nature portfolio

Peer Review File

Role of Thalamus in Human Conscious Perception Revealed by Low-Intensity Focused Ultrasound Neuromodulation

Corresponding Author: Dr Zirui Huang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this manuscript Jang and colleagues s sed ultrasound to investigate the role of thalamus in conscious perception. Overall the authors find thalamic sonication to have target specific and target independent effects on behavioral response to a threshold visual stimulus. I think the idea in the manuscript seems interesting but as explain below the statistical analysis of the LIFU data, at least as presented, makes very little sense to me which prevents me from being able to tell whether any of the conclusions in the manuscript are actually warranted. I'm perfectly open to the possibility that it is just not well described, and if it were better described it would make perfect sense, but, as is, I do not understand what was done and the little I see in the text/tables makes little sense to me in terms of its logic (e.g., the proliferation of follow-up tests regardless of significance of main effects or interactions in the initial analysis with, for the most part little correction for multiplicity). I try to explain below best I can.

Specific feedback:

As I said above, my main issue at this time is that I am not understanding the statistical approach, in multiple respects.

First, I'm just not sure what exactly was done when they say a "linear mixed effects ANOVA." Is this a repeated measures ANOVA with Target as the within factor and DC as the between subjects factor? Or did they do a mixed-effects linear model, and (I assume, since nothing is said about this..) they used subjects and target as random effects and DC as fixed effect? Since I just couldn't figure out what they did, in frustration I tried seeing if I could just reconstruct it form the data provided in the supplement (which is wonderful that they shared), but when I tried either approach I got completely different numbers. To be clear, I am in no way saying the authors did anything wrong, I'm just saying I have no idea what statistical analysis was carried out. There are about 3 lines (624-626) in which they describe the key analysis of the LIFU effects, but there are no details. This just needs to be explained more. Simply put, a methods section must allow a reader to be able to replicate an analysis – this methods section does not meet this bar.

Second, even conceding that the analysis presented is correct – whichever method was used – and taking the stated values in the manuscript as correct, I do not understand the logic that leads to the proliferation of tests (most uncorrected) even where no main effects or interactions were found. A conventional analysis would do something like an ANOVA first (as done here) and then would follow up any significant main effect and/or interaction with a post-hoc test. Here, instead, all effects, for all dependent variables, whether found significant or not in the ANOVA still get full post-hoc analysis. Why? For example, False Alarms, Decision Bias, and Accuracy for scrambled images show no main effect of DC, no main effect of region, and no significant interaction in the main ANOVA. Yet, they all still get re-analyzed with 64 Wilcoxon signed-rank tests (despite the non-significant interaction term; in fig 3), a 2-sided Mann-Whitney collapsing across regions (despite the non-significant main effect of DC in the main ANOVA; in Fig 4) and even this latter test gets followed up with its own Wilcoxon tests (again in Fig 4, despite the non-significant effect of DC). The vast majority of these tests are not justified by the analysis.

Similarly, the authors present the "region dependent" and "region independent" effects as separate sections and analysis, but I'm not sure they are. Doesn't a main effect of DC in the main ANOVA, in the absence of an interaction with Target, already imply a "region independent" effect? Why did the authors re-run the main effect of DC, in isolation, with the Mann-Whitney tests performed on all measures regardless of the ANOVA finding?

Further compounding the issue, most of the dependent measures are highly correlated, isn't this further inflating type I error given the separate/independent ANOVAs done here?

Third, there are some important confound variables that the authors collected (sex and whether the participant heard noise -- which is a pretty big topic in the field), shouldn't these also be entered as confound variables in the main ANOVA analysis?

As an additional point, looking at the data presented in the supplement (in the statistics tab) I am not sure I understand the degrees of freedom. I agree with the df(numerator), which should indeed be 1 for the intercept and DC, and 3 for target and the interaction, but how are the degrees of freedom of the error (I'm assuming that's what df2 is?) always 201 across all effects? Wouldn't the intercept and DC have a different df(error) compared to block/target and the interaction?

Also, what software was used to analyze the data?

Last, I do have one question regarding the experimental design: LIFU is administered in 30s blocks ON and OFF, which has been done by many others, but I'm not clear as to whether the task was performed only during the ON periods or during the OFF periods too? If the latter, is there any difference between the ON and OFF periods?

In short, at present I just have a hard time understanding what has been done. Without radically more clarity in the analysis, it is very hard to evaluate whether any of the conclusions here are justified.

Reviewer #2

(Remarks to the Author)
This study sought to elucidate how LIFU to different thalamic regions affects conscious perception. There are serious issues regarding the LIFU aspects of the study and the ability of this device to reach the intended target(s). More detail on modelling procedures for each individual is necessary. To this point, each target for each participant received a different amount of energy and this was not accounted for and thus no conclusion on the intervention between conditions can be made as is. Also, the distinction between 5% and 70% duty cycles is not well justified and unclear how and when these two energy levels were delivered. There is also some concern for potential carry-over effect of delivering multiple sessions of LIFU to each intended target within the same session. As such, these issues and points below need to be addressed before acceptance can be considered.
Abstract
Please include something about the connectivity data.
Introduction
Lines 73-74: Please change to “hypothesized to lead to …” and you could add the following reference. The influence of duty cycle for inhibition/excitation is not established.
Lines 74-75: Acknowledgement of previous human sub-cortical LIFU applications should be mentioned here especially including previous thalamus work.
Nakajima, K., Osada, T., Ogawa, A., Tanaka, M., Oka, S., Kamagata, K., ... & Konishi, S. (2022). A causal role of anterior prefrontal-putamen circuit for response inhibition revealed by transcranial ultrasound stimulation in humans. Cell Reports, 40(7).
Legon, W., Ai, L., Bansal, P., & Mueller, J. K. (2018). Neuromodulation with single-element transcranial focused ultrasound in human thalamus. Human brain mapping, 39(5), 1995-2006.
More justification or setup in the introduction regarding hypotheses regarding the chosen duty cycles should be offered.
Line 81. Replace stimulate with modulate
There is nothing in the introduction regarding the thalamic connectivity portion of the experiments that looks to be used to help explain differences in experimental results? More on this is needed.
Methods
Line 436-437: How were skull effects considered?
Line 440: This is not clear. Would not changing the duty cycle also change the pulse width? Please also use pulse duration instead of pulse width.
Lines 441-442: Please also provide the Isppa and are these values for outside or inside the head? Please also provide the incident pressure in kPa and estimated pressure at the focal zone in kPa.
Lines 460-481: I do not understand the purpose of these procedures. Please elaborate.
Lines 482 – 489: Please provide images of these models for each of the targeted thalamic regions in transverse or coronal and sagittal views either in supplemental or manuscript figure. Please identify how the skull was identified and acoustic properties assigned.
The purpose of the connectivity data is not clear.
Results
Figure 1 is well done and informative. Line 133 is not clear to me nor are the methods in this case. Was 5% and 70% duty interleaved within each of the LIFU ON blocks?
Lines 146-151: This section is unclear. You overlaid a beam profile (collected in free water or modelled? And modelled in free water?) on an MNI template brain? This is somewhat disingenuous as this would be a ‘stylized beam’ in an ideal circumstance and not indicative of that actual beam in an individual head. Please make this clear in the figure caption.
Lines 152-156: This is a highly unorthodox method. Do you not have the modelled beam for each participant? And did you collect MRIs for each participant? The employed method of confirming deviation is not acceptable. Please use the modelling data from each participant and do within subject estimates of beam relative to the target.
Lines 157 – 164: It is not clear what this post-hoc analysis was or how it was done. Please define focal center. That is a massive estimated intensity loss and does not seem to converge with current research. Please explain/define how these

methods were done.
Figure 2. A. Please explicitly state that this is not an actual acoustic model from an individual but a 'stylized beam' overlaid (please define in methods how this was properly scaled) on the template brain to explicitly only show/illlustrate the size of the beam (empirical free water? Modelled free water?) relative to the thalamus.
B. Please quantify or describe what beam center is. Pressure maximum?
K. I see that some individuals registered zeros is that correct? Why? And how was this data dealt with?
Because a standardized intensity was delivered outside the head this results in varying intensity in the head for each participant for each thalamic target. This variability must be accounted for as a covariate in all statistical models before any rational conclusions can be made.
Please provide size and depth of each thalamic target for each individual also based upon sex. Please also provide how (given the variability in depth of the targets) this was compensated for with the current transducer. Was a different offset used for each participant? How was this done?
There needs to be some comment somewhere on the potential carry-over effects of the protocol between the four regions. It would be helpful if some analysis of potential stimulation carry-over or accrual was present.
Discussion
Line 293-294. This statement is disingenuous. How can an effect from this experiment be associated with increased functional connectivity from another data set where there was no manipulation?
Lines 324-327: The references provided do not support this statement for inhibition vs. excitation for the duty cycles used. These statements need to be tempered and because you have no direct evidence of excitation vs inhibition for the results of this study. The 70% condition, in effect, delivered more energy than the 5% condition so could this not be a reason?
Lines 346 – 348: Same as above.
Line 352 - 355: Methodologically, this work does not represent a significant advancement in human LIFU studies as several LIFU studies that have previously be done are not acknowledged or discussed in particular a seminal study in human thalamus by Legon et al. 2018.
Line 375: Please also discuss these energy levels relative to other human LIFU studies. These Ispta values are extremely low.

Reviewer #3

(Remarks to the Author)
The study reveals that stimulating specific thalamic regions with low-intensity focused ultrasound (LIFU) affects conscious perception in distinct ways. Stimulation of the ventral anterior (VA) thalamus with a high duty cycle (70%) significantly enhanced sensitivity in object recognition, suggesting an excitatory effect that boosts perceptual processing in this transmodal-dominant area. Conversely, a low duty cycle (5%) induced a conservative decision bias, indicating a potential inhibitory effect on perception. Additionally, categorization accuracy for real images decreased under high duty cycle stimulation, possibly due to interactions with the visual network. These findings highlight both region-specific and general effects of LIFU on perception, offering new insights into the thalamus's role in conscious processing and supporting the potential of LIFU as a non-invasive tool for targeted brain modulation.

If the same Ispta was used across different duty cycle conditions, this would mean that a higher Isppa was applied under the 5% duty cycle condition. If so, it would be helpful to state this explicitly in the text. Additionally, in Fig. 1b, it would be clearer to represent the 5% duty cycle condition with a higher amplitude in the illustration to reflect this difference. This difference in amplitude could also explain the greater occurrence of auditory confounding artifacts observed at the 5% duty cycle, as the higher instantaneous pressure may have produced more noticeable sounds.

In Fig. 2a, the beam profile appears to be simulated, and it seems that there is no actual acoustic beam profile measured with a hydrophone or similar device. Including this measured data would strengthen the study by providing empirical validation of the beam's characteristics, such as focal precision and intensity distribution, especially through the human skull.

While the actual sonication was directed from the temporal lobe, Fig. 2b-e represents the acoustic focus as a central point, which may not fully capture the anatomical accuracy of targeting relative to the sonication direction. It would be beneficial to consider ways to improve this representation to better reflect the actual targeting accuracy based on the sonication approach.

While it is noted that "the order of stimulated regions was pseudo-randomized," it would be helpful to include information on how long it took to adjust the single-element transducer between different target locations. This detail would provide additional clarity on the practicality of the experimental setup.

The study lacks data on acoustic attenuation specific to the experimental setup, such as simulations or measurements. Including this information would enhance the understanding of ultrasound energy transmission through the skull and its impact on targeting accuracy.

To enhance the reporting on ultrasound stimulation conditions and improve alignment with safety standards, it would be beneficial to refer to Martin et al. (2024), ITRUSST Consensus on Standardised Reporting for Transcranial Ultrasound Stimulation and Aubry et al. (2023), ITRUSST Consensus on Biophysical Safety for Transcranial Ultrasonic Stimulation. Including these guidelines would help provide a more comprehensive report on stimulation parameters and reinforce the study's adherence to established safety standards.

It would be helpful to provide additional explanation regarding the statement, “The focal depth was also validated by Blatek Ultrasonic Transducers (Boalsburg, PA, USA).” Further details on the validation process, including methods or measurements used to confirm the focal depth, would clarify this aspect of the setup and enhance the reproducibility of the study.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I thank the authors for having put sincere effort into the response and I do think the methods are now a little clearer. Nonetheless, I still have some substantial concern regarding both the statistical analysis and the fact that I don't think the analysis and results support the key conclusions.

Methodological/statistical concerns:

1. First, I still have a concern relating to multiple comparisons. I'll start by saying that the authors now apply appropriate FDR correction to post-hoc analyses (I assume they used a Benjamini–Yekutieli procedure, which controls the FDR under highly correlated tests, as in the present case? Which FDR procedure is not said in the methods). Where I think an adjustment for multiplicity might still be needed is the fact that the exact same data are being analyzed twice: once in the region-specific analysis and once in the region-invariant analysis. It is exactly the same data going into the two analyses. To explain, consider the following: had any of the findings in the region-specific analysis shown that all/a majority of the LIFU blocks were different from baseline, it would have necessarily implied a significant difference in the region-invariant "baseline vs LIFU" analysis. So these two analyses cannot be independent of eachother. Rather, they seem to be a reanalysis of the same data (unlike, say, had they first looked at baseline vs LIFU, and then compared between the LIFU conditions, that would have been an orthogonal analysis of the data). I understand the authors "consider them" complementary, but still, the issue is just one of re-analyzing multiple times, in multiple different parallel ways, the exact same data.

2. Second, I have a number of concerns relating to the way in which covariates were adjusted for.
(i) The major point here, unless the authors just forgot to say this in the methods, is that the "de-noising" regression followed by a repeat of the Friedman ANOVAs was done with the incorrect degrees of freedom. To explain: In adjusting for these covariates, the authors didn't account (or don't say they did) for the fact that the residuals of the "denoising regression" have fewer degrees of freedom than the original data. Specifically, in a regression the DF of the residuals (i.e., df error) is N-i-1, with N being the number of observations, "i" being the number of confounders in the regression, and "1" for the intercept. Yet, as far as what is in the methods, the Friedman ANOVAs and Wilcoxon t-tests done on the residuals are evaluated at the nominal DF of N observations. Which seems incorrect. (Might be inconsequential, but this is an empirical question.) I know people use this procedure in fMRI all the time, but in that case the inference is then done at a higher level compared to the first-level "denoising" analysis, so the DF relevant to inference in fMRI comes from the number of subjects (either entirely, if in a random effects analysis, or mainly, if doing a mixed effects analysis) meaning the loss of DFs at the level of individual timeseries (i.e., volumes) is almost irrelevant. Here, instead the inference is done at the same level as the denoising regression, thus DFs should be appropriately accounted for. Of course the authors never even report the chi2 statistic of the Friedman ANOVA or the DF, so I'm left to infer.
(ii) Hadn't I had the benefit of reading the rebuttal letter, where this two-step procedure is explained clearly, I don't think I would have been able to understand what they did from the text in the current version of the methods. First they say that "To address potential confounders, we included six nuisance variables in the analysis." This is not true. They removed confounders from the data, using a linear regression (which is said later on) and then analyzed the residuals with the same strategy in the main analysis -- which is buried a few sentences later into this section. A few sentences later they do say ""We adjusted for these covariates using a linear regression to obtain residuals, followed by Friedman's tests on these residuals [...]" but then in the appendix I see they also used this for the Wilcoxon t-test, but that is not written in the methods. All I'm saying is that it would help if the methods described clesarly what was done so one could replicate them.
(iii) Considering the above, since the Friedman ANOVA is not very flexible and doesn't allow including covariates, why not use a linear mixed model, which can accommodate the non-normal data and covariates or a permutation-based ANCOVA (the widely used permuco package in R)? Either of these approaches would allow including the confounders in the main analysis and do away with having to do a separate convoluted analysis with incorrect degrees of freedom.
(iv) Missing data imputation: there is one throw away sentence in the methods that reads "Missing data were imputed using group means or modal categories (for noise awareness), ...". Was the data imputation done in the main analysis too or only in the 'confounder' analysis? What variables were imputed, only the confounders themselves or also the dependent variables (the sensitivity, criterion, etc for each nucleus)? Was this done also in the main analysis?

Other concerns (about results/interpretation/conclusion):

1. With respect to the region-invariant analysis, it is odd that none of the findings in the region-invariant analysis have a correspective in the region-specific analysis. Meaning, neither real images at 70% duty cycle, nor hit rate and criterion at 5% of duty cycle have a correspective significant finding in the more granular region-specific analysis. This must imply that the finding at the region-invariant level are just a side-effect of the reduction of the spread around the median consequent to the

averaging of multiple LIFU blocks (and makes me wonder whether the baseline averaging is having a similar smoothing effect on the region-specific analysis -- meaning, would the one result on sensitivity at 70% duty cycle no longer be there in the absence of baseline averaging?). Now, generally that's exactly why we average, to decrease noise. But here what is being averaged is the result from sonication of nuclei which the authors then go on to conclude are very different because of their cytoarchitectonics and connectivity. It feels like the authors want to have their cake and to eat it too. Either there is a specific role of VA (and, as they conclud, matrix cells/multimodal connectivity) or there is a generalized effect of tFUS on image detection (and then it's not about cytoarchitectonics or connectivity). I know the authors say the two analysis are "complementary" but to me they just appear contradictory (or at least, the significant finding in the baseline vs LIFU just seems completely contrary to the stated conclusion).

