processes instantiated through hierarchical ventral pathway interactions.

Supplementary Materials

S1. Dynamic causal modeling structure optimization

Below are two figures showing the structures used for building and optimizing the dynamic causal model based on the no-tFUS successful stopping event related potentials. The full steps of the process are described in methods in the main text. Figure S1 shows the tested structures and possible connections. Figure S2 shows outputs for comparing these models using Bayesian model comparisons, as well as model comparisons for which nodes receiving thalamic input, and tests for the type of connection projecting from cortical to deep areas (forward, backward, or both).

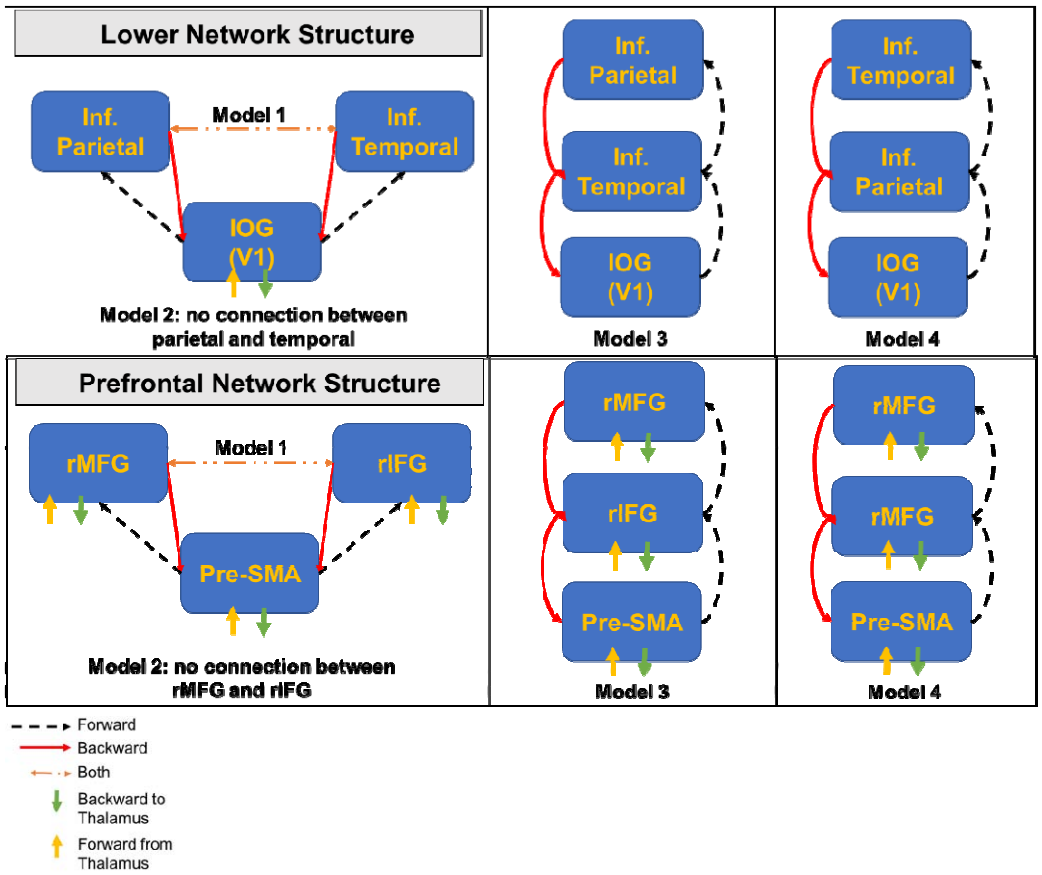


Figure S1. Differences in dynamic causal model spaces that were compared for the lower network structure (top panel) and prefrontal network structure (bottom panel). The legend on the bottom shows how different connections are coded in these putative model spaces.

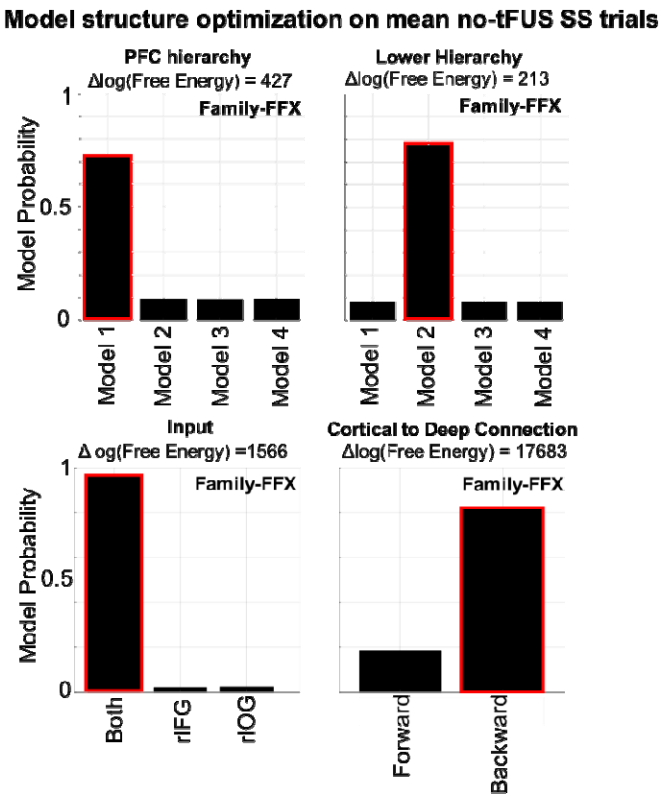


Figure S2. Model probabilities for the different model tests. Each of these tests was examined on the mean ERP data for No-tFUS SS trials using family-wise fixed effect Bayesian model selection. Above each plot shows the model component being optimized. For example, the first bar plot shows the different prefrontal hierarchy arrangements. Above each plot is the difference in log-free energy between the best family and the 2nd best. The log difference of free energies approximates a log-Bayes factor (Penny, 2012). It is considered positive evidence in favor of a model if this value is > 5. Evidence in favor of the winning models (bars enclosed in red) was very strong. The plots on the right show the different structural arrangements used to examine the prefrontal and lower-level hierarchical structure.