2. As mentioned above, what is the effect of baseline averaging? Is the result just a side effect of artificially reducing the variance of the baseline? I think it should be very strightforward for the authors to repeat the Friedman ANOVa (only for d') with the two separate baselines and show that the averaging is inconsequential (that could be a very simple table in the appendix), but at least the reader wouldn't have to be left wondering if the artificial procedure is the reason for the one significant effect upon which much/most of the conclusion is predicated.

3. I don't think the conclusion follows from either the analysis or the results. To explain, the authors want to conclude that there is a "unique role for the matrix-rich transmodal region of thalamus" and that this research "provides a valuable insight into the unimodal transmodal functional organization of thalamus and its relationship with human conscious perception". Later in the discussion, the authors conclude by saying that "these results provide insights into the unimodal transmodal functional organization of the thalamus and its relationship to conscious perception." Yet, that's what the results show, and I'm not even sure that the analysis they did allows for this conclusion. To explain:
(i) With respect to the former, it is hard to draw the conclusion that what matters is matrix rich nuclei that have transmodal connectivity when DA has no effect on visual awareness whatsoever (in fact, it is as different from VA as is the baseline) despite having a similar pattern of cross-modal connectivity to VA and despite also being matrix rich and consciousness-sensitive (as shown in their prior work; reference #20). The only conclusion one can draw from the region specific analysis is that VA only might have an effect on the sensitivity of detection. This, of course, does not match the much more encompassing speculations in the conclusion about cytoarchitectonics and connectivity (which, again, does not fit with the result of DA).
(ii) With respect to the latter, I think the whole region invariant analysis is contrary to the intended conclusion. A significant finding there implies the effect is nonspecific, particularly given that none of the findings in the region invariant analysis then have a correspective to show that there is some kind of specificity in the region specific analysis (meaning, none of the findings in the region invariant analysis are matched by any kind of difference between baseline and any of the nuclei, implying the result is likely just due to averaging and dispersion reduction). To be clear, it's perfectly fine to also have region invariant results, they just don't allow making the conclusion that there's anything specific or any dissociation between the nuclei, all the less that this has anything to do with being matrix rich or having transmodal connectivity. If the authors truly wanted to make this case, they could have looked at whether the average amount of matrix rich cells in each nucleus (from their prior work) or the amount of transmodal versus unimodal connectivity assessed here is the thing that correlates with observing an effect of lifu on the task, or something of the sort. In other words, while I believe that VA has had a differential effect (pending the reanalysis with the two baselines), the link from there to the story about cytoarchitectonics and connectivity is not supported by these data.

4. Similarly, later on in the discussion, the authors say that the nonspecific decrease in category accuracy might be due to the fact that the 70% duty cycle stimulation might have led to consistent co-activation of the visual cortex doing the stimulation. Yet, positive stimulation of DP, which is uniquely connected primarily to visual cortex, didn't have any such effact. In light of this, the interpretation doesn't seem to follow logically from the results. Doesn't the absent finding in DP sonication contradict the point being made here?

Reviewer #2

(Remarks to the Author)
The authors made a strong effort to address most major points raised in the reviews which is much appreciated. Of particular note, the authors thoroughly addressed the statistical rigor concerns and made substantial revisions that appear to resolve most of the issues. The authors adequately addressed issues including the justification and description of the duty cycles and the integration of the connectivity data. The authors somewhat addressed the potential issue of carry-over effects but did not perform more rigorous assessments such as sonication order effect per region which could still be important given potential accumulating or bleeding over effects per region given their small size and close proximity. The authors also partially addressed the targeting and intensity issues. While statistical regression of confounds is a good step there is still a lack of individualized skull modelling which remains a significant limitation of this study and should be explicitly acknowledged. Some other small issues still remain including the manuscript still presents stylized beam data (and only free water data) and not individual specific data which weakens any precise claims; there is still no indication that individual anatomical MR were used which is a major limitation for thalamic targeting and there is still insufficient detail on whether the transducer offset or orientation was individually adjusted for each participants' anatomy. Overall, the authors did a good job to address most major points. However, the biggest unresolved limitation remains the lack of subject-specific beam modeling and skull compensation, which undermines the claim of precise region-specific effects especially for individual thalamic nuclei that are small, deep and in close proximity. This may not invalidate the results entirely, but it weakens causal claims about localization and mechanism.

Reviewer #3

(Remarks to the Author)
While the free-field acoustic beam profile shown in Supplementary Fig. 5 provides useful baseline information, it would be more intuitive and informative to present it as a two-dimensional (2D) axial-lateral intensity map rather than as separate 1D line plots. Additionally, when defining the acoustic focus, it is standard practice to use the full width at half maximum (FWHM) of the intensity distribution, corresponding to the -3 dB boundary where intensity exceeds 50% of the maximum value, rather than the -6 dB (25% intensity) threshold currently applied. According to the ITRUSST practical guide (Murphy et al., 2025, Clinical Neurophysiology), the FWHM of the intensity profile should be used to define focal dimensions in transcranial ultrasonic stimulation studies. Therefore, we suggest reanalyzing and presenting the focal dimensions based on the FWHM criterion to align with established standards. Similarly, in Fig. R11, when visualizing beam-present regions, it would enhance technical rigor and clarity if the focus boundaries were defined and visualized according to the FWHM standard.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I sincerely thank the authors for having patiently gone along with my sometimes "detailed" feedback. This latest revised version is very strong methodologically and in terms of the connect between the specific statistical tests conducted and the conclusions drawn. I think this is work that the authors and the field will be very proud of.

Martin Monti

Reviewer #3

(Remarks to the Author)
While the authors present a well-structured behavioral study, the methodological pipeline for targeting deep brain regions—particularly thalamic subnuclei—has inherent precision limitations that should be acknowledged more explicitly. Specifically, the use of pseudo-anatomical MR images derived from linearly rescaled MNI templates does not account for individual variability in skull morphology and deep brain anatomy. The scalp-surface–based neuronavigation registration (fiducial + ICP) introduces additional alignment errors that may accumulate at subcortical depths. Furthermore, acoustic beam modeling was performed on segmented pseudo-MRI using homogenized skull properties, without subject-specific CT or MRI-based modeling, thus failing to capture individual focal shifts and intensity variations. Given the close proximity (3–6 mm) of thalamic subnuclei and their deep location, these factors—combined with single-focus sonication without adaptive beam correction—can result in cumulative targeting errors of several millimeters. This limits the strength of claims regarding precise region-specific neuromodulatory effects at the individual level.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In this manuscript Jang and colleagues used ultrasound to investigate the role of thalamus in conscious perception. Overall the authors find thalamic sonication to have target specific and target independent effects on behavioral response to a threshold visual stimulus. I think the idea in the manuscript seems interesting but as explain below the statistical analysis of the LIFU data, at least as presented, makes very little sense to me which prevents me from being able to tell whether any of the conclusions in the manuscript are actually warranted. I'm perfectly open to the possibility that it is just not well described, and if it were better described it would make perfect sense, but, as is, I do not understand what was done and the little I see in the text/tables makes little sense to me in terms of its logic (e.g., the proliferation of follow-up tests regardless of significance of main effects or interactions in the initial analysis with, for the most part little correction for multiplicity). I try to explain below best I can.

Response:

We thank the reviewer for a critical review of our manuscript. We appreciate your thorough evaluation and acknowledge your primary concern regarding the clarity and logic of our statistical analyses. You raise valid points about the need for a better explanation of our statistical approaches. We take these concerns very seriously and have substantially revised our statistical methodology and its presentation, as detailed in our point-by-point responses below.

Question 1:

As I said above, my main issue at this time is that I am not understanding the statistical approach, in multiple respects.

First, I'm just not sure what exactly was done when they say a "linear mixed effects ANOVA." Is this a repeated measures ANOVA with Target as the within factor and DC as the between subjects factor? Or did they do a mixed-effects linear model, and (I assume, since nothing is said about this..) they used subjects and target as random effects and DC as fixed effect? Since I just couldn't figure out what they did, in frustration I tried seeing if I could just reconstruct it form the data provided in the supplement (which is wonderful that they shared), but when I tried either approach I got completely different numbers. To be clear, I am in no way saying the authors did anything wrong, I'm just saying I have no idea what statistical analysis was carried out. There are about 3 lines (624-626) in which they describe the key analysis of the LIFU effects, but there are no details. This just needs to be explained more. Simply put, a methods section must allow a reader to be able to replicate an analysis – this methods section does not meet this bar.

Response:

We appreciate the reviewer's concern regarding the clarity and validity of our original statistical analysis. To address this, we substantially revised our statistical pipeline and verified our findings. We provide a detailed explanation of our new methods as below:

1. Normality assessment: We first began by testing the distributions of our perceptual outcomes pooled from the entire subjects and blocks using Anderson-Darling test. This analysis revealed that most outcomes were not normally distributed, except for criterion and categorization accuracy for scrambled images (*p*-values reported in Supplementary Table 3). Based on these results, we opted to use non-parametric tests (i.e., Friedman's ANOVA and Wilcoxon signed-rank tests) throughout our main analysis.

Supplementary Table 3: Anderson-Darling test results. Significances (FDR-corrected $p < 0.05$) are highlighted as red.

Perceptual outcome	Raw *p*-value	FDR-corrected *p*-value
Hit rate	0.0087	0.0131
False alarm rate	0.0005	0.0012
Sensitivity *d'*	0.0006	0.0012
Criterion *c*	0.0694	0.0833
Real image accuracy	0.0005	0.0012
Scrambled image accuracy	0.3142	0.3142

2. Analysis of region-specific effects: Our primary interest is in perceptual outcome changes (e.g., sensitivity and criterion) relative to baseline. Therefore, we chose to include the baseline as an additional condition. After confirming that two baselines do not show significant differences in all outcomes (Supplementary Fig. 3), we merged the two baselines by averaging. We then applied Friedman's test to the five repeated measures (baseline-average plus four thalamic targets) for each outcome. For post-hoc comparisons, we used Wilcoxon signed-rank tests with FDR correction for multiple comparisons. This analysis identified a significant sensitivity *d'* difference between the baseline and VA sonication at 70% DC (see updated Fig. 3 and Fig. R1).

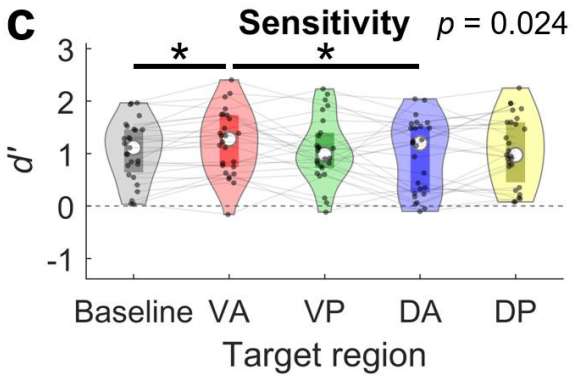


Fig. R1: Region-specific sensitivity values during human thalamic LIFU administration (adopted from Fig. 3c). The *p*-value of Friedman's ANOVA is provided. Asterisks denote significant differences in post-hoc pairwise comparisons (FDR-corrected $p < 0.05$; Wilcoxon signed-rank test, two-sided).

3. Analysis of region-invariant effects: Our initial approach involved pooling baseline-subtracted outcomes and testing against zero with a Wilcoxon signed-rank test. However, we recognized that this method fails to account for subject-specific variability. To improve on this, we averaged perceptual outcomes across the four LIFU-ON blocks. Comparison of the average values of LIFU-OFF (i.e., baseline) vs. LIFU-ON conditions with a Wilcoxon signed-rank test reproduced the results of our previous analysis. We observed a decrease in accuracy for real images at 70% DC and significant changes in hit rate and criterion *c* at 5% DC (see Fig. R2 and updated Figure 4).

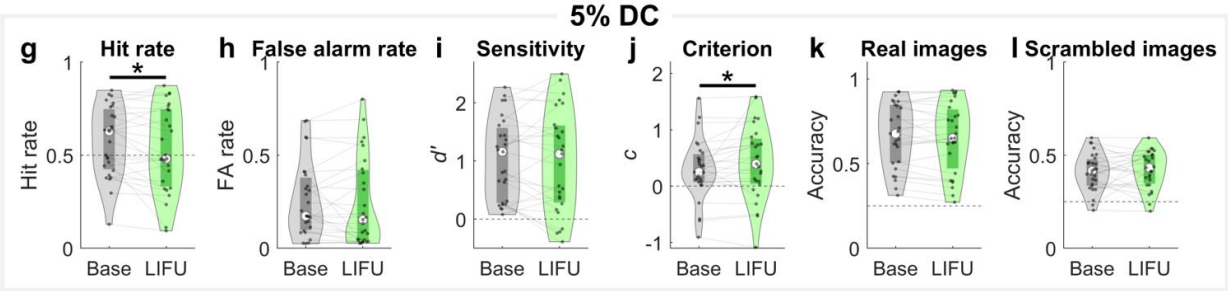


Fig. R2: Region-invariant outcome values during human thalamic LIFU administration (adopted from Fig. 4). Asterisks denote significant differences ($p < 0.05$; Wilcoxon signed-rank test, two-sided).

To ensure full reproducibility, we have attached complete outcome data and MATLAB code for statistical testing as supplementary material. We have also expanded the Methods section to incorporate this detailed specification.

"We initially assessed the normality of each perceptual outcome pooled across the entire dataset using the Anderson-Darling test. Most outcomes did not meet the criteria for normal distribution, except for criterion and categorization accuracy for scrambled images (Supplementary Table 3). Consequently, we utilized non-parametric statistical tests (Friedman's and Wilcoxon signed-rank tests). Our primary analysis investigated region-specific and region-invariant perceptual outcome changes, relative to baseline. Given no significant differences between Baseline-1 and Baseline-2 across all measures (Supplementary Fig. 3), we averaged these into a single baseline measure.

For region-specific analysis, we conducted Friedman's tests using five repeated measures (baseline-average plus four LIFU-ON blocks). Outcomes showing significance in the Friedman's tests were further analyzed through two-sided Wilcoxon signed-rank post-hoc comparisons, with FDR correction for multiple comparisons at α = 0.05. Region-invariant analysis involved averaging perceptual outcomes across the four LIFU-ON blocks per subject and comparing to the baseline averages using two-sided Wilcoxon signed-rank tests." (Line 750, Page 23)

Question 2:

Second, even conceding that the analysis presented is correct – whichever method was used – and taking the stated values in the manuscript as correct, I do not understand the logic that leads to the proliferation of tests (most uncorrected) even where no main effects or interactions were

found. A conventional analysis would do something like an ANOVA first (as done here) and then would follow up any significant main effect and/or interaction with a post-hoc test. Here, instead, all effects, for all dependent variables, whether found significant or not in the ANOVA still get full post-hoc analysis. Why? For example, False Alarms, Decision Bias, and Accuracy for scrambled images show no main effect of DC, no main effect of region, and no significant interaction in the main ANOVA. Yet, they all still get re-analyzed with 64 Wilcoxon signed-rank tests (despite the non-significant interaction term; in fig 3), a 2-sided Mann-Whitney collapsing across regions (despite the non-significant main effect of DC in the main ANOVA; in Fig 4) and even this latter test gets followed up with its own Wilcoxon tests (again in Fig 4, despite the non-significant effect of DC). The vast majority of these tests are not justified by the analysis.

Response:

Thank you for your valuable feedback and suggestions. We have revised our analytical approach to directly address your concern about the proliferation of statistical tests.

In the updated analyses, we have significantly limited the number of statistical tests to ensure rigor and reduce the risk of Type I errors. Specifically, for region-specific analyses, we now begin with an omnibus Friedman's test. Only outcomes showing significance in this initial omnibus test undergo post-hoc pairwise comparisons. Consequently, only sensitivity (*d'*) at 70% DC showed significance, and we have restricted subsequent pairwise testing exclusively to this outcome. No additional post-hoc analyses were conducted for outcomes that did not meet this initial criterion.

Additionally, we conducted a region-invariant analysis to evaluate overall effects across conditions. This region-invariant approach, performed after outcome averaging (comprising only 12 additional tests), provides complementary evidence regarding the presence or absence of overall effects—even when the region-specific analysis fails to detect them. When an outcome is nonsignificant in the region-specific ANOVA but significant in the region-invariant analysis, it may be due to several factors:

1. Increased statistical power through averaging: Averaging across regions can reduce variability and enhance statistical power, thereby detecting subtle overall effects that might be masked when analyzing each region separately.

2. Heterogeneity across regions: Overall effects might exist but could be diluted at the regional level by variability or inconsistent patterns. Averaging helps clarify these overarching patterns.

We believe this revised and targeted analytical framework justifies the modest number of additional tests conducted. Our adjustments aim to enhance analytical clarity, reduce the risk of Type I errors, and promote transparency and replicability of our findings.

Question 3:

Similarly, the authors present the "region dependent" and "region independent" effects as separate sections and analysis, but I'm not sure they are. Doesn't a main effect of DC in the main ANOVA, in the absence of an interaction with Target, already imply a "region independent" effect?

Why did the authors re-run the main effect of DC, in isolation, with the Mann-Whitney tests performed on all measures regardless of the ANOVA finding?

Response:

In our revised analysis, we treat region-specific and region-invariant effects as complementary rather than redundant. In our region-specific analysis, the average baseline is included as an additional condition and Friedman's test is applied to compare the five repeated measures, thereby assessing local, region-by-region effects. In contrast, our region-invariant analysis further averaged the perceptual outcomes across the four LIFU-ON blocks, yielding two measures per subject.

Although omnibus Friedman's ANOVA might imply an overall effect, this was not observed for sensitivity at 70% DC. The region-invariant analysis did not reveal an overall effect after averaging, implicating that the sensitivity increase is confined to VA stimulation. Conversely, for some outcomes, the averaging process uncovered overall differences that were not apparent in the region-specific omnibus test, likely because averaging reduces regional variability and noise.

Question 4:

Further compounding the issue, most of the dependent measures are highly correlated, isn't this further inflating type I error given the separate/independent ANOVAs done here?

Response:

Thank you for highlighting this point. We acknowledge that our outcome measures, such as sensitivity *d'* and criterion *c*, are derived directly from hit and false alarm rates, resulting in inherent correlations among these metrics. To illustrate these interdependencies, we have provided the correlation matrix in Supplementary Table 4 below:

Supplementary Table 4: Spearman correlations among six perceptual outcome measures.

	FA rate	Sensitivity *d'*	Criterion *c*	Real image accuracy	Scrambled image accuracy
Hit rate	0.3591	0.3948	-0.7611	0.5100	0.1539
FA rate		-0.6753	-0.8623	-0.3643	-0.1067
Sensitivity *d'*			0.2434	0.7749	0.2215
Criterion *c*				-0.0384	-0.0017
Real image accuracy					0.4057

Based on this, we suspect that the correlation between hit rate and criterion likely underlies their region-invariant effects co-observed at DC 5%. We now clearly describe this relationship in the revised manuscript.

"Given that criterion is directly derived from hit and false alarm rates (Equation 4) and criterion and hit rate are indeed highly correlated (Spearman correlation = -0.7611; Supplementary Table 4), these two findings likely arise from their inherent interdependence." (Line 269, Page 10)

Please note that our revised analysis strategy has substantially limited the number of tests. Specifically, we now perform 12 Friedman's tests (for region-specific effects; only one significance) and 22 Wilcoxon tests (10 for the post-hoc tests of 70% DC sensitivity and 12 for the region-invariant analysis; five significances overall). Also, FDR correction has been applied to the post-hoc tests.

We believe that our explicit discussion of the inherent correlations among outcomes and newly streamlined testing strategy would address the reviewer's concerns.

Question 5:

Third, there are some important confound variables that the authors collected (sex and whether the participant heard noise -- which is a pretty big topic in the field), shouldn't these also be entered as confound variables in the main ANOVA analysis?

Response:

We thank the reviewer for highlighting the importance of including relevant confound variables such as sex and noise awareness in our analyses. In the updated analysis, we not only accounted for sex and noise awareness but also included lateral deviation, age, sonication order, and tissue-induced attenuation ratio as confounds. Given that our primary region-specific analysis relies on Friedman's test (which does not directly accommodate covariates), we addressed the reviewer's suggestion through a two-step approach:

1. Handling missing data: Missing values were imputed using mean or the most common category (in the case of noise awareness). For the baseline condition, where lateral deviation and estimated intensity were undefined, we used the mean values across all LIFU-ON conditions.

2. Confound regression: We regressed out the confounding variables (sex, noise awareness, lateral deviation, age, stimulation order, and tissue attenuation) from our outcome measures, using a standard linear regression model. Subsequently, we performed Friedman's test on the residuals to determine whether significant differences remained after accounting for these confounds.

The results for confound-adjusted region-specific analysis are summarized in Supplementary Table 5 (see below).

Supplementary Table 5: Region-specific analysis after confounder regression. Post-hoc pairwise comparison was conducted with Wilcoxon signed rank test (two-sided). Significances ($p < 0.05$) are highlighted as red.

Friedman's ANOVA (*p*-values)		
Perceptual outcome	**70% DC**	**5% DC**
Hit rate	0.0155	0.4579
FA rate	0.3644	0.1648
Sensitivity *d'*	0.0050	0.7235
Criterion *c*	0.8648	0.3883
Real accuracy	0.0559	0.6327
Scrambled accuracy	0.8202	0.5196

Post-hoc pairwise comparison (*p*-values)				
70% DC	**Hit rate**		**Sensitivity *d'***	
	Raw	**FDR-corrected**	**Raw**	**FDR-corrected**
Baseline vs. VA	0.0324	0.1621	0.0003	0.0027
Baseline vs. VP	0.9678	0.9678	0.8612	0.8824
Baseline vs. DA	0.3395	0.6790	0.7570	0.8824
Baseline vs. DP	0.5098	0.8411	0.8191	0.8824
VA vs. VP	0.0128	0.1281	0.0110	0.0275
VA vs. DA	0.0780	0.1950	0.0035	0.0117
VA vs. DP	0.0544	0.1812	0.0032	0.0117
VP vs. DA	0.6766	0.8411	0.6964	0.8824
VP vs. DP	0.6377	0.8411	0.8824	0.8824
DA vs. DP	0.7570	0.8411	0.7982	0.8824

Post-hoc Wilcoxon tests (FDR-corrected) further indicated hit rate and sensitivity (*d'*) differences specifically for the 70% DC condition. For *d'*, VA showed additional significant differences with VP and DP, which was not shown in confound-unadjusted analysis, confirming that these effects persisted even after adjusting for confounders.

For the region-invariant analysis, the confound-adjusted Friedman's test yielded the following:

Supplementary Table 6: Region-invariant analysis after confounder regression. Pairwise comparison was conducted with Wilcoxon signed rank test (two-sided). Significances ($p < 0.05$) are highlighted as red.

Perceptual outcome	**70% DC**	**5% DC**
Hit rate	0.9234	0.0488
FA rate	0.6309	0.5806
Sensitivity *d'*	0.4711	0.3871
Criterion *c*	0.9044	0.0461
Real accuracy	0.0517	0.3246
Scrambled accuracy	0.3246	0.9044

Here too, the 5% DC condition showed significant differences in hit rate and criterion, with

marginal difference (p = 0.0517) observed in the categorization accuracy of real images. Overall, the inclusion of these confounds did not change our primary conclusions. We have detailed this additional analysis in the main manuscript.

"We performed a control analysis to account for potential confounding factors that might contribute to the observed LIFU effects. Specifically, we evaluated both subject-specific and block-specific nuisance variables. The subject-specific factors included age, sex, and noise awareness (i.e., whether participants were aware of transducer noise during LIFU-ON blocks), while the block-specific factors encompassed lateral deviation, sonication order, and the beam attenuation ratio. To mitigate any influence from these variables, we regressed them out and re-conducted both region-specific and region-invariant analyses. We confirmed that these factors had a negligible impact on our primary results, suggesting that the observed changes in perceptual outcomes were not attributable to these confounds (Supplementary Table 5, 6)." (Line 328, Page 12)

"To address potential confounds, we included six nuisance variables in the analysis. Subject-specific factors correspond to age, sex, and the awareness of transducer noise during LIFU-ON blocks (coded as 0 for unaware, 0.5 for unsure, and 1 for aware). Block-specific factors such as lateral deviation, sonication order, and tissue attenuation ratio (in dB) were also incorporated. Missing data were imputed using group means or modal categories (for noise awareness), and lateral deviation and tissue attenuation ratio averaged across LIFU-ON conditions were used for baseline. In region-invariant analyses, we did not include sonication order as a confound. We adjusted for these confounding variables using linear regression to obtain residuals, followed by Friedman's tests on these residuals to confirm if our main findings remained significant (Supplementary Table 5, 6)." (Line 763, Page 23)

Question 6:

As an additional point, looking at the data presented in the supplement (in the statistics tab) I am not sure I understand the degrees of freedom. I agree with the df(numerator), which should indeed be 1 for the intercept and DC, and 3 for target and the interaction, but how are the degrees of freedom of the error (I'm assuming that's what df2 is?) always 201 across all effects? Wouldn't the intercept and DC have a different df(error) compared to block/target and the interaction?

Response:

Thank you for raising this point. In our previous manuscript, we used a linear mixed effects model and performed ANOVA with MATLAB's 'anova()' function. According to the web documentation (https://www.mathworks.com/help/stats/linearmixedmodel.anova.html), the error degrees of freedom in that context are calculated as $n-p$ (where n is the total number of observations and p is the number of estimated coefficients). This is why, regardless of whether an effect is numeric (like the intercept or DC) or categorical (like target or the interaction, which use indicator variables), the error df remained constant (e.g., 201).

In our updated analysis, however, we are using the Friedman's test. Here, the degrees of freedom are entirely determined by the study design:

- Treatment (Columns) df: $k - 1$, where k is the number of conditions.
- Error df: $(n - 1)(k - 1)$, reflecting the blocking by subjects and the within-subject variability.

- Total df: $n \times k - 1$.

Thus, the error df of 201 in the previous manuscript arises from the model-based estimation and remains uniform across effects, whereas in the Friedman's test, the error df is computed based solely on the experimental design. We have updated the Source Data file accordingly.

Question 7:

Also, what software was used to analyze the data?

Response:

We used MATLAB R2024a for the entire data analysis and statistical testing. We explicitly mentioned this in the revised manuscript.

> *"Template rescaling and its post-hoc validation were performed with custom-built code and SPM12 (https://www.fil.ion.ucl.ac.uk/spm) both executed on MATLAB R2024a. Neuronagivation processes including transducer targeting and coordinate recording were conducted on proprietary Brainsight software v2.5.4 (https://www.rogue-research.com/downloads/). Head segmentation with CHARM was conducted from the SimNIBS v4.1.0 (https://simnibs.github.io/simnibs/build/html/index.html). Ultrasound beam simulation was performed on BabelBrain v0.4.3 (https://proteusmrighifu.github.io/BabelBrain/). The 3D renderings of brain and beam overlaps were performed using MRIcroGL v1.2.20220720 (https://www.nitrc.org/projects/mricrogl). The staircase and main task were based on a custom-modified version of the public code for the near-threshold behavioral paradigm (https://github.com/BiyuHeLab/NatCommun_Levinson2021/) utilizing the Psychophysics Toolbox executed on MATLAB R2024a. For fMRI data preprocessing and voxel-level analysis, we employed the AFNI software suite (linux_ubuntu_16_64; http://afni.nimh.nih.gov/). All other analyses (e.g., random number generation, perceptual outcome calculation, statistical tests, and network-level analysis) and visualizations were conducted on MATLAB R2024a."* (Line 774, Page 23)

Question 8:

Last, I do have one question regarding the experimental design: LIFU is administered in 30s blocks ON and OFF, which has been done by many others, but I'm not clear as to whether the task was performed only during the ON periods or during the OFF periods too? If the latter, is there any difference between the ON and OFF periods?

In short, at present I just have a hard time understanding what has been done. Without radically more clarity in the analysis, it is very hard to evaluate whether any of the conclusions here are justified.

Response:

Thank you for raising this point. The choice of 30-second epochs was based on previous literature which showed significant LIFU effects (Badran et al., 2020; Cain et al., 2021; Mahdavi et al., 2023). The visual task (lasting about 12 minutes) was performed continuously during both the ON and

OFF epochs. We explicitly clarified this by modifying the subplot of Fig. 1b and main text.

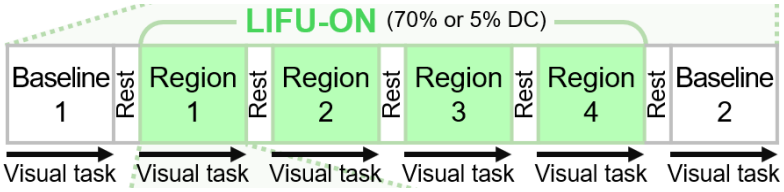


Fig. R3: Modified version of the subplot in Fig. 1b. We clearly denoted that visual tasks were performed continuously within one block.

"Visual tasks were conducted continuously within each block." (Line 149, Page 5)

We acknowledge that a direct analysis comparing trials during LIFU-ON vs. OFF epochs would be interesting. However, because the visual task was self-initiated by participants and LIFU administration was manually triggered by the experimenter after the task onset, we did not achieve precise synchronization between LIFU cycles and behavioral responses. As a result, it is not possible to reliably separate responses during LIFU-ON and OFF epochs with the current dataset. We plan to incorporate automated synchronization in future studies to address this limitation more directly.

Reviewer #2 (Remarks to the Author):

This study sought to elucidate how LIFU to different thalamic regions affects conscious perception. There are serious issues regarding the LIFU aspects of the study and the ability of this device to reach the intended target(s). More detail on modelling procedures for each individual is necessary. To this point, each target for each participant received a different amount of energy and this was not accounted for and thus no conclusion on the intervention between conditions can be made as is. Also, the distinction between 5% and 70% duty cycles is not well justified and unclear how and when these two energy levels were delivered. There is also some concern for potential carry-over effect of delivering multiple sessions of LIFU to each intended target within the same session. As such, these issues and points below need to be addressed before acceptance can be considered.

Response:

We sincerely thank the reviewer for critical assessment of our study's methodological framework, particularly regarding the LIFU targeting and sonication parameters. We take these methodological issues very seriously and have thoroughly addressed each point in our detailed responses below.

Question 1:

Abstract - Please include something about the connectivity data.

Response:

Thank you for the suggestion. We have added the relevant sentence to the revised Abstract:

"Complementary connectivity analysis of a large-scale functional magnetic resonance imaging dataset confirms strong transmodal connectivity of the VA thalamus with frontoparietal and default-mode networks." (Line 30, Page 1)

Question 2:

Introduction - Lines 73-74: Please change to "hypothesized to lead to …" and you could add the following reference. The influence of duty cycle for inhibition/excitation is not established.

Response:

We revised our expression accordingly and cited the suggested references.

Question 3:

Lines 74-75: Acknowledgement of previous human sub-cortical LIFU applications should be mentioned here especially including previous thalamus work. Nakajima, K., Osada, T., Ogawa, A., Tanaka, M., Oka, S., Kamagata, K., ... & Konishi, S. (2022). A causal role of anterior prefrontal-putamen circuit for response inhibition revealed by transcranial ultrasound stimulation in humans.

Cell Reports, 40(7). Legon, W., Ai, L., Bansal, P., & Mueller, J. K. (2018). Neuromodulation with single-element transcranial focused ultrasound in human thalamus. Human brain mapping, 39(5), 1995-2006.

Response:

Thank you for your suggestion. We cited these references and explicitly mentioned them in the revised manuscript:

"Pioneering work has applied LIFU to human subcortical regions, including the thalamus, putamen, and basal ganglia." (Line 67, Page 2)

Question 4:

More justification or setup in the introduction regarding hypotheses regarding the chosen duty cycles should be offered.

Response:

Thank you for raising this important issue of duty cycle choice. We had careful consideration on duty cycle choice. We elaborated our logic in a separate paragraph in the Introduction (see below). Based on these observations, we employed 70% and 5% DC to hypothetically induce excitation and inhibition, respectively.

"LIFU has been shown to modulate neural activity in a bidirectional manner, with the directionality governed by sonication parameters such as the duty cycle (DC; the percentage of time the ultrasound is actively transmitted within each pulse cycle). Early work using neuroimaging, electrophysiological, and behavioral measures indicates that higher DC values generally elicit excitatory effects, whereas lower DC values result in inhibition. This pattern is supported by both computational models and in vivo experiments, which demonstrate that 70% DC effectively lowers neuronal activation thresholds and facilitates excitation. In contrast, 5% DC is associated with diminished evoked potentials and suppressed blood oxygen level-dependent (BOLD) responses consistent with functional inhibition. These observations align with the neuronal intramembrane cavitation excitation (NICE) model, which posits that the value of DC determines the direction of neuromodulation." (Line 74, Page 2)

Question 5:

Line 81. Replace stimulate with modulate

Response:

Thanks for pointing this out. We modified the expression as suggested. Moreover, throughout the text and figures, we replaced the word "stimulate" to more neutral terms (e.g., "target", "administer", or "modulate").