S3. Go Reaction Times

We addressed whether tFUS (real or sham) exerted any effects on simply responding to the Go signal. We analyzed this by extracting the ex-gaussian-based mean RT using a maximum likelihood approach (Lacoutoure and Cousineau, 2008). The means were calculated separately for all subjects, and then analyzed using a 3 × 2 mixed-design ANOVA with factors of the 3 groups, and tFUS state: No-tFUS and Go-tFUS trials. This analysis allowed us to assess if “going”, independent of “stopping”, was altered by potential tFUS auditory artifacts, stimulation

to unrelated areas (S1 group), or whether tFUS to rIFG also influenced Go RT processes independent of a stopping context signal. Table 1 shows the per group mean difference of the RT between No-tFUS and Go-tFUS per group. Across groups, there was a consistent negative difference indicating that RTs during tFUS may have been slightly shorter. These differences in means seemed marginally larger for the rIFG group. However, the ANOVA did not reveal any significant effect of tFUS, group, or their interaction (all $p > 0.05$). These results suggest that neither tFUS (rIFG and S1 groups) nor auditory factors alone (sham rIFG group) altered Go RTs independent of a stop signal.

S2. Signal-respond RTs and context independence

Calculating a measure of stopping latency (SSRT) based on the independent race-model (Logan and Cowan, 1984; Bissett and Logan, 2014) assumes that signal-respond RTs are Go processes resulting from a censoring of the Go RT distribution. Testing this assumption predicts that (a) mean signal-respond RT should be faster than mean Go RT, and (b) during fixed-SSD paradigms like the one employed here, signal-respond RT should increase with longer SSDs because there are more failed inhibitory responses. Before testing (a) and (b), we wanted to discern whether our tFUS manipulation exerted any effects on the signal-respond RTs. Based on the context-independence assumption of the race model that signal-respond RTs are Go processes escaping inhibition, and (2) if the tFUS-driven changes in inhibition (P(respond|signal) (Figure 3) are due to changes in inhibition, we should expect no differences in signal-respond RTs between conditions or groups. We used a mixed-design ANOVA to examine the subject level signal-respond RT means with Group (3 levels) and tFUS (3 levels: no-tFUS, Go-tFUS, Stop-tFUS). We found no significant interactions or effects. Importantly, follow-up t-tests revealed no differences across tFUS conditions within the rIFG group (mean signal-respond RT: no-tFUS: 375 ±18 ms; Go-tFUS: 368 ±13 ms; Stop tFUS: 353 ±13 ms). This

supports the conclusion that tFUS did not alter the mean of Go processes that escaped inhibition.

Given the lack of difference in mean signal-respond RTs (across all SSDs) across tFUS conditions, we collapsed these RT means across tFUS conditions and compared them to each subject's mean Go RT with a mixed-design ANOVA. There was no interaction or Group differences ($p > 0.05$). As expected, the mean signal-respond RT (all groups: 365 ±15) was significantly shorter ($F(1,50) = 89.77$, $p < 0.0001$) than the mean Go RT (all groups: 436 ±22). Next, we examined whether the signal-respond RT increased with increasing SSD by regressing all subjects' signal-respond RTs (collapsed across tFUS conditions) onto their SSDs to obtain a single-slope parameter. This revealed a significant regression slope of 0.44 ($p < 0.001$), indicating signal-respond RTs did increase with increasing SSD.

Having confirmed that tFUS did not alter signal-respond RTs or the Go RTs, we sought to test the context-independence assumption of the race model. This has been done in several ways (see Bissett, 2014 for a non-parametric approach). The standard approach for examining this assumption is comparing predicted signal-respond RTs from fitting the independent race model to the observed signal-respond RTs (Verbruggen and Logan, 2009). The independence assumption is typically assessed by showing that observed signal-respond RTs and those predicted by the independent race model are not different. Because we used a parametric (ex-gaussian) based approach to estimate the SSRT (Matzke et al, 2013), we verified these derived fits by using a posterior predictive model comparison to the observed data. The models were used to simulate 500 predictive distributions of signal-respond RTs to estimate the absolute goodness of fit (Gelman & Hill, 2007) to each individual subject's signal-respond RT. This approach generates *p*-values that test for the difference in the predicted and observed signal-respond RTs at each SSD level. The typical metric for assuming goodness of fit is that the *p*-value is close to 0.5, while being below and above 0.05 and 0.95, respectively, is considered a

poorly predictive model. As pointed out by Matzke et al. (2013), these estimates are most stable for SSDs in which several signal-respond RTs are observed. Therefore, we analyzed each subject's p-values averaged across the two highest SSDs.

The mixed-design ANOVA across groups and tFUS conditions did not reveal any effects or interaction on the *p*-values. This result also agrees with the analysis showing no differences in signal-respond RTs between tFUS conditions, indicating the Bayesian procedure for estimating SSRT provided accurate predictions of signal-respond RTs. For all participants, the *p*-values were in the range of 0.1 to 0.9 with a mean of 0.48 and standard deviation of 0.2.

S3. SSRT Variability was not altered by tFUS.

One possible driver of increased response inhibition performance in the rIFG tFUS group is that it may have reduced within-subject SSRT variability. Thus, because tFUS did not alter mean or variance of Go RTs, if it reduced SSRT variability, this would increase the likelihood of successful inhibition. Using the same statistical approach for the mean SSRT in the main text, we found no significant effects of SSRT variability between any of the conditions (all $p > 0.05$).

S4. SS (No-tFUS) – SS (Stop-tFUS) whole-brain regression table

The table below lists the peak coordinates of clusters of the difference in evoked activity of No-tFUS and Stop-tFUS SS with the between-subjects changes in SSRT. These coordinates were used to identify which areas exhibited differential activation with respect to the speed of stopping across subjects. Coordinates for relevant regions of interest were used as prior locations in subsequent dynamic causal modeling in the main text.

Regression of ΔSSRT and SS (No-tFUS) – SS (Stop-tFUS)					
Time-Window (ms)	Regression Direction	Region	Peak Coordinates (X,Y,Z MNI)	Z-value	Extent (Voxels)
-40:20	Positive	R Supramarginal	56, -40, 16	2.05	316
		L Supramarginal	-56, -50, 18	1.94	123
		R Inferior Temporal	40, -4, -40	1.92	45
	Negative	R Inferior Frontal	40, 30, 10	1.94	55
20:100	Negative	R Inferior Frontal	56, 28, 12	2.64	455
100:180	Positive	R Superior Frontal	26, 50, 18	3.18	568
		R Paracentral Lobule	6, -20, 68	2.98	389
		L Paracentral Lobule	-6, -20, 68	2.94	563
		L Superior Frontal	22, 52, 28	2.74	574
	Negative	L Supramarginal	-60, -28, 38	1.76	52
180:260	Negative	R Inferior Occipital	46, -82, -6	4.12	41
		R Inferior Temporal	58, -40, -24	3.62	552
		L Inferior Temporal	-58, -38, -18	3.55	342
		R Middle Temporal	62, -2, -24	3.16	22
260:340	Positive	L inferior Frontal	-46, 14, 32	4.25	39
		L Middle Frontal	-30, 22, 34	2.84	236
		R Middle Frontal	26, 14, 42	2.61	442
	Negative	R Angular	44, -62, 38	3.62	62
		L Postcentral	-42, -34, 52	2.91	92
		R Middle Temporal	50, 8, -30	2.81	73

Table S1. Significant regions of regression for the change in tFUS induced SSRT.

S5. Whole-brain SPM analysis

In this analysis, we performed a whole-brain SPM analysis for the (1) positive t-contrast for SS (No-tFUS) – SS (Stop-tFUS) trials to examine where activity was larger for tFUS compared to No-tFUS trials, (2) SS-US main effects, and (3) tFUS main effect. We present

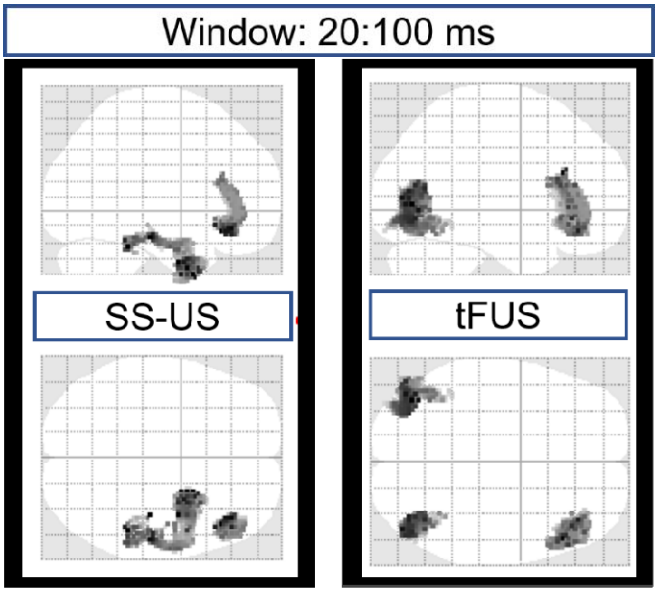
these results in this sequential order. In all of the analysis below, we used a cluster-forming voxel threshold of $p < 0.005$

SS-SS t-contrast. In the stop-signal locked window (-40: 20 ms), we found differences in left supramarginal gyrus. In the N100 window (20:100 ms), the contrast was significant, as anticipated, in rIFG. This result supports our scalp ERP measures and previous results indicating that an N100 occurs in the rIFG that is predictive of stop success. In the N200 window (100:180 ms) we found increased source activity in tFUS trials in bilateral clusters with locations indicative of pre-SMA (Mayka et al., 2006). Consideration of the fourth time window (180:260 ms), which overlapped with the SSRT, revealed that only the right inferior occipital area produced a larger response in tFUS trials. In the last window (260:340 ms), we found clusters of activity larger in tFUS trials for left superior occipital, right cuneus, and anterior cingulate. Because this window always occurred after the SSRT for both No-tFUS and tFUS conditions, it is likely these changes in activity represent a component of performance monitoring rather than inhibition *per se*. Examination of the contrast for larger No-tFUS SS trial activity revealed this contrast was only significant in the latest time-window (260:340 ms). This contrast indicated source ERP activity was larger in No-tFUS trials in bilateral postcentral gyrus clusters. An important result from these contrasts is the primary areas in which activity was larger during tFUS trials. We found that both rIFG and pre-SMA activity coincided with increased tFUS-related stopping. Interestingly, this analysis confirmed that rIFG differences occurred before those in pre-SMA. However, a temporal precedent of change in ERP does not necessitate that stopping-related changes occurred in rIFG before pre-SMA. The source cluster MNI locations and cluster sizes are presented in Table S2.