Question 6:

There is nothing in the introduction regarding the thalamic connectivity portion of the experiments that looks to be used to help explain differences in experimental results? More on this is needed.

Response:

Thanks for the suggestion. We added a sentence describing the thalamocortical connectivity analysis at the last paragraph of the Introduction.

"To interpret the observed effects at a systems level, we conduct complementary thalamocortical connectivity analyses." (Line 92, Page 3)

Question 7:

Methods - Line 436-437: How were skull effects considered?

Response:

We conducted a post-hoc analysis to estimate the skull effects. We employed a template rescaling method whereby the MNI template brain was adjusted to match each participant's head size using three measured dimensions (i.e., width, depth, and side-to-top distance). With the CHARM method in SimNIBS pipeline, we segmented skin, skull, and brain, and used BabelBrain software to simulate ultrasound intensity and thermal profiles. We also identified the skull-induced beam shortening (Supplementary Fig. 7) and intensity attenuation compared to a free-field reference (Fig. 2c). This approach allowed us to account for skull effects in our simulation and ensure accurate estimation of LIFU effects.

Question 8:

Line 440: This is not clear. Would not changing the duty cycle also change the pulse width? Please also use pulse duration instead of pulse width.

Response:

We apologize for the mistake. We revised with correct information.

"… pulse duration: 100 ms; duty cycle: 70% and 5% (yielding pulse width of 70 and 5 ms, respectively)" (Line 522, Page 17)

Question 9:

Lines 441-442: Please also provide the Isppa and are these values for outside or inside the head? Please also provide the incident pressure in kPa and estimated pressure at the focal zone in kPa.

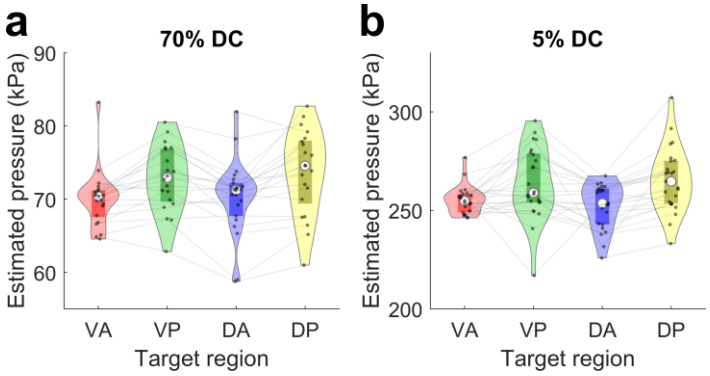
Response:

The free-field I_{sppa} ($I_{sppa.0}$) values are 1.03 W cm^{-2} (70% DC) and 14.4 W cm^{-2} (5% DC). The incident pressure ($P_{r.0}$) values are 197 kPa (70% DC) and 710 kPa (5% DC). These values are without any attenuations (i.e., outside the head).

Based on intensity attenuation ratio (Fig. 2c) and the principle that acoustic intensity is proportional to the square of pressure, we estimated the acoustic pressures to be 72.6 ± 4.2 kPa for a 70% DC and 258.2 ± 18.3 kPa for a 5% DC (mean ± SD). These results are mentioned in the main manuscript. Condition-specific distributions are provided in Supplementary Fig. 6, and a complete parameter summary adhering to ITRUSST reporting guidelines is included in Supplementary Table 1.

"This yielded estimated pressure of 72.6 ± 4.2 kPa for 70% DC and 258.2 ± 18.3 kPa for 5% DC." (Line 183, Page 6)

"non-derated spatial-peak temporal average intensity ($I_{spta.0}$): 1.02 W cm^{-2}; derated spatial-peak temporal average intensity ($I_{spta.3}$): 0.72 W cm^{-2}; non-derated spatial-peak pulse-average intensity ($I_{sppa.0}$): 1.03 W cm^{-2} (70% DC) and 14.4 W cm^{-2} (DC 5%); non-derated incident pressure ($P_{r.0}$): 197 kPa (70% DC) and 710 kPa (5% DC); derated incident pressure ($P_{r.3}$): 165 kPa (70% DC) and 619 kPa (5% DC)." (Line 524, Page 17)



Supplementary Fig. 1: Estimated tissue pressure. Region-specific pressure values for **a** 70% and **b** 5% duty cycles (DCs). Boxes indicate interquartile ranges. For panel a, VA: *n* = 19; VP: *n* = 19; DA: *n* = 18; DP: *n* = 18. For panel b, VA: *n* = 19; VP: *n* = 19; DA: *n* = 21; DP: *n* = 22. Median values are marked by white circles. VA: ventral anterior thalamus; VP: ventral posterior thalamus; DA: dorsal anterior thalamus; DP: dorsal posterior thalamus.

Supplementary Table 1: A report summary based on ITRUSST reporting guidelines.

Transducer and drive system description				
Transducer manufacturer and model number	BrainSonix, BXPulsar 1002			
Transducer center frequency	659.6 kHz			
Transducer geometry	Aperture: 61.5 mm			
Drive system components	All-in-one system (POC-155, Advantech)			
Drive system settings				
Operating frequency	650 kHz			
Output level settings	0.72 W cm^{-2} ($I_{spta.3}$)			
Focal position settings	Fixed			
Free field acoustic parameters				
Spatial-peak pressure amplitude	197 kPa (70% DC) and 710 kPa (5% DC)			
Position of spatial-peak pressure amplitude	80.7 mm			
Size of focal volume	42.5 mm (length), 4.4 mm (width)			
Description	Focal depth was validated by Blatek Industries, Inc. with ASTM E1065/E1065M Standard guidelines. Focal volume size was measured with a bilaminar membrane hydrophone with a 0.4 mm diameter active receive aperture (data provided by BrainSonix Corporations).			
Pulse timing parameters				
	Duration	Ramp duration	Ramp shape	Repetition interval/frequency
70% DC Pulse Pulse train Pulse train repeat	70 ms 30 s 11.5 min	0 0 0	Rectangular Rectangular Rectangular	100 ms/10 Hz 60 s/0.017 Hz
5% DC Pulse Pulse train Pulse train repeat	5 ms 30 s 11.5 min	0 0 0	Rectangular Rectangular Rectangular	100 ms/10 Hz 60 s/0.017 Hz
***In situ* estimates of exposure parameters**				
Estimated *in situ* pressure amplitude at the target	72.6 ± 4.2 kPa (70% DC, mean ± SD) 258.2 ± 18.3 kPa (5% DC, mean ± SD)			
Estimated *in situ* mechanical index	0.20 (70% DC); 0.77 (5% DC)			
Estimated temperature rise	0.23 ± 0.04 °C (mean ± SD)			
Description	Pressure and temperature rise were simulated *in silico* on BabelBrain. Mechanical index was calculated from $P_{r.3}$, without considering skull attenuation.			

Question 10:

Lines 460-481: I do not understand the purpose of these procedures. Please elaborate.

Response:

The purpose of these procedures was to achieve accurate LIFU targeting without requiring individual anatomical imaging, thereby streamlining the experimental process. We measured head dimensions (width, depth, and side-to-top distance) with a digital caliper and used these

data to rescale a standard MNI template to each participant's head size. This template rescaling method allowed us to generate individualized brain images for LIFU targeting, eliminating the need for subject-specific CT or MRI scans. We validated this approach on 13 subjects by comparing coordinates from the rescaled template to those from actual T1 images, finding minimal deviations (2.15 ± 0.82 mm) and confirming targeting accuracy. We elaborated the purpose of this method both in Results and Discussion sections.

"As a proof of concept, we used a pseudo-anatomical reference for beam targeting and post-hoc analysis (see Methods: Template rescaling method for details). This was achieved by rescaling the MNI template image to match each participant's head size measured before the main task (Fig. 1c). We validated this approach on a separate dataset (n = 13) by comparing actual anatomical images with their rescaled pseudo-anatomical images, finding an average deviation of 2.15 ± 0.82 mm (mean ± SD; Supplementary Fig. 4)." (Line 165, Page 6)

"Thinking of future clinical applications, we also mitigated the reliance on anatomical imaging (e.g., computed tomography or T1 MRI) by using a template rescaling method. We verified that this method meets the desired targeting accuracy in two ways. First, using a separate MRI dataset, we confirmed small deviations of approximately 2 mm relative to the actual T1 image (Supplementary Fig. 4). Second, we observed that effect size increased as the beam-target lateral deviation decreased, which is a good indicator that our method effectively estimates the target location within acceptable error range (Supplementary Fig. 11a). Our approach could expedite the experimental procedures and enable LIFU application in populations where anatomical scans are challenging, such as infants, individuals with implanted medical devices, or those with severe claustrophobia." (Line 456, Page 16)

Question 11:

Lines 482 – 489: Please provide images of these models for each of the targeted thalamic regions in transverse or coronal and sagittal views either in supplemental or manuscript figure. Please identify how the skull was identified and acoustic properties assigned.

Response:

In the updated Fig. 2, we generated individual simulated beam trajectories and overlayed the "beam-present" regions (defined as regions exceeding a normalized intensity of 25%) as well as the outlines of the four thalamic targets (see Fig. R4). Also, individual-level trajectories are now presented in Supplementary Fig. 8-10. Exemplary mosaic views are shown below (see Fig. R5).

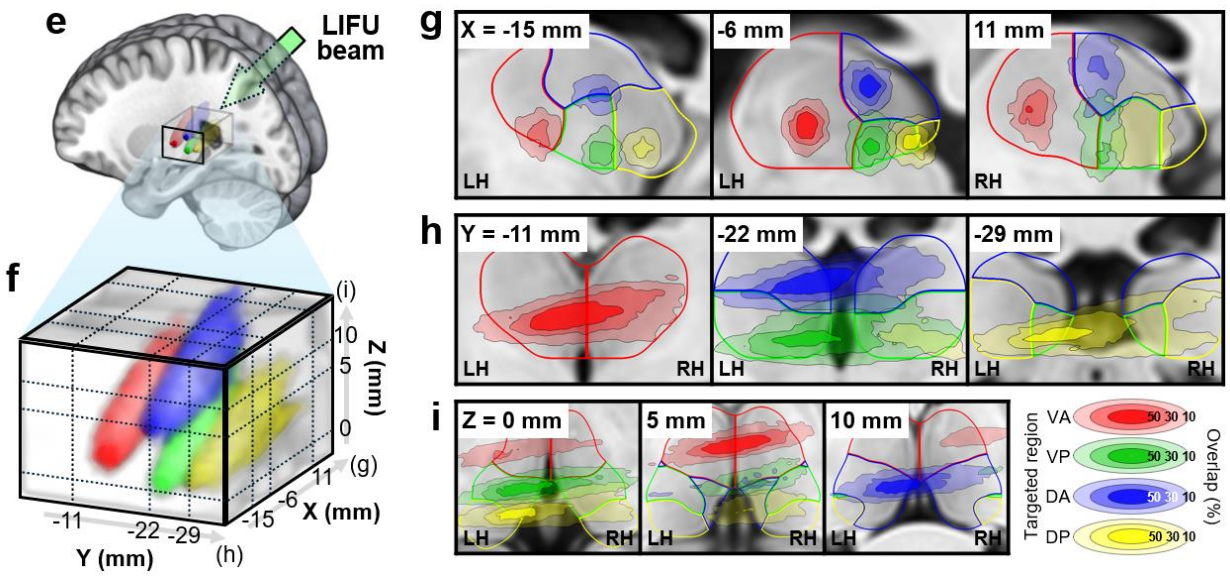


Fig. R4: Post-hoc LIFU beam analysis (adopted from Fig. 2). e Three-dimensional rendering of the beam overlaps. For each LIFU-ON block, the "beam-present" region includes all voxels with > 25% of the simulated relative beam intensity. The overlap percentage at a given voxel is the fraction of participants for whom that voxel is classified as beam-present. Colored regions highlight voxels where the overlap exceeds 10%. An arrow indicates the LIFU beam originating from the right side. **f** Zoomed-in view of panel e, with dashed lines indicating the locations of cross-sections. **g-i** Cross-sectional views in the **g** sagittal, **h** coronal, and **i** axial planes, displaying overlap percentages and outlines of bilateral thalamic regions. Higher opacity reflects greater overlap.

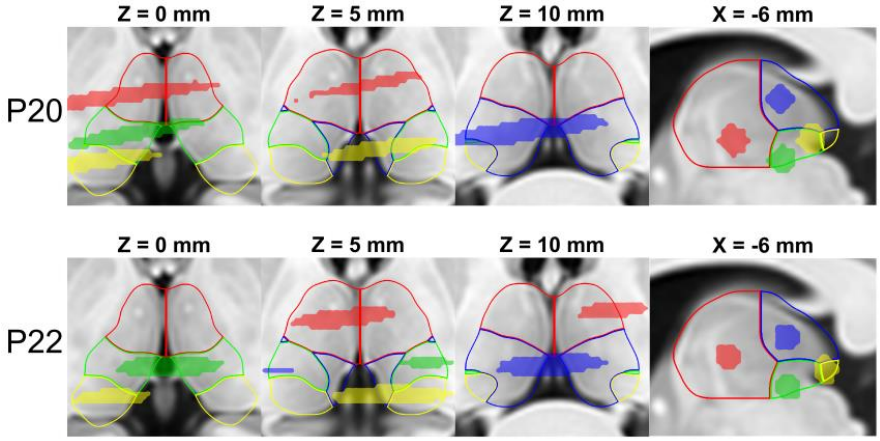


Fig. R5: Two examples of individual-level mosaic view of beam trajectories. Colored regions are the "beam-present" regions, exceeding 25% of relative intensity. Color coding for the thalamic regions is as follows: VA (red), VP (green), DA (blue), and DP (yellow) thalamus.

Within the pseudo-anatomical images, we segmented skin, skull, and brain with CHARM (Complete Head Anatomy Reconstruction Method) contained in SimNIBS pipeline. The acoustic properties assignment was conducted with BabelBrain software (Pichardo, 2023), through the following main procedures. First, simplified masks were created using segmented MRI data (generated by CHARM). The proportion of trabecular bone was set to 80%. These masks assign homogenized acoustic properties (density, longitudinal and shear speeds, and attenuation) to tissues, with predefined acoustic properties derived from literature values (Table R1).

Table R1: Material properties used in BabelBrain, adopted from Pichardo (2023). The *f* denotes frequency in MHz.

Material	Density (kg/m^3)	Longitudinal speed (m/s)	Shear speed (m/s)	Longitudinal attenuation (Np/m/MHz)	Shear attenuation (Np/m/MHz)
Water	1000	1500	-	-	-
Skin	1116	1537	-	4.6×*f*	-
Brain	1041	1562	-	6.9×*f*	-
Cortical bone	1896.5	2415+120×*f*	1369+346×*f*	163×*f*	329×*f*
Trabecular bone	1738.0	2063+282×*f*	1206+212×*f*	162×*f*	329×*f*

The assigned acoustic properties are employed in the BabelViscoFDTD library, a GPU-accelerated solver of viscoelastic wave equations. Based on acoustic intensity distributions calculated in transcranial modeling, thermal effects are estimated using the Bio Heat Transfer Equation (BHTE). See Pichardo (2023) for full details. We also incorporated this procedure in the Methods section.

"Simulation of ultrasound intensity and thermal profiles was performed using a two-step segmentation and modeling approach. We initially segmented the pseudo-anatomical image into scalp, skull, and brain employing the Complete Head Anatomy Reconstruction Method (CHARM). For a more accurate segmentation, we used a fat-suppression method (https://github.com/ProteusMRIgHIFU/EnhanceTUS_Zadeh_JNE). For cases where participant-specific ultrasound trajectories extended beyond the headspace due to anatomical variability, the label assignment of scalp was manually extended outward by 1 cm, and the mesh was recalculated with CHARM, ensuring that all recorded head-transducer contact points lie within the adjusted headspace.

Subsequently, acoustic property assignments and ultrasound beam simulations were performed using an open-source software named BabelBrain. Simplified masks based on the CHARM-generated MRI segmentations were utilized, with trabecular bone proportion set at 80%. These masks assigned homogenized acoustic properties, including density, longitudinal and shear speeds of sound, and attenuation, derived from established literature values. Simulations were conducted up to 100 mm below the transducer with a spatial resolution of six points-per-wavelength.

The tissue attenuation ratio was determined by how much acoustic energy is lost when ultrasound propagates through tissue compared to a water reference (exported as "RatioLosses" variable in BabelBrain software). In short, the software selects the plane at which the maximum tissue pressure occurs. It then computes the acoustic energy in this plane as the sum over all voxels of the squared pressure amplitude. Similarly, the code processes the free-field pressure amplitude data at the same plane. The

tissue attenuation ratio is then defined as the ratio of the tissue acoustic energy to the water acoustic energy at that specific plane.