SS (No-tFUS) – SS (Stop-locked tFUS)					
Time-Window (ms)	t-contrast Direction	Region	Peak Coordinates (X,Y,Z MNI)	Z-value	Extent (Voxels)
20:100	tFUS > No-tFUS	Right Inferior Frontal	44, 28, 5	2.8	979
100:180	tFUS > No-tFUS	L pre-SMA	-10, -10, 50	2.7	1108
		R pre-SMA	6, 22, 50	2.67	1069
		R DLPFC	30, 28 40	2.91	850
180:260	tFUS > No-tFUS	Right Inferior Occipital	50, -70, 8	2.62	47
260:340	tFUS > No-tFUS	Left Superior Occipital	-18, -82, 34	2.34	594
		Right Cuneus	18, -70, 32	2.31	259
		Anterior Cingulate	8, 6, 24	2.01	253
	tFUS < No-tFUS	Left Postcentral	-18, -38, 68	2.02	578
		Right Postcentral	26, -24, 70	1.99	362
P. Uncorrected Voxel < 0.005 and P. Cluster Uncorrected < 0.05					

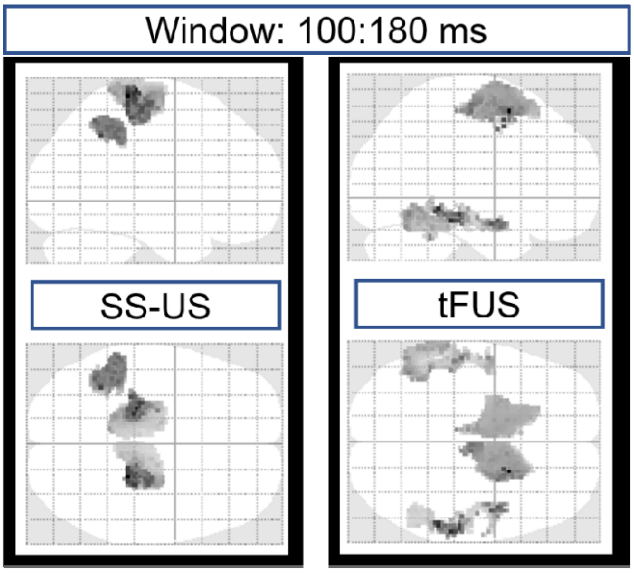
Table S2. MNI locations for SS-SS contrasts

SS-US and tFUS F-contrast. Here we show brain areas that exhibited differential evoked activation according to (1) successful compared to unsuccessful stopping (SS – US F-contrast), and (2) which areas were modulated by tFUS (tFUS F-contrast). We computed these whole-brain SPM contrasts using a flexible factorial model to implement a repeated-measures ANOVA. We examined these contrasts over 3 time windows that covered the time from stop-signal onset till after the range of subject SSRTs. The main result from all of these analyses is that the only brain area exhibiting overlap between a significant SS-US and tFUS contrast was right inferior frontal gyrus during the 20:100 ms time window. This result supports the accuracy of our tFUS manipulation and its effects on inhibition performance. These plots and tables are in Figures S3-5 below.



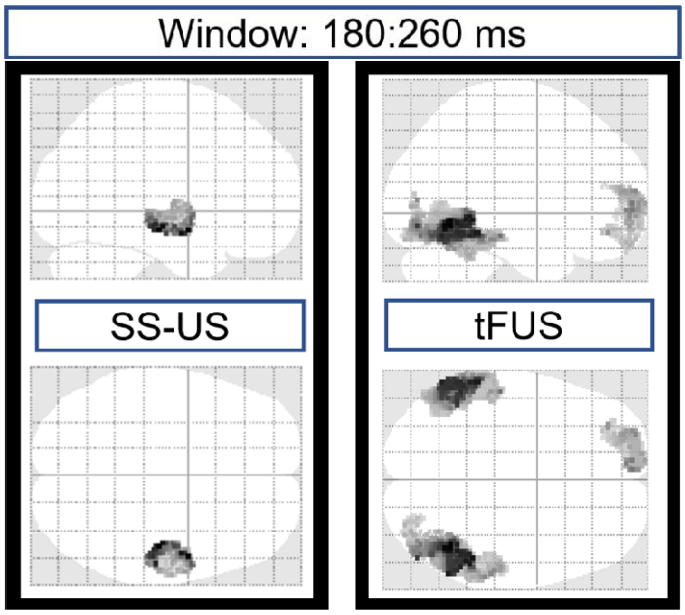
FWE *P*	Cluster Extant	Uncorr. *P*	Peak-F	X	Y	Z	Cluster Peak
F-contrast: Successful Stop – Unsuccessful Stop (SS-US)							
<0.005	893	0.0001	16.35	50	-40	-26	R Inferior Temporal
<0.05	586	0.0001	15.16	54	26	18	R Inferior Frontal
F-contrast: non-tFUS – stop-tFUS (tFUS)							
<0.01	620	0.0001	16.65	-42	-72	-8	L Inferior Occipital
<0.005	899	0.0001	16.09	52	18	22	R Inferior Frontal:
<0.01	699	0.0001	15.52	48	-74	-10	R Inferior Occipital

Figure S3.



FWE *P*	Cluster Extant	Uncorr. *P*	Peak-F	X	Y	Z	Cluster Peak
F-contrast: Successful Stop – Unsuccessful Stop (SS-US)							
<0.001	1147	<0.0001	24.87	24	-34	68	R Postcentral
<0.0005	1279	<0.0001	21.85	-22	-28	72	L Pre/Postcentral
<0.05	479	<0.005	19.23	-36	-52	42	L Inferior Parietal
F-contrast: non-tFUS – stop-tFUS (tFUS)							
<0.0001	1337	<0.0001	18.41	22	8	56	R pre-SMA
=0.2	343	<0.01	16.15	52	18	22	R Middle Temporal
<0.05	590	<0.001	13.18	48	-74	-10	L Middle Temporal
<0.001	1113	<0.0001	11.99	-6	-28	56	L Precentral

Figure S4.



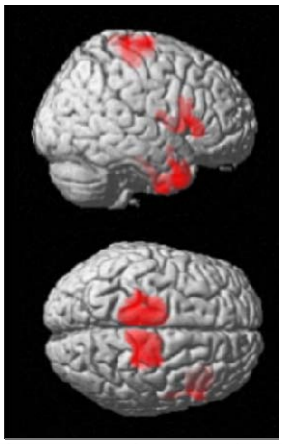
FWE *P*	Cluster Extant	Uncorr. *P*	Peak-F	X	Y	Z	Cluster Peak
F-contrast: Successful Stop – Unsuccessful Stop (SS-US)							
<0.05	497	<0.005	16.83	46	-20	-14	R Middle Temporal
F-contrast: non-tFUS – stop-tFUS (tFUS)							
<0.0001	1702	<0.00001	19.64	42	-58	-10	R Inferior Temporal
<0.001	1270	<0.0001	17.92	-58	-40	-14	L Middle Temporal
<0.05	564	<0.005	13.49	-4	60	0	L Medial Superior Frontal

Figure S5.

Interaction t-contrasts. The most common contrast in analyzing stop-signal neural data involves comparing successful to unsuccessful stop activation (SS – US) as we did for the above whole-brain analysis. The rationale is based on the assumption that areas directly related to stopping/inhibition are more 'potently' active in successful trials, and that failed stop trials (US) also reflect go activity according to the independent race model of Logan and Cowan (1984). This choice of contrast also stems from the fact that typical stop-signal tasks do not offer a second set of SS trials for comparison (which our experiment does). Therefore, this raises the following question: Did tFUS yield changes in stopping by merely altering what would have been US Stop-tFUS trials? If tFUS merely raised the overall level of activity across all stop trials, (i.e.,

SS and US), then we should expect no interactions. We examined the whole-brain SPM interaction across the four time windows used in the previous analysis. In each window, we examined the t-contrast interaction that compared for a bigger SS-US difference in either Stop-tFUS or No-tFUS trials. We only found these effects for the 100:180 ms and 180:260 ms window. The interaction t-contrast SPMs and corresponding tables are below in Figures S6-7.

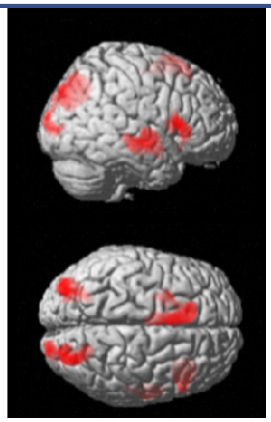
Window: 100:180 ms



FWE *P*	Cluster Extent	Uncorr. *P*	Peak-T	X	Y	Z	Cluster Peak
F-Contrast: Interaction SS-US (stop-tfUS) > SS-US (non-tFUS)							
<0.001	1364	<0.0001	4.42	52	-18	-12	L Postcentral
<0.005	1153	<0.00001	4.41	42	-90	4	R Middle Occipital
<0.005	988	<0.00001	4.30	32	-4	38	R Inferior Temporal
<0.05	670	<0.005	3.70	44	14	10	R Frontal Operculum /Inferior Frontal /Pars Opercularis

Figure S6.

Window: 180:260 ms



FWE *P*	Cluster Extent	Uncorr. *P*	Peak-T	X	Y	Z	Cluster Peak
F-Contrast: Interaction SS-US (stop-tfUS) > SS-US (non-tFUS)							
<0.01	706	<0.001	4.21	52	-18	-12	R Middle Temporal
<0.005	984	<0.00001	3.90	42	-90	4	R Middle Occipital
<0.05	567	<0.0001	3.73	46	16	4	R Inferior Frontal Gyrus / Pars Opercularis
<0.005	822	<0.001	3.61	-2	14	64	pre-SMA/SMA
<0.005	780	<0.005	3.25	-34	-68	28	L Middle Occipital

Figure S7.

We note that we only found significant contrasts in the latter two time windows, i.e., 100-180 and 180-260 ms from the Go signal. In both contrasts, the SS-US difference was larger in the Stop-tFUS conditions. Generally, these results agree with that of the main text examining just the SS-SS contrast and other studies (Aron and Poldrack, 2006; Boehler et al., 2010). However, as Boehler et al. (2010) point out, this contrast is conservative which lends itself to identifying areas primarily involved in successful stopping. It is also unlikely that this conservative contras can identify areas involved in the broader stopping network. It is therefore not surprising that we did not find difference in parietal cortices, for example. Importantly, though, these results demonstrating mainly activation differences in inferior frontal and occipital cortices is almost identical to that found in Boehler et al. (2010). This indicates that our results generally agree with previous studies using the SS – US contrast.

ROI Interaction analysis

In the main text we presented a time-based source analysis of the main ROIs of interest, including Right IFG, Right pre-SMA, Right DLPFC, Left M1, and Right Inferior Temporal cortices. In the main text's analysis, we compared SS trials across tFUS conditions. To address the change in time-course activation at the ROI level, we computed the interaction of SS-US (no-tFUS) and SS-US (Stop-tFUS) trials using the same MNI locations as the main text (Fig. S6).