Thermal effects resulting from ultrasound exposure were estimated using the Bio Heat Transfer Equation (BHTE). Baseline temperature for thermal modeling was set at 37 ºC. During the simulation, BabelBrain software calculates the temperature evolution in the tissue (i.e., skin, skull, brain, and target point) both during ON and OFF epochs, which are then concatenated, yielding an array of 4 × 72,000 size. Maximal temperature was defined as the maximal value within this array." (Line 577, Page 18)

Question 12:

The purpose of the connectivity data is not clear.

Response:

The connectivity analysis was performed to provide mechanistic context for the behavioral effects, by identifying the large-scale neural networks associated with each targeted thalamic region. At the network level, we identified a graded unimodal-transmodal connectivity gradient. At the voxel level, VA thalamus was strongly connected with the medial and dorsolateral prefrontal cortices.

These connectivity profiles help explain our findings. For example, VA thalamus showing dominant connections with frontoparietal and default-mode networks may support its unique role in modulating sensitivity. The region-invariant decrease in categorization accuracy at 70% duty cycle may stem from the common connectivity with the visual cortex, while the conservative decision bias observed at 5% duty cycle is consistent with an inhibitory influence. As such, the connectivity data provides neurobiological insight underlying the observed effects of LIFU on conscious perception.

Question 13:

Results - Figure 1 is well done and informative. Line 133 is not clear to me nor are the methods in this case. Was 5% and 70% duty interleaved within each of the LIFU ON blocks?

Response:

We apologize that our description was unclear. For one participant, only one duty cycle (either 70% or 5%) was used consistently. We revised the caption of Figure 1 and relevant Methods section for clarification.

"Each ON epoch consists of 300 pulses (pulse repetition frequency = 10 Hz) at a fixed duty cycle and temporal-average intensity ($I_{spta.3}$ = 0.72 W cm^{-2})." (Line 150, Page 5)

"Within one participant, only one duty cycle (either 70% or 5%) was applied throughout the four LIFU-ON blocks." (Line 629, Page 20)

Question 14:

Lines 146-151: This section is unclear. You overlaid a beam profile (collected in free water or modelled? And modelled in free water?) on an MNI template brain? This is somewhat disingenuous as this would be a 'stylized beam' in an ideal circumstance and not indicative of that actual beam in an individual head. Please make this clear in the figure caption.

Response:

We thank the reviewer for this comment. To clarify, the beam profile previously presented in Fig. 2a was not an overlay of a free-water measurement but an *in-silico* simulation that incorporated brain tissue acoustic properties. To address this unclarity, we included the free-field beam profile in Supplementary Fig. 5 and removed previous Fig. 2a. Also, we calculated the simulated beam depth (Supplementary Fig. 7).

Question 15:

Lines 152-156: This is a highly unorthodox method. Do you not have the modelled beam for each participant? And did you collect MRIs for each participant? The employed method of confirming deviation is not acceptable. Please use the modelling data from each participant and do within subject estimates of beam relative to the target.

Response:

We appreciate the reviewer's concern regarding the reliance on rescaled template (i.e., pseudo-anatomical) images. However, we note that we performed beam modeling in an individual manner, using the pseudo-anatomical images of each subject. Our approach was intended to eliminate the need for individual anatomical imaging, thereby enhancing clinical applicability and expediting experimental procedures.

While fully acknowledging its limitation, to validate this method, we obtained the head size and actual T1 images of 13 subjects (previously five; we increased the number of samples during our revision process) from an individual dataset. We measured the Euclidean distances of the thalamic regions in pseudo-anatomical and actual T1 images and found an average deviation of 2.15 ± 0.82 mm (Supplementary Fig. 4). We also observed that the effect of LIFU increased as the recorded lateral deviation decreased (Supplementary Fig. 10), further supporting the robustness and precision of our targeting approach. While modeling with actual anatomical data would be ideal, our current strategy represents a pragmatic balance between accuracy and feasibility and is well supported by our validation data.

Question 16:

Lines 157 – 164: It is not clear what this post-hoc analysis was or how it was done. Please define focal center. That is a massive estimated intensity loss and does not seem to converge with current research. Please explain/define how these methods were done.

Response:

Thank you for pointing this out. In the previous manuscript, the focal center was defined by the "ideal" beam center, located 80 mm from the transducer surface. To ensure rigor, we re-ran the entire post-hoc simulation and observed beam shortening caused by the skull (reported in Supplementary Fig. 7). While the original representation of ideal beam center captured the successful "orientation" of the beams, we acknowledge it may not accurately represent the current simulation. Therefore, we replaced this analysis with "beam-present" region—voxels exceeding 25% of the maximum relative intensity—and examined their overlap across participants to confirm successful targeting.

For the estimated intensity, we re-ran the analysis and discovered a -8.78 ± 0.56 dB energy loss mainly due to the skull. The tissue attenuation ratio was obtained by calculating the acoustic energy at the plane of maximal tissue pressure and comparing it with a corresponding free-field simulation. Our observed attenuation ratio aligns with previous empirical and simulation results in humans. For example, Lord et al. (2024) reported 90% energy loss at 500 kHz, Legon et al. (2024) mentioned approximately 75% loss at 700 kHz, and Chen et al. (2023) found about 95% energy loss at 650 kHz.

Question 17:

Figure 2. A. Please explicitly state that this is not an actual acoustic model from an individual but a 'stylized beam' overlaid (please define in methods how this was properly scaled) on the template brain to explicitly only show/illlustrate the size of the beam (empirical free water? Modelled free water?) relative to the thalamus.

Response:

We thank the reviewer for the opportunity to clarify. The beam profile presented in Fig. 2a was not derived from free-water modeling but from an *in silico* simulation that incorporated brain tissue acoustic properties. This simulation was intended solely to illustrate the relative size and shape of the LIFU beam with respect to the thalamus. However, to prevent any ambiguity, we removed Fig. 2a and instead included the free-field beam profile in Supplementary Fig. 5. Also, we included the simulated individual beam profiles in Supplementary Fig. 8-10.

Question 18:

B. Please quantify or describe what beam center is. Pressure maximum?

Response:

In our original Fig. 2b, we defined the "estimated" beam center as a 3D coordinate 80 mm from the transducer, recorded from the neuronavigation system. However, due to skull-induced beam shortening, this idealized focal point may no longer be accurate. Consequently, we now report the

"beam-present" region, comprising all voxels above 25% of the maximum relative intensity, and then assessed the overlap of these voxels across participants to confirm targeting.

Question 19:

K. I see that some individuals registered zeros is that correct? Why? And how was this data dealt with?

Response:

Thank you for pointing this out. Previously, the relative intensity was calculated based on the maximum-intensity point within the simulation. However, we discovered that certain technical issues in those calculations produced near-zero intensity values for some participants. Specifically, skull-induced beam aberration can broaden the beam, causing reduced intensity at the original focal center. To address this, we re-ran the entire simulation and computed the attenuation ratio by examining the acoustic energy at the plane of maximal tissue pressure and comparing it with the equivalent free-field simulation. We believe this new approach better reflects relative intensity.

Question 20:

Because a standardized intensity was delivered outside the head this results in varying intensity in the head for each participant for each thalamic target. This variability must be accounted for as a covariate in all statistical models before any rational conclusions can be made.

Response:

Thank you for the comment. In the revised manuscript, we incorporated the tissue attenuation ratio for each thalamic region (in dB) as one of the confounds and repeated our statistical analysis after regressing them. We confirmed that this did not change our primary conclusions (Supplementary Table 5, 6).

Supplementary Table 5: Region-specific analysis after confounder regression. Post-hoc pairwise comparison was conducted with Wilcoxon signed rank test (two-sided). Significances ($p < 0.05$) are highlighted as red.

Friedman's ANOVA (*p*-values)		
Perceptual outcome	**70% DC**	**5% DC**
Hit rate	0.0155	0.4579
FA rate	0.3644	0.1648
Sensitivity *d'*	0.0050	0.7235
Criterion *c*	0.8648	0.3883
Real accuracy	0.0559	0.6327
Scrambled accuracy	0.8202	0.5196

Post-hoc pairwise comparison (*p*-values)				
70% DC	**Hit rate**		**Sensitivity *d'***	
	Raw	**FDR-corrected**	**Raw**	**FDR-corrected**
Baseline vs. VA	0.0324	0.1621	0.0003	0.0027
Baseline vs. VP	0.9678	0.9678	0.8612	0.8824
Baseline vs. DA	0.3395	0.6790	0.7570	0.8824
Baseline vs. DP	0.5098	0.8411	0.8191	0.8824
VA vs. VP	0.0128	0.1281	0.0110	0.0275
VA vs. DA	0.0780	0.1950	0.0035	0.0117
VA vs. DP	0.0544	0.1812	0.0032	0.0117
VP vs. DA	0.6766	0.8411	0.6964	0.8824
VP vs. DP	0.6377	0.8411	0.8824	0.8824
DA vs. DP	0.7570	0.8411	0.7982	0.8824

Supplementary Table 6: Region-invariant analysis after confounder regression. Pairwise comparison was conducted with Wilcoxon signed rank test (two-sided). Significances ($p < 0.05$) are highlighted as red.

Perceptual outcome	70% DC	5% DC
Hit rate	0.9234	0.0488
FA rate	0.6309	0.5806
Sensitivity *d'*	0.4711	0.3871
Criterion *c*	0.9044	0.0461
Real accuracy	0.0517	0.3246
Scrambled accuracy	0.3246	0.9044

Question 21:

Please provide size and depth of each thalamic target for each individual also based upon sex.

Response:

As requested, we estimated the size (i.e., volume) and depth of the four target regions for every participant (Fig. R6, 7). The volume was calculated with the pseudo-anatomical images. For the depth calculation, we defined two parallel planes oriented perpendicular to the beam vector; one passing through the head-transducer contact point (i.e., the transducer plane) and the other passing through the center-of-mass of the thalamic region. The depth was then measured as the distance between these planes. For both size and depth, we identified significant differences between male and female for all regions (FDR-corrected $p < 0.05$, Mann-Whitney U-test).

This differences in size and depth are likely originating from the sex difference in head width (Fig. R8; $p = 0.0034$, Mann-Whitney U-test). To address its potential influences, we included sex during confound regression and confirmed that it does not significantly alter our primary conclusions (see Supplementary Table 5,6; attached above).

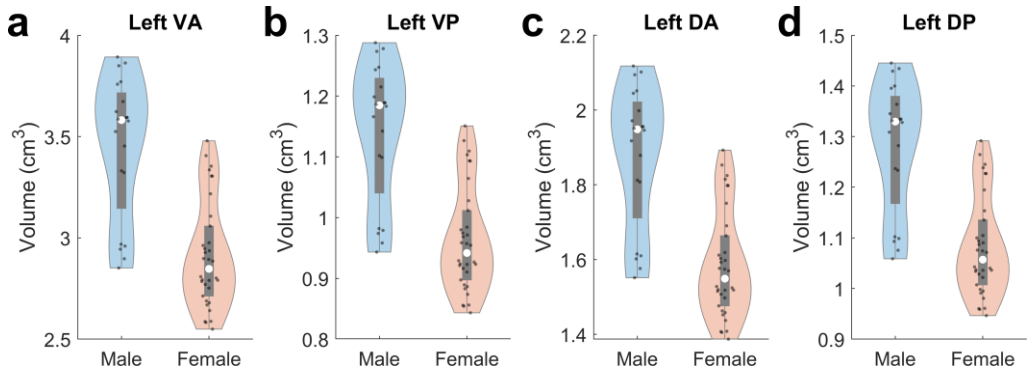


Fig. R6: Estimated volume of four thalamic regions, including (a) VA, (b) VP, (c) DA, and (d) DP for each sex.

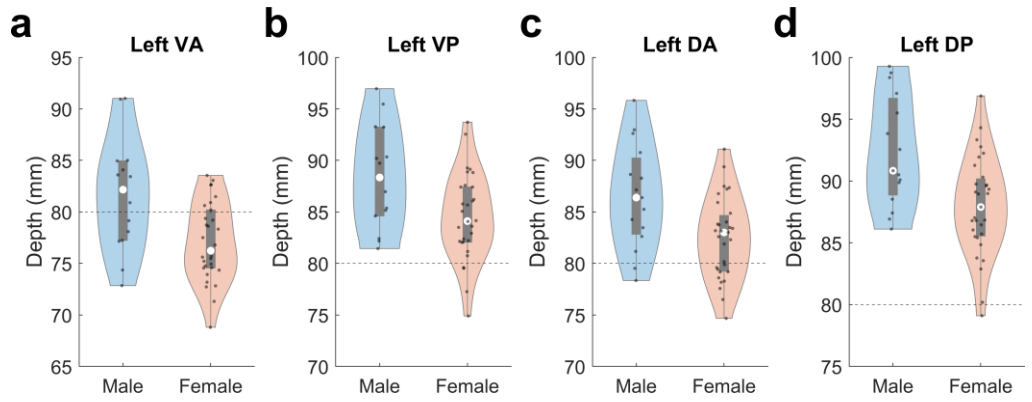


Fig. R7: Estimated depth of four thalamic regions, including (a) VA, (b) VP, (c) DA, and (d) DP for each sex. The depth was defined as the perpendicular distance between the plane through the head-transducer contact and the plane through the region's center along the beam direction. Dash lines indicate theoretical focal depth (80 mm).

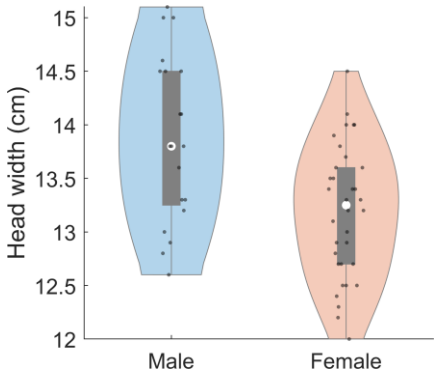


Fig. R8: Measured head width for each sex. Head width is defined by the Euclidean distance between left and right supratragic notches.

Question 22:

Please also provide how (given the variability in depth of the targets) this was compensated for with the current transducer. Was a different offset used for each participant? How was this done?

Response:

We tried to account for individual differences in target location by rescaling the MNI template to each participant's head size. Given the transducer's bullet-like, elongated beam profile, we did not adjust the *z*-axis offset in on an individual basis. Instead, we made our best effort to minimize the lateral deviation to ensure optimal alignment. Lateral deviation was defined as the Euclidean distance between the beam vector and the center of mass of the target thalamic region, and was collected from the neuronavigation system.

To further verify the targeting in an individual level, we generated mosaic views of pseudo-anatomical images, overlayed with the "beam-present" regions for each beam (i.e., voxels with > 25% relative intensity) and the outlines of four thalamic targets. Full figures are included as Supplementary Fig. 8-10. Exemplary mosaic views are shown below in Fig. R9.

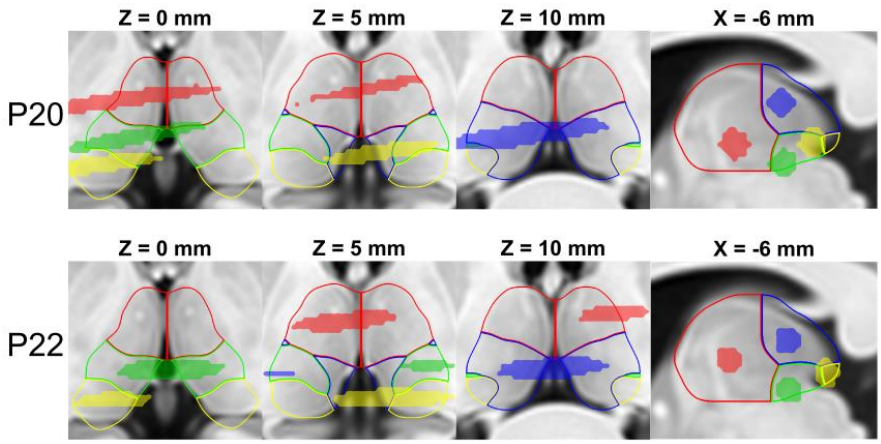


Fig. R9: Two examples of individual-level mosaic view of beam trajectories. Colored regions are the "beam-present" regions, exceeding 25% of relative intensity. Color coding for the thalamic regions is as follows: VA (red), VP (green), DA (blue), and DP (yellow) thalamus.

Question 23:

There needs to be some comment somewhere on the potential carry-over effects of the protocol between the four regions. It would be helpful if some analysis of potential stimulation carry-over or accrual was present.

Response:

We appreciate the reviewer's concern regarding potential region-by-region accrual effects. To address this, for region-specific analysis, we incorporated the sonication order of each region as

one of the confounding variables. We regressed it out prior to performing Friedman's test and verified that the sonication order did not significantly alter our main findings.

Also, as we noted in the main manuscript, we analyzed baseline measures taken at the beginning and end of the experimental session and found no significant differences in perceptual outcomes (Supplementary Fig. 3). This further indicates that there is likely no detectable cumulative or carry-over effect between regions within our experimental paradigm. We have updated the manuscript text to highlight these points.