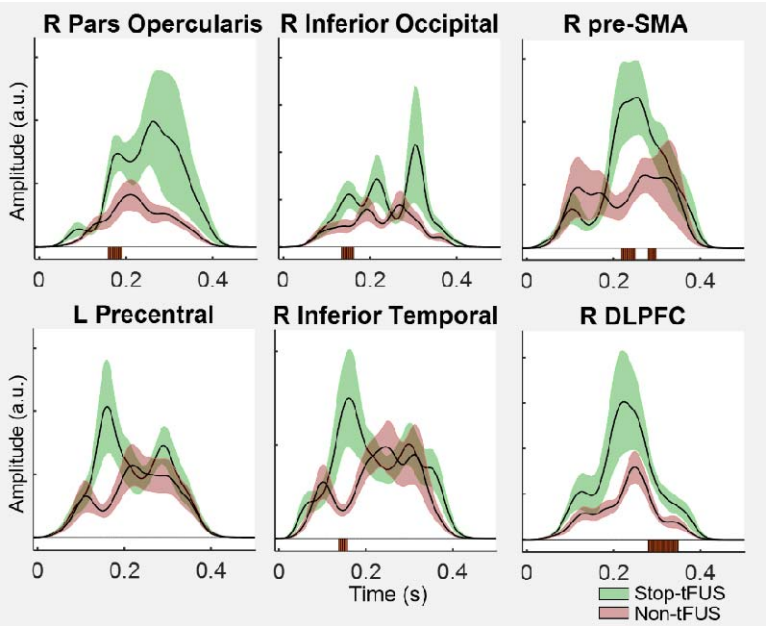


Figure S8. Source-base evoked power time-series for different regions-of-interst listed in the figure. Each ROI was extracted with a radius of 6mm.

Only three of these ROIs are different before the minimum SSRT across tFUS conditions (185 ms). These include pars opercularis, Right Inferior Occipital, and Right inferior temporal ROIs. This result is mostly in line fMRI (Boehler et al., 2010; Li et al., 2006) and EEG/MEG (Boehler et al., 2009; Wessel and Aron, 2015) studies. Although previous work has also identified Right

temporal cortex (Rae et al., 2015; Xu et al., 2017). However, previous findings of differential temporal cortex activity have been based on comparing SS to Go trials rather than the SS – US (and interaction) comparison we used here. Given the DCM and SSRT regression results indicating changes in temporal cortex activity – as well as the temporal cortex's role in visual signal in both the top-down and bottom-up processing directions – these results suggest the success of stopping may rely on information transfer of node between sensory and prefrontal ventral areas, e.g., rIFG. Finally, an interesting result is the lack of difference in r pre-SMA before SSRT. Notably, there was differential activation but only after SSRT. This result speaks to the broader debate of whether r pre-SMA or rIFG lead to stopping in a serial process fashion (Obeso et al., 2013; Watanabe et al., 2015). For example, some studies have suggested that information for stopping passes through r pre-SMA onto rIFG and vice versa in other accounts. These serial accounts, though, forego three factors. First, visual information regarding stop contexts are passed up both the ventral and dorsal pathways which inherently project separately to rIFG and r pre-SMA, respectively. Second, the prefrontal cortex is likely arranged hierarchically with both areas connected with basal ganglia structures, such as the STN and striatum. The third comparison is based on considering the current EEG-DCM (main test) results to those from other fMRI-DCM and other connectivity results (Jahfari et al., 2017) on inhibition. Specifically, the previous bilinear DCM models used in fMRI accounts of response inhibition are unlikely to capture the fast-timescale processes that underlies stopping processes. Along this same line, though, when considering connectivity as a property of inhibition processes, caution should also be used in interpreting temporal precedence of control between brain areas based solely on activation. Thus, the above results indicate that the set of expected ROIs were differentially modulated by tFUS during SS stopping.

Control analysis examining possible auditory effects of TFUS

Recent work examining the effects of TFUS on cortical activity has employed animal models to consider the possibility that TFUS may alter activity in auditory pathways. Using a single-element transducer and optical imaging on a mouse model, Guo et al. (2018) showed that, regardless of the transducer placement and target, TFUS caused activation of the auditory pathway. They suggested that this activity may spread cortically and induce artifactual effects of TFUS in cortical areas not directly targeted by TFUS. It remains unknown how auditory pathway activation via TFUS would yield our behavioral and neural effects. Nevertheless, it is critically important to quantify the extent to which auditory pathway activation might have affected our results. To address this question, we analyzed source-localized evoked results across the rIFG group, as well as rS1 and rIFG auditory (sham rIFG) control groups. Because our goal was to determine whether tFUS altered the evoked activity with respect to No-tFUS, we compared the time courses of source power for stop-locked data by using a source-based ROI of right auditory cortex. We used the source location of X: 46, Y: -14, Z: 8 for right auditory cortex, which was obtained from Rademacher et al. (2001). These locations were used to extract the eigenvariate time-course after source localization in a sphere with radius of 8 mm to be conservative. These source-time courses were converted to a pseudo activation using the (exp(SOURCE)+exp(-SOURCE))/2 transform. We used this procedure to ensure we could properly detect differences regardless of the ERP activity sign. These time series are shown below in Figure. S7

If the behavioral results in the rIFG group were merely the result of changes in auditory pathway activity, we should at least see a difference when contrasting the SS-US trials for No-tFUS and Stop-tFUS conditions in the rIFG group, or at least across groups. Paired samples t-tests did not reveal any significant differences (after false discovery rate correction of $p < 0.05$) when comparing the time courses of the No-tFUS and Stop-tFUS conditions (Figure S7). Therefore, given the spatiotemporal fidelity of EEG, we conclude that our results were not contaminated by tFUS artifacts activating the auditory pathway.

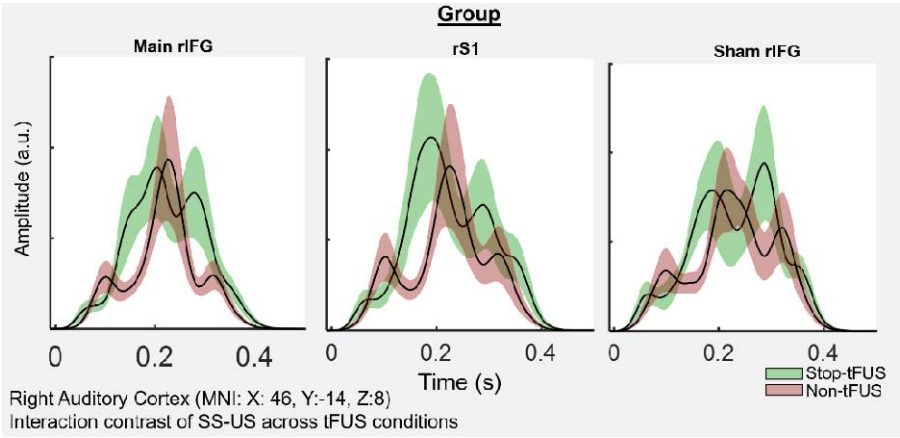


Figure S7. Group source-evoked power time series from right auditory cortex.

Overall, these results suggest that, although TFUS may exert an auditory pathway effects measurable in mouse single neurons and LFPs, (1) this activity may not be measurable at the macroscopic level, and (2) support the notion that changes in auditory cortical activity cannot account for our TFUS-related neural and behavioral effects. Additionally, we note that Guo et al. (2018) found that the auditory pathway activation was accompanied by startle-like reflexes. Our behavioral results (Figure 2) are not compatible with a startle release reflex. If that were the case, tFUS auditory-related startle activity would likely predict shorter RTs during Go trials and failed inhibition Stop trials. The effects of tFUS on our behavioral responseses are not compatible with the involvement of startle reflexes, as we found no tFUS effects on the Go RTs. Therefore, we conclude that the effects of tFUS on stopping behavior, nor processing by the cortical and subcortical nodes of the inhibition network, were not induced by artifactual stimulation of the auditory pathway.

S6. DCM family model hypothesis spaces and results

Below we show the resultant family model posterior probabilities for the family comparisons over the 4 factors (3 levels each) for both parametric empirical Bayesian (PEB) GLMs (Figure S8-9).

PEB GLM model 1 examined the changes in mean connectivity from No-tFUS US to SS trials. Changes in mean connectivity represent the gain on connectivity to represent the No-tFUS SS trials. PEB GLM model 2 examined the effects of tFUS-induced changes in mean connectivity and changes in connectivity that accompanied change in SSRT across subjects. Figures S8 and S9 show the changes in connections and family model probabilities for PEB model 1. Figure S10 tested the exact same factor space and the results are graphically shown as bar plots to represent the family probability of each model separately for the tFUS-induced mean and SSRT change in successful inhibition connectivity.

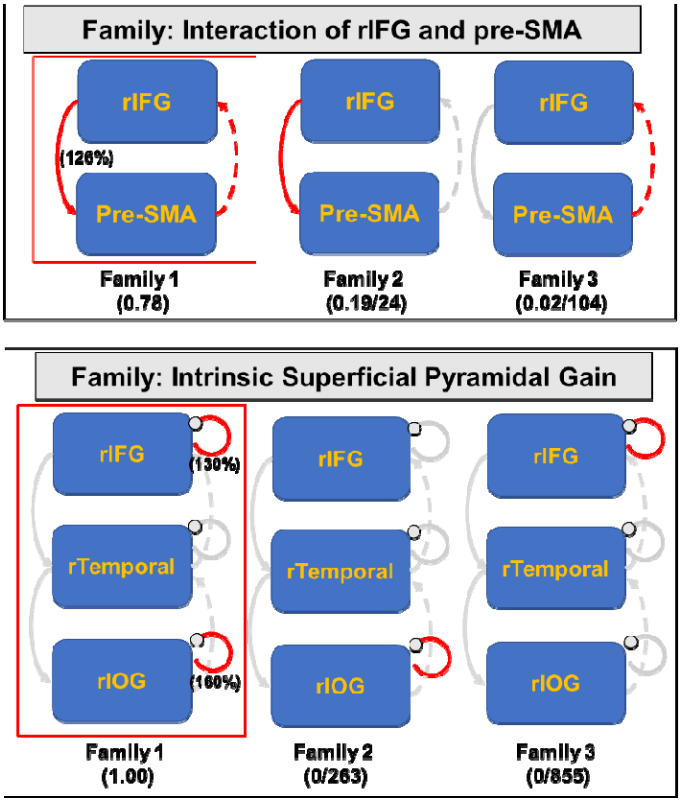


Figure S8. Hypothesis space and results of the Family based PEB Bayesian model comparison for different hypotheses. Family-based posterior probability (F-Pp) and the log-bayes factor with respect to the winning model are listed below each model and were computed as the difference in free energy between families (F-Pp/Log-Bayes). The winning model family is enclosed in a red box. The outcomes of these plots can be interpreted of as revealing the modulatory parameters connections with very strong, positive evidence being different between US and SS trials. Parameters estimates with a greater than 95% posterior probability in these families are presented in parentheses next to the modulated connection. The parameters are presented in exponential form of percentage gain. Values above 100% equates to a parameter increase in

SS trials compared to US (and the opposite for values below 100%). Parameters in-active in each model are in a gray color. The top Panel shows the hypothesis test of rIFG and pre-Sma interactions. The bottom panel shows the hypothesis test of intrinsic gains modulation.