"We examined whether LIFU induced any carry-over effects within our approximately one-hour experimental timeframe by comparing Baseline-1 and Baseline-2. For both duty cycles, we found no significant differences in any perceptual outcome ($p > 0.05$, Wilcoxon signed-rank test, two-sided; Supplementary Fig. 3). These results suggest that LIFU did not produce detectable long-lasting or cumulative effects on visual perception." (Line 136, Page 4)

"… while the block-specific factors encompassed lateral deviation, sonication order, and the beam attenuation ratio. To mitigate any influence from these variables, we regressed them out and re-conducted both region-specific and region-invariant analyses. We confirmed that these factors had a negligible impact on our primary results, suggesting that the observed changes in perceptual outcomes were not attributable to these confounds (Supplementary Table 5, 6)." (Line 331, Page 12)

Question 24:

Discussion - Line 293-294. This statement is disingenuous. How can an effect from this experiment be associated with increased functional connectivity from another data set where there was no manipulation?

Response:

We thank the reviewer for pointing this out. We agree that the expression 'increased' was erroneous. Our original intention was to suggest the association with 'higher' functional connectivity between VA and transmodal networks as confirmed in complementary fMRI analysis, and we never intended to argue that their functional connectivity actually 'increases' upon LIFU modulation. We modified the figure caption accordingly.

"This is associated with higher functional connectivity between the VA thalamus and transmodal areas compared to other thalamic nuclei." (Line 352, Page 13)

Question 25:

Lines 324-327: The references provided do not support this statement for inhibition vs. excitation for the duty cycles used. These statements need to be tempered and because you have no direct evidence of excitation vs inhibition for the results of this study. The 70% condition, in effect, delivered more energy than the 5% condition so could this not be a reason? / Lines 346 – 348: Same as above.

Response:

We appreciate the reviewer's observation regarding the interpretation of the results in our study. Since we maintained a fixed temporal-average intensity, the "overall energy" delivered per unit area over the sonication duration should be the same. However, the pulse-average intensity and instantaneous pressure were indeed higher in the 5% duty cycle condition. These differences imply that the distinct pulse characteristics may yield different neuromodulatory effects. This is what neuronal intramembrane cavitation excitation (NICE) predicts, which suggests that various types of neurons react differentially to different duty cycles (Plaksin et al., 2016).

However, we fully acknowledge that our data cannot conclusively ascribe the observed outcomes to actual excitation or inhibition due to lack of concurrent functional imaging. This revised interpretation is now reflected in the updated manuscript text.

"LIFU has been shown to modulate neural activity in a bidirectional manner, with the directionality governed by sonication parameters such as the duty cycle (DC; the percentage of time the ultrasound is actively transmitted within each pulse cycle). Early work using neuroimaging, electrophysiological, and behavioral measures indicates that higher DC values generally elicit excitatory effects, whereas lower DC values result in inhibition. This pattern is supported by both computational models and in vivo experiments, which demonstrate that 70% DC effectively lowers neuronal activation thresholds and facilitates excitation. In contrast, 5% DC is associated with diminished evoked potentials and suppressed blood oxygen level-dependent (BOLD) responses consistent with functional inhibition. These observations align with the neuronal intramembrane cavitation excitation (NICE) model, which posits that the value of DC determines the direction of neuromodulation." (Line 74, Page 2)

"Our study has limitations. First, we did not perform functional imaging; while our findings are consistent with previously suggested duty cycle-dependent bidirectionality of LIFU effects, we acknowledge that further investigation using concurrent functional imaging is necessary to conclusively prove the existence and directionality of neural activity and connectivity changes." (Line 466, Page 16)

Question 26:

Line 352 - 355: Methodologically, this work does not represent a significant advancement in human LIFU studies as several LIFU studies that have previously be done are not acknowledged or discussed in particular a seminal study in human thalamus by Legon et al. 2018.

Response:

We thank the reviewer for highlighting the need to acknowledge earlier human LIFU works. We agree that previous studies have played a crucial role in advancing human ultrasound neuromodulation. In response, we have revised the manuscript to clarify that (1) studies including Legon et al. (2018) demonstrated effective single-target thalamic modulation and (2) our study extends these findings by enabling multi-regional stimulation with sub-centimeter precision in human. We have also augmented our discussion with additional references that further underscore the feasibility and evolving potential of focused ultrasound for deep brain neuromodulation in humans. These revisions ensure that our work is accurately contextualized within the broader literature.

"Our work also represents several refinements in the field of human LIFU research. First, in contrast to prior human studies that employed only one duty cycle, we incorporated two duty cycles (70% vs. 5%) into our experimental design and found duty cycle-dependent effects. Second, while previous LIFU investigations successfully demonstrated targeted neuromodulation of the human thalamus, our work extends these findings by multi-regional stimulation with sub-centimeter precision. The average lateral deviation of the beam from the target was approximately 2 mm (Fig. 2b), smaller than beam width of about 5 mm. This was comparable to or better than those in previous works (3–9 mm), reinforcing the robustness and precision of our targeting approach. Third, while regional targeting within the subcortex has been demonstrated in rodents, our targeting approach successfully distinguished various thalamic sub-regions in human, suggesting promising avenues for translational scientific investigations and clinical applications (Fig. 2e-m)." (Line 444, Page 15)

Question 27:

Line 375: Please also discuss these energy levels relative to other human LIFU studies. These Ispta values are extremely low.

Response:

Thank you for the comment. In the Discussion section of the updated manuscript, we explicitly discussed this issue and compared it with other studies, attached below.

"Second, our simulations suggest significant energy loss (about 87%), mainly due to skull, with potential additional attenuation by residual air pockets in hairs. Despite such loss, previous human studies using comparable or lower intensities have still demonstrated significant neuromodulatory effects. In a recent rodent study, a temporal-average intensity as low as 0.0075 W cm^{-2} (roughly 100-fold lower than ours; about 10-fold lower after considering skull-induced attenuation) was sufficient to induce a detectable calcium signal. Nonetheless, advanced approaches such as direct attenuation compensation can help mitigate these issues." (Line 470, Page 16)

Reviewer #3 (Remarks to the Author):

The study reveals that stimulating specific thalamic regions with low-intensity focused ultrasound (LIFU) affects conscious perception in distinct ways. Stimulation of the ventral anterior (VA) thalamus with a high duty cycle (70%) significantly enhanced sensitivity in object recognition, suggesting an excitatory effect that boosts perceptual processing in this transmodal-dominant area. Conversely, a low duty cycle (5%) induced a conservative decision bias, indicating a potential inhibitory effect on perception. Additionally, categorization accuracy for real images decreased under high duty cycle stimulation, possibly due to interactions with the visual network. These findings highlight both region-specific and general effects of LIFU on perception, offering new insights into the thalamus's role in conscious processing and supporting the potential of LIFU as a non-invasive tool for targeted brain modulation.

Response:

Thank you for the reviewer's insightful summary. We appreciate your comments and have carefully addressed the concerns as below.

Question 1:

If the same Ispta was used across different duty cycle conditions, this would mean that a higher Isppa was applied under the 5% duty cycle condition. If so, it would be helpful to state this explicitly in the text. Additionally, in Fig. 1b, it would be clearer to represent the 5% duty cycle condition with a higher amplitude in the illustration to reflect this difference. This difference in amplitude could also explain the greater occurrence of auditory confounding artifacts observed at the 5% duty cycle, as the higher instantaneous pressure may have produced more noticeable sounds.

Response:

We fully agree. As the reviewer correctly mentioned, non-derated spatial-peak pulse-average intensity ($I_{sppa.0}$) values are 1.03 W cm^{-2} for 70% DC and 14.4 W cm^{-2} for 5% DC. This likely caused a larger sound from the transducer for 5% DC. To clarify, we also confirmed that regressing the awareness of this transducer noise as one of the confounds did not change our primary conclusions (Supplementary Table 5, 6). In the revised manuscript, we explicitly stated this information both in Results, Conclusion, and Methods sections. We also updated Fig. 1b as suggested.

"The pulse-averaged intensity was higher at 5% DC." (Line 151, Page 5)

"A 5% DC generated more noticeable noise from the transducer due to the higher pulse-averaged intensity than 70% DC (14.4 W cm^{-2} for 5% DC vs. 1.03 W cm^{-2} for 70% DC), given a fixed temporal-average intensity." (Line 411, Page 15)

"Since we used a fixed $I_{spta.3}$ for two duty cycles, we note that pulse-average intensity (I_{sppa}) was higher at DC 5% condition." (Line 528, Page 17)

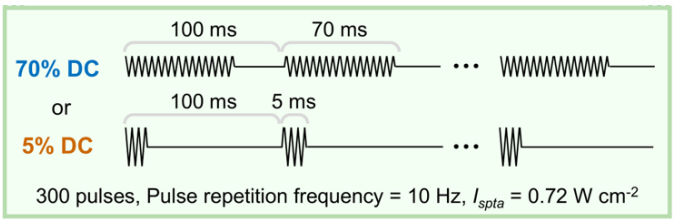


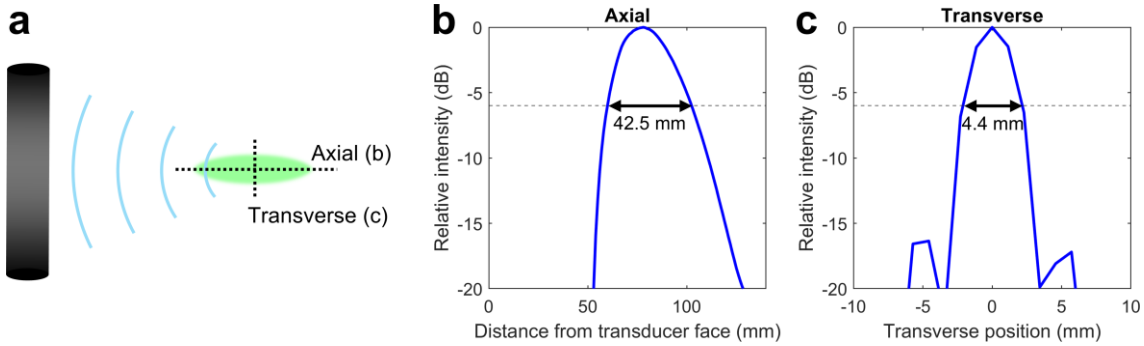
Fig. R10: A modified subplot of Fig. 1b. The pulse trains of two duty cycles are depicted with different heights, emphasizing distinct pulse-averaged intensities.

Question 2:

In Fig. 2a, the beam profile appears to be simulated, and it seems that there is no actual acoustic beam profile measured with a hydrophone or similar device. Including this measured data would strengthen the study by providing empirical validation of the beam's characteristics, such as focal precision and intensity distribution, especially through the human skull.

Response:

Thank you for your comment regarding the need to include a measured beam profile. In response, we have obtained experimentally measured acoustic field data from the device manufacturer. Specifically, they measured the beam characteristics of a 80 mm transducer using a bilaminar membrane hydrophone with a 0.4 mm diameter active receive aperture under free-field conditions (i.e., without the presence of a skull). These measurements include both the axial and transverse beam profile. The -6 dB focal length and width were measured as 42.5 mm and 4.4 mm, respectively. We have now incorporated this result in Supplementary Fig. 5.



Supplementary Fig. 2: Free-field acoustic profile of the transducer. a Schematic illustration of intensity measurements. **b, c** Relative beam intensity across **b** axial and **c** transverse directions. Dashed lines indicate -6 dB, equivalent to 25% of the initial intensity. Axial and transverse focal extent (-6 dB) was determined to be 42.5 mm and 4.4 mm, respectively. This data was provided by BrainSonix Corporations.

"In a free-field condition, our transducer produces a bullet-like beam shape with an aspect ratio of approximately 10, measuring 42.5 mm in length and 4.4 mm in width (defined by -6 dB attenuation; Supplementary Fig. 5)." (Line 172, Page 6)

Question 3:

While the actual sonication was directed from the temporal lobe, Fig. 2b-e represents the acoustic focus as a central point, which may not fully capture the anatomical accuracy of targeting relative to the sonication direction. It would be beneficial to consider ways to improve this representation to better reflect the actual targeting accuracy based on the sonication approach.

Response:

Thank you for your suggestion. We agree that representing the acoustic focus as a single dot may not fully capture the directional profile of the sonication beam. To address this concern, we generated individual mosaic views of pseudo-anatomical images. These views overlay the "beam-present" regions (defined as voxels exceeding a relative intensity of 25%) as well as the outlines of the four thalamic targets. This approach provides a more detailed visualization of the beam's "bullet-like" profile and confirms accurate targeting of each thalamic region. Full figures are now presented in Supplementary Fig. 8-10. Exemplary mosaic views are shown below.

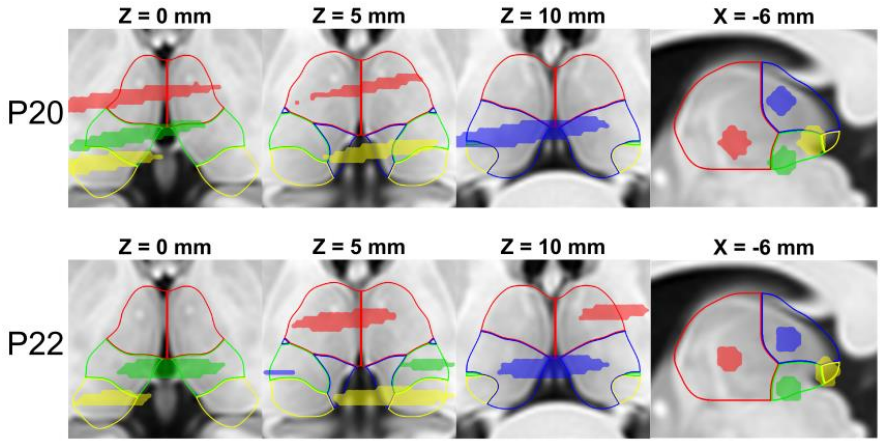


Fig. R11: Two examples of individual-level mosaic view of beam trajectories. Colored regions are the "beam-present" regions, exceeding 25% of relative intensity. Color coding for the thalamic regions is as follows: VA (red), VP (green), DA (blue), and DP (yellow) thalamus.

Question 4:

While it is noted that "the order of stimulated regions was pseudo-randomized," it would be helpful to include information on how long it took to adjust the single-element transducer between different target locations. This detail would provide additional clarity on the practicality of the experimental setup.

Response:

Thank you for pointing out this important aspect of the experiment. We allowed the participants to rest between the blocks (3.8 ± 2.4 min), during which we adjusted the transducer position. After the transducer was successfully repositioned and the participants stated that they are willing to

start the next block, we proceeded. However, the exact length of the rest differed for every block, partly because in some cases the experimenters needed additional time for transducer readjustment. To provide more transparency, we provide the length of rest period between each block in the Source Data file. We also modified Fig. 1b and included the relevant information in Methods.

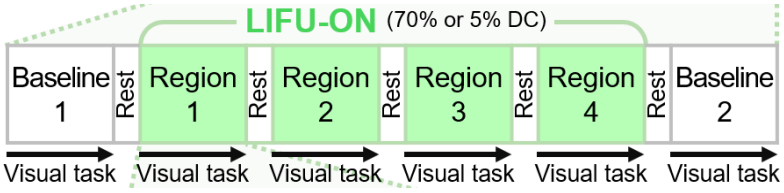


Fig. R12: A modified subplot of Fig. 1b, highlighting the existence of rests between blocks.

"Participants were allowed to rest between blocks (3.8 ± 2.4 min; mean ± SD; see Source Data), during which the transducer was readjusted. We also measured the eye-to-screen distance and readjusted the position of the laptop if necessary (see below). After the transducer was successfully repositioned and the participants were willing to proceed, the visual task was self-initiated by participants and LIFU administration was manually triggered by the experimenter after the task onset." (Line 630, Page 20)

Question 5:

The study lacks data on acoustic attenuation specific to the experimental setup, such as simulations or measurements. Including this information would enhance the understanding of ultrasound energy transmission through the skull and its impact on targeting accuracy.

Response:

We thank the reviewer for raising this issue. We estimated ultrasound energy transmission and attenuation by simulations. Specifically, we used the SimNIBS pipeline to segment the pseudo-anatomical images into skin, skull, and brain compartments and then applied BabelBrain software to simulate the ultrasound intensity. Specifically, the tissue attenuation ratio is now obtained by calculating the acoustic energy at the plane of maximal tissue pressure and comparing it with a corresponding free-field simulation. We reported the updated tissue attenuation ratio values in Fig. 2c. To ensure clarity, we substantially updated the Methods section regarding the simulation process.