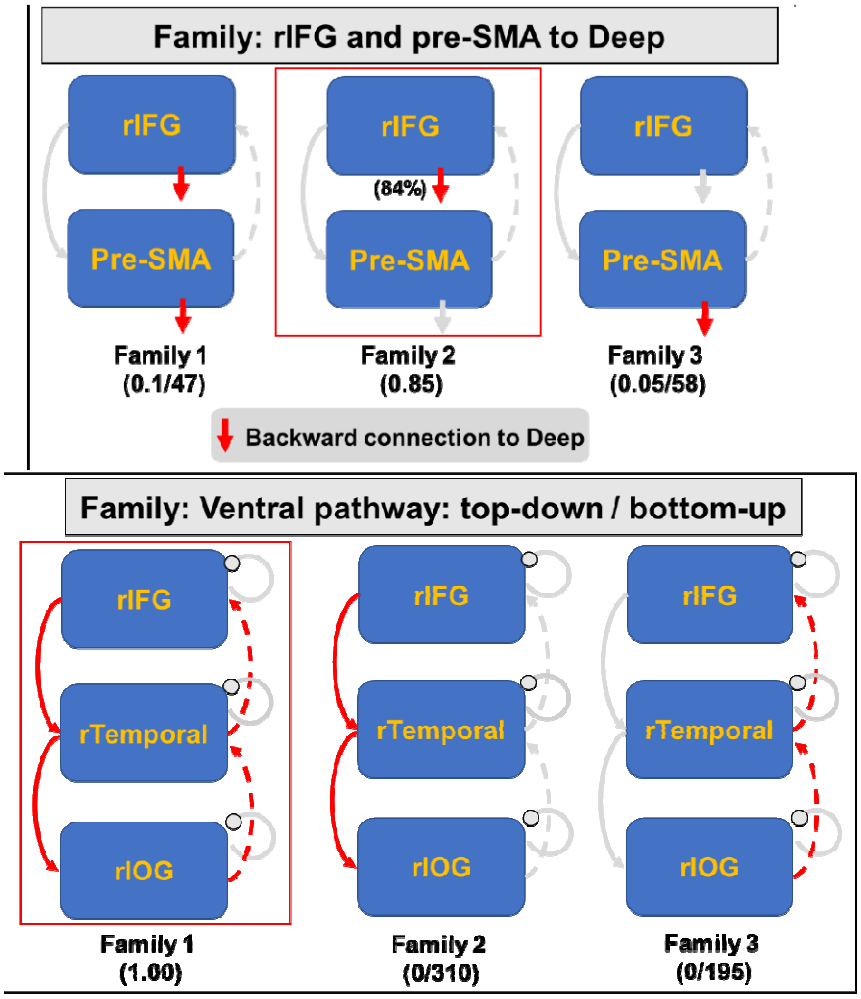


Figure S9. Hypothesis space and results of the Family based PEB Bayesian model comparison for 3 different families comparing hypothetical different interactions between rIFG and the deep node, pre-SMA and the deep node, or both interacting with the deep pathway. The plot is in the same format as Figure S8. Both of these nodes had backward, inhibitory connections with the deep pathway. The top panel compares families comparing the rIFG and pre-SMA to deep backwards connection. The bottom panel compares families testing for differences in top-down v bottom-up connections along the ventral pathway.

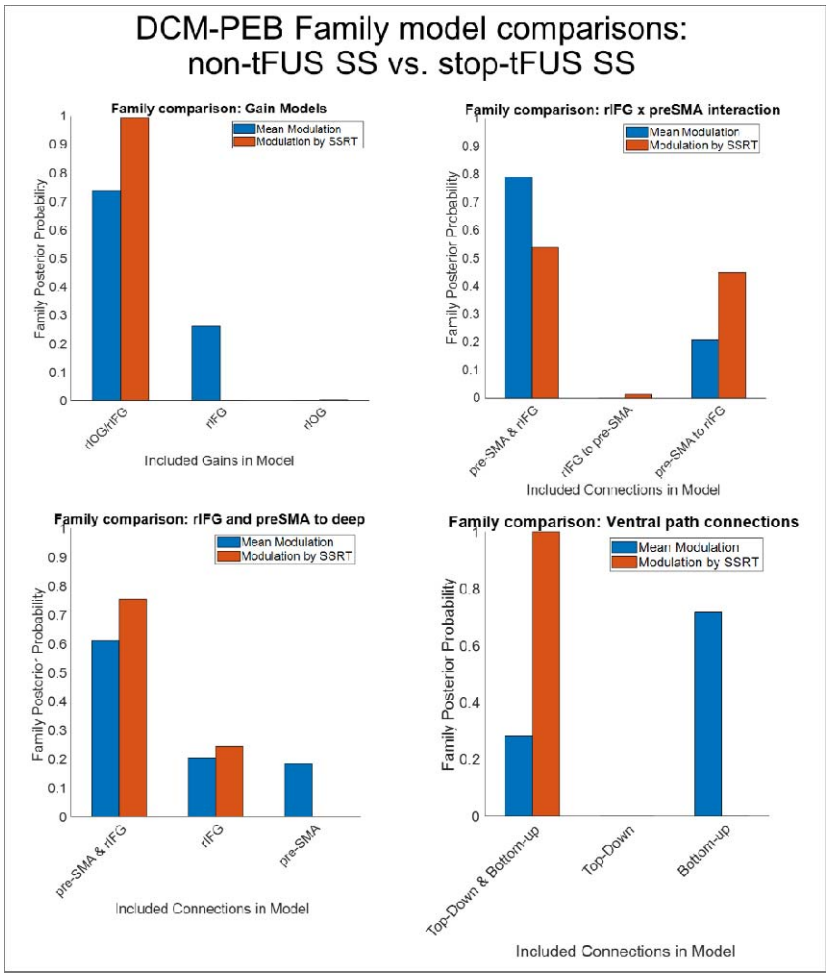


Figure S10. Resultant family posterior probabilities for the comparisons of each factor. Each bar plot shows the marginal probability of each family marginalized separately for changes in the mean connectivity (blue bars) and connectivity changes predicted by tFUS induced changes in SSRT (orange bars). The Bayesian model averaged parameters per connection are presented in the main text.