"Simulation of ultrasound intensity and thermal profiles was performed using a two-step segmentation and modeling approach. We initially segmented the pseudo-anatomical image into scalp, skull, and brain employing the Complete Head Anatomy Reconstruction Method (CHARM). For a more accurate segmentation, we used a fat-suppression method (https://github.com/ProteusMRIgHIFU/EnhanceTUS_Zadeh_JNE). For cases where participant-specific ultrasound trajectories extended beyond the headspace due to anatomical variability, the label assignment

of scalp was manually extended outward by 1 cm, and the mesh was recalculated with CHARM, ensuring that all recorded head-transducer contact points lie within the adjusted headspace.

Subsequently, acoustic property assignments and ultrasound beam simulations were performed using an open-source software named BabelBrain. Simplified masks based on the CHARM-generated MRI segmentations were utilized, with trabecular bone proportion set at 80%. These masks assigned homogenized acoustic properties, including density, longitudinal and shear speeds of sound, and attenuation, derived from established literature values. Simulations were conducted up to 100 mm below the transducer with a spatial resolution of six points-per-wavelength.

The tissue attenuation ratio was determined by how much acoustic energy is lost when ultrasound propagates through tissue compared to a water reference (exported as "RatioLosses" variable in BabelBrain software). In short, the software selects the plane at which the maximum tissue pressure occurs. It then computes the acoustic energy in this plane as the sum over all voxels of the squared pressure amplitude. Similarly, the code processes the free-field pressure amplitude data at the same plane. The tissue attenuation ratio is then defined as the ratio of the tissue acoustic energy to the water acoustic energy at that specific plane." (Line 577, Page 18)

Question 6:

To enhance the reporting on ultrasound stimulation conditions and improve alignment with safety standards, it would be beneficial to refer to Martin et al. (2024), ITRUSST Consensus on Standardised Reporting for Transcranial Ultrasound Stimulation and Aubry et al. (2023), ITRUSST Consensus on Biophysical Safety for Transcranial Ultrasonic Stimulation. Including these guidelines would help provide a more comprehensive report on stimulation parameters and reinforce the study's adherence to established safety standards.

Response:

Thank you for your suggestion. After re-running the thermal simulation, we ensured that the estimated temperature rise is approximately 0.2 ºC and never exceeded 0.5 ºC (Fig. 2d), well below the recommendations suggested by Aubry et al. (2023). Following Martin et al. (2024), We included a report summary table containing key information of our study as Supplementary Table 1 and referred to them in the manuscript.

"The maximal temperature rise simulated across skin, skull, and brain was 0.23 ± 0.04 °C (mean ± SD) and never exceeded 0.5 °C (Fig. 2d), complying with ITRUSST risk recommendations." (Line 185, Page 6)

"A comprehensive parameter report summary based on the ITRUSST reporting guidelines can be found in Supplementary Table 1." (Line 533, Page 17)

Supplementary Table 1: A report summary based on ITRUSST reporting guidelines.

Transducer and drive system description				
Transducer manufacturer and model number	BrainSonix, BXPulsar 1002			
Transducer center frequency	659.6 kHz			
Transducer geometry	Aperture: 61.5 mm			
Drive system components	All-in-one system (POC-155, Advantech)			
Drive system settings				
Operating frequency	650 kHz			
Output level settings	0.72 W cm^{-2} ($I_{spta.3}$)			
Focal position settings	Fixed			
Free field acoustic parameters				
Spatial-peak pressure amplitude	197 kPa (70% DC) and 710 kPa (5% DC)			
Position of spatial-peak pressure amplitude	80.7 mm			
Size of focal volume	42.5 mm (length), 4.4 mm (width)			
Description	Focal depth was validated by Blatek Industries, Inc. with ASTM E1065/E1065M Standard guidelines. Focal volume size was measured with a bilaminar membrane hydrophone with a 0.4 mm diameter active receive aperture (data provided by BrainSonix Corporations).			
Pulse timing parameters				
	Duration	Ramp duration	Ramp shape	Repetition interval/frequency
70% DC Pulse Pulse train Pulse train repeat	70 ms 30 s 11.5 min	0 0 0	Rectangular Rectangular Rectangular	100 ms/10 Hz 60 s/0.017 Hz
5% DC Pulse Pulse train Pulse train repeat	5 ms 30 s 11.5 min	0 0 0	Rectangular Rectangular Rectangular	100 ms/10 Hz 60 s/0.017 Hz
***In situ* estimates of exposure parameters**				
Estimated *in situ* pressure amplitude at the target	72.6 ± 4.2 kPa (70% DC, mean ± SD) 258.2 ± 18.3 kPa (5% DC, mean ± SD)			
Estimated *in situ* mechanical index	0.20 (70% DC); 0.77 (5% DC)			
Estimated temperature rise	0.23 ± 0.04 °C (mean ± SD)			
Description	Pressure and temperature rise were simulated *in silico* on BabelBrain. Mechanical index was calculated from $P_{r.3}$, without considering skull attenuation.			

Question 7:

It would be helpful to provide additional explanation regarding the statement, “The focal depth was also validated by Blatek Ultrasonic Transducers (Boalsburg, PA, USA).” Further details on the validation process, including methods or measurements used to confirm the focal depth, would clarify this aspect of the setup and enhance the reproducibility of the study.

Response:

We apologize for the unclarity. The focal depth validation was performed using a standardized pulse-echo testing procedure in accordance with ASTM E1065/E1065M Standard guidelines. The general process is summarized as follows:

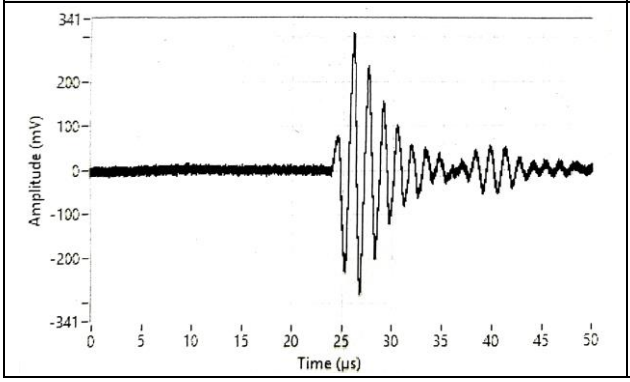
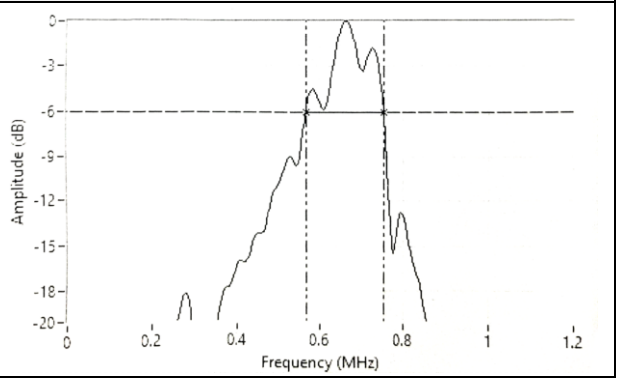
A 1” steel target is used as the reflector and positioned at the transducer’s expected focal point (80 mm). After pulsing, the steel target provides an echo signal, and the system records the echo amplitude as the distance is varied. The focal point is identified at the position where the echo amplitude reaches its maximum. Other measurements, such as the -20 dB pulse length and the frequency bandwidth, were also recorded.

To ensure transparency and reproducibility of our study, we have attached the complete test data report of our transducer as Supplementary Table 2 (see below), containing all details, signal waveform, and frequency spectrum. In the main text, we also clarified that this test was compliant with ASTM E1065/E1065M guidelines.

“The free-field focal depth is 80.7 mm, validated by Blatek Industries, Inc. (Boalsburg, PA, USA), following ASTM E1065/E1065M Standard guidelines (see Supplementary Table 2 for test data report).” (Line 513, Page 17)

Supplementary Table 2: Test data report of the transducer used in the study.

Transducer information	
Part number	AT29823
Description	BrainSonix, 650 kHz, 61.5 mm diameter, 80 mm focal depth, 10 ft cable (BNC)
Test data	
Serial number	0919-5441
Test time	7:14 AM 12/29/2021
Sensitivity	588 mV
Frequency bandwidth	28%
-20dB pulse length	9.201 µs
Center frequency	659.6 kHz
Test target	1” steel @ focal
Test path length	80.7 mm
Equipment used	
Oscilloscope model	LeCroy WS452
Oscilloscope calibration due date	6/30/2022
Pulser model	5072PR
Pulser calibration due date	07/19/2022
Pulser/receiver settings	
Energy level	4
Damping	3 (50 Ohm)
Gain	-23 dB

Signal waveform	Frequency spectrum
	

RESPONSE TO REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

I thank the authors for having put sincere effort into the response and I do think the methods are now a little clearer. Nonetheless, I still have some substantial concern regarding both the statistical analysis and the fact that I don't think the analysis and results support the key conclusions.

Response:

We acknowledge the reviewer's concerns regarding the statistical analysis and the alignment of our results with the key conclusions. We have thoroughly re-evaluated these aspects and improved our statistical methodology to ensure that the analysis robustly supports our findings.

Question 1:

First, I still have a concern relating to multiple comparisons. I'll start by saying that the authors now apply appropriate FDR correction to post-hoc analyses (I assume they used a Benjamini–Yekutieli procedure, which controls the FDR under highly correlated tests, as in the present case? Which FDR procedure is not said in the methods).

Response:

We apologize for the unclarity. We now clarified it in the Methods section.

For the entire FDR correction steps, we used Benjamini-Hochberg procedure using MATLAB's 'mafdr' function. (Page 22, Line 759)

Question 2:

Where I think an adjustment for multiplicity might still be needed is the fact that the exact same data are being analyzed twice: once in the region-specific analysis and once in the region-invariant analysis. It is exactly the same data going into the two analyses. To explain, consider the following: had any of the findings in the region-specific analysis shown that all/a majority of the LIFU blocks were different from baseline, it would have necessarily implied a significant difference in the region-invariant "baseline vs LIFU" analysis. So these two analyses cannot be independent of each other. Rather, they seem to be a reanalysis of the same data (unlike, say, had they first looked at baseline vs LIFU, and then compared between the LIFU conditions, that would have been an orthogonal analysis of the data). I understand the authors "consider them" complementary, but still, the issue is just one of re-analyzing multiple times, in multiple different parallel ways, the exact same data.

Response:

We thank the reviewer for raising this important methodological concern about analyzing the same data in multiple ways. We acknowledge that our initial approach to conducting separate target-

specific and target-invariant analyses on identical dataset was problematic, as these analyses were not truly independent.

To address this concern, we have fundamentally restructured our analytical approach to eliminate the issue of two dependent analyses. We now analyze the data sequentially rather than in parallel. Our revised approach begins with a single **omnibus F-test**, testing whether there are any differences among five targeting conditions (baseline plus four thalamic regions). This test serves as a global screen controls family-wise error rate at the outcome level.

For outcomes that pass the omnibus test after FDR correction, we proceed with target-specific post-hoc comparisons, applying additional FDR correction within this family of contrasts. Notably, we have eliminated the prior target-invariant "baseline vs. LIFU" analysis entirely, as any general stimulation effects should be inherently captured by the omnibus model and follow-up contrasts. This approach ensures that each dataset undergoes a single, unified analytical pathway, thereby addressing the reviewer's concern about multiple re-analysis.

Our statistical implementation involves fitting a single linear mixed-effects model specified as:

$$Outcome \sim Target + 6\ Covariates + (1|Subject)$$

where *Target* includes baseline and four thalamic regions (total five categories) and *Covariates* encompass age, sex, noise awareness, beam lateral deviation, sonication order, and the beam attenuation ratio. Analyses were performed using MATLAB's 'fitlme' and 'anova' functions with Satterthwaite approximation.

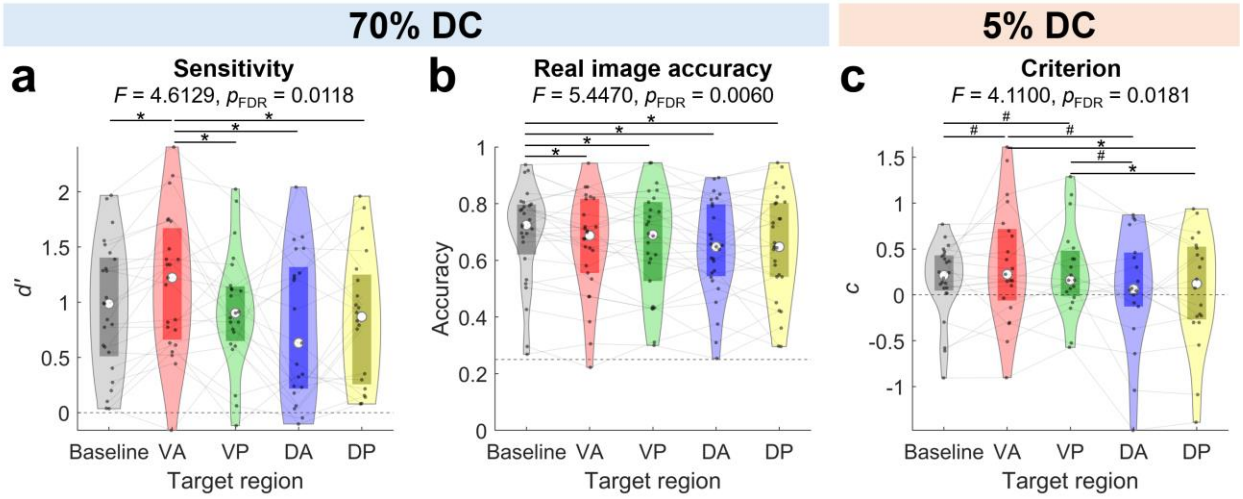
The results of omnibus ANOVA were as follows:

Updated Table 1: Omnibus ANOVA for the effect of sonication target region on behavioral outcomes.

DC	Perceptual outcome	df_1	df_2	*F*-value	*p*-value	FDR-corrected *p*-value
70%	Hit rate	4	107	1.3407	0.2596	0.5192
	FA rate	4	110	0.6525	0.6263	0.8351
	Sensitivity *d'*	4	90	4.6129	0.0020	0.0118
	Criterion *c*	4	89	0.5325	0.7122	0.8546
	Real image accuracy	4	107	5.4470	0.0005	0.0060
	Scrambled image accuracy	4	111	0.2665	0.8989	0.9553
5%	Hit rate	4	104	2.3734	0.0570	0.1710
	FA rate	4	104	2.0005	0.0999	0.2398
	Sensitivity *d'*	4	78	0.7452	0.5641	0.8351
	Criterion *c*	4	77	4.1100	0.0045	0.0181
	Real image accuracy	4	103	0.9165	0.4574	0.7841
	Scrambled image accuracy	4	106	0.1658	0.9553	0.9553

Notes: significances (FDR-corrected $p < 0.05$) are highlighted as red. DC: duty cycle; df: degree of freedom; FA: false alarm.

Since the omnibus *F*-test evaluates the overall Target effect, all subsequent analyses are conditional on omnibus significance. Afterwards, for three significant conditions, we conducted post-hoc pairwise comparisons to test target-specific effects, as shown in updated Fig. 3. This revised analytical strategy is more conservative than our original approach provides stronger control of Type I error, while preserving statistical power to detect meaningful effects.



Updated Fig. 3: Target-specific changes in perceptual outcomes during human thalamic LIFU administration. Perceptual outcomes that demonstrated significant omnibus effects, measured at **a-b** 70% and **c** 5% duty cycle, including **a** sensitivity *d′*, **b** categorization accuracy for real images, and **c** criterion *c*, are shown. Asterisks and pounds denote significant differences in post-hoc pairwise comparisons (asterisk: FDR-corrected $p < 0.05$; pound: FDR-uncorrected $p < 0.05$). Boxes indicate interquartile ranges. Median values are marked by white circles.

Full details of this revised approach have been incorporated into the Results and Methods sections of the updated manuscript.

In the updated Results:

> To determine whether LIFU-targeting of different thalamic regions produces differential effects on visual perception, we conducted omnibus *F*-tests using linear mixed-effects (LME) models for each perceptual outcome and duty cycle condition. (Page 8, Line 223)
>
> The omnibus analysis revealed significant regional effects for three conditions (Table 1). At 70% DC, we observed significant omnibus effects for sensitivity *d′* ($F(4,90) = 4.6129$, $p_{FDR} = 0.0118$) and real-image accuracy ($F(4,107) = 5.4470$, $p_{FDR} = 0.0060$). At 5% DC, criterion *c* showed a significant omnibus effect ($F(4,77) = 4.1100$, $p_{FDR} = 0.0181$). Other perceptual measures, including hit rate, false alarm rate, and scrambled image accuracy, did not demonstrate significant effects after FDR correction at either duty cycle condition. (Page 8, Line 229)

In the updated Methods:

> We implemented a sequential analytical approach beginning with omnibus testing using linear mixed-effects (LME) models. This approach was chosen because LME models can directly accommodate covariates, handle non-normal data distributions, and properly account for missing outcome data through restricted maximum likelihood estimation. Our statistical implementation involved fitting a single linear mixed-effects model specified as:
>
> $$Outcome \sim Target + Covariates + (1|Subject)$$
>
> where *Target* includes baseline and four LIFU regions (five categories total) and *Covariates* encompass the six nuisance variables described above. For each perceptual outcome, we first conducted an omnibus F-test to determine whether there were any differences among the five experimental conditions using MATLAB's 'fitlme' and 'anova' functions with Satterthwaite approximation.
>
> Only outcomes that demonstrated significant omnibus effects after FDR correction proceeded to post-hoc analysis. For significant outcomes, we conducted pairwise comparisons between each LIFU condition and baseline using linear contrasts, with additional FDR correction applied within this family of contrasts at $\alpha = 0.05$. This approach ensures that each dataset undergoes a single, unified analytical pathway while controlling family-wise error rate at the outcome level. (Page 22, Line 744)

Question 3:

Second, I have a number of concerns relating to the way in which covariates were adjusted for.

The major point here, unless the authors just forgot to say this in the methods, is that the "de-noising" regression followed by a repeat of the Friedman ANOVAs was done with the incorrect degrees of freedom. To explain: In adjusting for these covariates, the authors didn't account (or don't say they did) for the fact that the residuals of the "denoising regression" have fewer degrees of freedom than the original data. Specifically, in a regression the DF of the residuals (i.e., df error) is N-i-1, with N being the number of observations, "i" being the number of confounders in the regression, and "1" for the intercept. Yet, as far as what is in the methods, the Friedman ANOVAs and Wilcoxon t-tests done on the residuals are evaluated at the nominal DF of N observations. Which seems incorrect. (Might be inconsequential, but this is an empirical question.) I know people use this procedure in fMRI all the time, but in that case the inference is then done at a higher level compared to the first-level "denoising" analysis, so the DF relevant to inference in fMRI comes from the number of subjects (either entirely, if in a random effects analysis, or mainly, if doing a mixed effects analysis) meaning the loss of DFs at the level of individual timeseries (i.e., volumes) is almost irrelevant. Here, instead the inference is done at the same level as the denoising regression, thus DFs should be appropriately accounted for. Of course the authors never even report the chi2 statistic of the Friedman ANOVA or the DF, so I'm left to infer.

Hadn't I had the benefit of reading the rebuttal letter, where this two-step procedure is explained clearly, I don't think I would have been able to understand what they did from the text in the current version of the methods. First they say that "To address potential confounders, we included six nuisance variables in the analysis." This is not true. They removed confounders from the data, using a linear regression (which is said later on) and then analyzed the residuals with the same

strategy in the main analysis -- which is buried a few sentences later into this section. A few sentences later they do say ""We adjusted for these covariates using a linear regression to obtain residuals, followed by Friedman's tests on these residuals [...]" but then in the appendix I see they also used this for the Wilcoxon t-test, but that is not written in the methods. All I'm saying is that it would help if the methods described clesarly what was done so one could replicate them.

Response:

We thank the reviewer for raising these important concerns regarding covariate adjustment, degrees of freedom, and methodological clarity.

Regarding the degree-of-freedom issue, we acknowledge that our previous two-step approach of performing denoising regression followed by non-parametric tests was methodologically flawed. As the reviewer correctly points out, this procedure fails to account for the reduced degrees of freedom in the residuals, yet we conducted subsequent tests as if we had the full *N* observations.

To address this concern, we have completely restructured our analytical approach (described in our response to Question 2), eliminating the problematic two-step procedure entirely. Our revised analysis uses a single-step linear mixed-effects modeling framework that properly accounts for all covariates and their associated degrees of freedom simultaneously. This approach incorporates covariates directly into the model rather than through separate preprocessing, and the Satterthwaite approximation automatically calculates appropriate degrees of freedom that account for all model parameters including covariates. As evidence of proper degrees of freedom handling, our omnibus ANOVA table (Updated Table 1 provided above) shows denominator degrees of freedom varying across outcomes (ranging from 80 to 109), reflecting automatic adjustment for missing data patterns and model complexity.

We believe this revision fully resolves the degrees of freedom issue and provides a more statistically rigorous and transparent analytical framework. In addition, we have substantially revised the Methods section to clearly and accurately describe the revised procedure, ensuring full replicability.

Question 4:

Considering the above, since the Friedman ANOVA is not very flexible and doesn't allow including covariates, why not use a linear mixed model, which can accommodate the non-normal data and covariates or a permutation-based ANCOVA (the widely used permuco package in R)? Either of these approaches would allow including the confounders in the main analysis and do away with having to do a separate convoluted analysis with incorrect degrees of freedom.

Response:

We thank the reviewer for this excellent methodological suggestion. We have adopted exactly this approach in our revised analysis. As described in our responses to Questions 2 and 3, we now use linear mixed-effects models throughout our analysis, which directly incorporate covariates without the need for separate preprocessing steps. We believe this revised approach addresses

the concerns you raised and substantially improves the validity and transparency of our statistical inference.

Question 5:

Missing data imputation: there is one throw away sentence in the methods that reads "Missing data were imputed using group means or modal categories (for noise awareness), ...". Was the data imputation done in the main analysis too or only in the 'confounder' analysis? What variables were imputed, only the confounders themselves or also the dependent variables (the sensitivity, criterion, etc for each nucleus)? Was this done also in the main analysis?

Response:

We apologize for the lack of clarity in our original description of missing data handling. To clarify, imputation was applied only to covariates, not to any of the dependent perceptual outcome variables (sensitivity, criterion, hit rate, etc.). Imputation was necessary due to missing or undefined entries. Specifically, lateral deviation and attenuation ratio were not applicable to baseline (LIFU-OFF) blocks and were missing for some participants due to recording or simulation limitations; noise awareness responses were missing for a subset of participants. We used simple imputation methods with mean imputation for lateral deviation (2.251 mm) and attenuation ratio (-8.7812 dB), and modal imputation for noise awareness (most frequent response: 1, indicating awareness of transducer noise). No imputation was performed on any dependent variables. Missing outcome data were handled directly within the mixed-effects modeling framework, which accommodates unbalanced data via restricted maximum likelihood estimation. We have clarified these procedures in the revised Methods section.

> Missing data for noise awareness was imputed using modal category (noise awareness: 1). Missing or baseline values of lateral deviation and attenuation ratio were imputed using group means (lateral deviation: 2.251 mm; attenuation ratio: -8.7812 dB). Given no significant differences between Baseline-1 and Baseline-2 across all measures (Supplementary Fig. 3), we averaged these into a single baseline measure. No imputation was performed for the main perceptual outcome data. (Page 22, Line 738)

Question 6:

With respect to the region-invariant analysis, it is odd that none of the findings in the region-invariant analysis have a correspective in the region-specific analysis. Meaning, neither real images at 70% duty cycle, nor hit rate and criterion at 5% of duty cycle have a correspective significant finding in the more granular region-specific analysis. This must imply that the finding at the region-invariant level are just a side-effect of the reduction of the spread around the median consequent to the averaging of multiple LIFU blocks (and makes me wonder whether the baseline averaging is having a similar smoothing effect on the region-specific analysis -- meaning, would the one result on sensitivity at 70% duty cycle no longer be there in the absence of baseline averaging?). Now, generally that's exactly why we average, to decrease noise. But here what is being averaged is the result from sonication of nuclei which the authors then go on to conclude

are very different because of their cytoarchitectonics and connectivity. It feels like the authors want to have their cake and to eat it too. Either there is a specific role of VA (and, as they conclud, matrix cells/multimodal connectivity) or there is a generalized effect of tFUS on image detection (and then it's not about cytoarchitectonics or connectivity). I know the authors say the two analysis are "complementary" but to me they just appear contradictory (or at least, the significant finding in the baseline vs LIFU just seems completely contrary to the stated conclusion).

Response:

We thank the reviewer for this insightful observation regarding the apparent disconnect between the target-invariant and target-specific findings. Upon implementing the revised statistical framework (as described in our responses to Questions 2 and 3), we revisited our post-hoc results and found a more nuanced and internally consistent pattern of effects than originally reported.

First, with respect to sensitivity d' at 70% DC, our post-hoc analyses reveal a clear target-specific effect centered on VA. Specifically, VA stimulation significantly enhanced sensitivity relative to baseline ($p_{FDR} < 0.05$) and outperformed all other target regions (VA vs. VP, DA, DP: all $p_{FDR} < 0.05$, Table R1). This pattern suggests that the omnibus effect was primarily driven by VA and directly supports our hypothesis regarding the role of VA (and its associated matrix cell architecture and transmodal connectivity; see our response for Question 8 below) in perceptual enhancement. Thus, rather than being target-invariant, this result is strongly VA-specific.

Table R1: Post-hoc pairwise comparison for 70% DC sensitivity.

Comparison	*F*-value	*p*-value	FDR-corrected *p*-value	Direction
Baseline vs. VA	7.3460	0.0079	0.0246	VA > Baseline
Baseline vs. VP	0.4758	0.4919	0.7379	-
Baseline vs. DA	0.0003	0.9866	0.9866	-
Baseline vs. DP	0.0467	0.8293	0.9135	-
VA vs. VP	7.2724	0.0082	0.0246	VA > VP
VA vs. DA	13.7511	0.0003	0.0034	VA > DA
VA vs. DP	9.9654	0.0021	0.0119	VA > DP
VP vs. DA	0.8079	0.3709	0.6181	-
VP vs. DP	0.3813	0.5383	0.7530	-
DA vs. DP	0.0875	0.7680	0.8861	-

Notes: direction of change is noted for $p < 0.05$. Red indicates FDR-corrected $p < 0.05$.

Second, for real image categorization accuracy at 70% DC, we observed a consistent decrease across all four regions relative to baseline ($p_{FDR} < 0.05$; Table R2). This finding suggests a general disruptive effect of high-duty-cycle thalamic sonication on visual categorization. While less regionally selective than the effect on sensitivity, this result still aligns with a broader interpretation: that higher duty-cycle thalamic sonication may exert non-specific effects across thalamic targets.

Table R2: Post-hoc pairwise comparison for 70% DC real image accuracy.

Comparison	*F*-value	*p*-value	FDR-corrected *p*-value	Direction
Baseline vs. VA	9.6224	0.0024	0.0119	Baseline > VA
Baseline vs. VP	12.2472	0.0007	0.0049	Baseline > VP
Baseline vs. DA	16.6370	0.0001	0.0012	Baseline > DA
Baseline vs. DP	19.3209	0.0000	0.0007	Baseline > DP
VA vs. VP	0.2969	0.5869	0.7655	-
VA vs. DA	2.0056	0.1593	0.2986	-
VA vs. DP	2.7777	0.0982	0.1963	-
VP vs. DA	0.6387	0.4257	0.6722	-
VP vs. DP	1.4150	0.2365	0.4174	-
DA vs. DP	0.1061	0.7452	0.8861	-

Notes: direction of change is noted for $p < 0.05$. Red indicates FDR-corrected $p < 0.05$.

Third, at 5% DC, the effects on criterion *c* were more complex (Table R3; Tables R1-3 are also incorporated as Updated Supplementary Table 5). We observed elevated response criteria in both VA and VP compared to baseline and to dorsal targets (DA, DP), although many of these pairwise comparisons were only marginally significant after FDR correction. This partial pattern suggests a regionally heterogeneous but not uniformly target-invariant effect, warranting cautious interpretation.

Table R3: Post-hoc pairwise comparison for 5% DC criterion.

Comparison	*F*-value	*p*-value	FDR-corrected *p*-value	Direction
Baseline vs. VA	4.6718	0.0333	0.0832	VA > Baseline
Baseline vs. VP	4.1937	0.0435	0.0931	VP > Baseline
Baseline vs. DA	0.0347	0.8526	0.9135	-
Baseline vs. DP	0.1309	0.7183	0.8861	-
VA vs. VP	0.0012	0.9725	0.9866	-
VA vs. DA	5.6630	0.0194	0.0529	VA > DA
VA vs. DP	8.3334	0.0049	0.0182	VA > DP
VP vs. DA	4.3694	0.0394	0.0909	VP > DA
VP vs. DP	9.3500	0.0029	0.0125	VP > DP
DA vs. DP	0.3560	0.5522	0.7530	-

Notes: direction of change is noted for $p < 0.05$. Red indicates FDR-corrected $p < 0.05$.

We agree with the reviewer that our earlier framing of a strict dichotomy between "region-specific" and "region-invariant" effects was overly simplistic. Our revised analysis supports a more nuanced interpretation: (1) VA shows distinct enhancement of sensitivity, suggesting a specific functional role consistent with its known anatomy; (2) some effects (e.g., decreased real image accuracy at

high DC) reflect generalized impact of LIFU across thalamic regions; (3) effects at lower duty cycles are variable and may reflect subtle differences in regional responsiveness that are not uniformly robust.

We have revised the manuscript to reflect this more integrative and cautious interpretation and have removed prior claims of unequivocal target-invariant effects at 5% DC.

At 70% DC, LIFU targeting of the VA nucleus specifically enhanced sensitivity d' compared to baseline (VA vs. baseline: p_{FDR} = 0.0246; paired Cohen's d = 0.9164) and significantly outperformed all other target regions (Fig. 3a; VA vs. VP: p_{FDR} = 0.0246, VA vs. DA: p_{FDR} = 0.0034, VA vs. DP: p_{FDR} = 0.0119; see Supplementary Table 5 for full statistics). This pattern demonstrates that the omnibus effect was driven primarily by VA-specific enhancement rather than a generalized effect. (Page 9, Line 247)

For categorization accuracy of real images, a consistent decrease across all four thalamic regions relative to baseline was observed (Fig. 3b; baseline vs. all four regions: $p_{FDR} < 0.05$; average paired Cohen's d = -0.4545; Supplementary Table 5), suggesting a target-invariant disruptive effect of high-DC thalamic sonication on visual categorization tasks.

At 5% DC, the effects of LIFU on criterion c revealed a more complex pattern. We observed elevated criterion (i.e., shift toward a conservative stance) in both VA and VP sonication compared to baseline and to dorsal targets (DA and DP), although many of these pairwise comparisons reached only marginal significance after FDR correction (Fig. 3c; VA vs. DP and VP vs. DP: $p_{FDR} < 0.05$; Supplementary Table 5). This nuanced pattern suggests regionally heterogeneous effects that are not uniformly robust across all comparisons. (Page 9, Line 262)

Question 7:

As mentioned above, what is the effect of baseline averaging? Is the result just a side effect of artificially reducing the variance of the baseline? I think it should be very strightforward for the authors to repeat the Friedman ANOVa (only for d') with the two separate baselines and show that the averaging is inconsequential (that could be a very simple table in the appendix), but at least the reader wouldn't have to be left wondering if the artificial procedure is the reason for the one significant effect upon which much/most of the conclusion is predicated.

Response:

We thank the reviewer for highlighting this important methodological concern regarding the potential impact of baseline averaging on our results. Our decision to average the two baseline conditions was guided by two key considerations beyond variance reduction. First, baseline averaging reduces the total number of statistical comparisons, thereby helping to control the Type I error rate more effectively. Second, it improves the reliability of the baseline estimate by doubling the number of trials (from 90 to 180 responses per participant), which is particularly beneficial given the variability inherent in psychophysical measures.

To address whether our significant effects are artifacts of reduced baseline variance, we conducted the suggested control analysis by repeating the omnibus ANOVA for each baseline (Baseline-1 and Baseline-2) separately, using the same linear mixed-effects model framework.

The results show that the main findings remain robust across both baselines (Updated Supplementary Table 3; see below).

Updated Supplementary Table 3: Omnibus ANOVA test replicated with individual baselines without averaging.

Baseline-1						
DC	**Outcome**	**df_1**	**df_2**	***F*-value**	***p*-value**	**FDR-corrected *p*-value**
70%	Hit rate	4	108	1.1169	0.3524	0.7049
	FA rate	4	109	0.7117	0.5856	0.8378
	Sensitivity *d′*	4	90	4.2123	0.0036	0.0215
	Criterion *c*	4	89	0.4816	0.7492	0.8990
	Real image accuracy	4	107	4.2912	0.0029	0.0215
	Scrambled image accuracy	4	111	0.1927	0.9418	0.9657
5%	Hit rate	4	104	1.6848	0.1591	0.3970
	FA rate	4	104	1.6581	0.1654	0.3970
	Sensitivity *d′*	4	78	0.6504	0.6283	0.8378
	Criterion *c*	4	77	3.0564	0.0215	0.0861
	Real image accuracy	4	103	0.7181	0.5814	0.8378
	Scrambled image accuracy	4	105	0.1431	0.9657	0.9657
Baseline-2						
DC	**Outcome**	**df_1**	**df_2**	***F*-value**	***p*-value**	**FDR-corrected *p*-value**
70%	Hit rate	4	105	1.2346	0.3007	0.6015
	FA rate	4	107	0.6043	0.6604	0.8075
	Sensitivity *d′*	4	84	3.4286	0.0120	0.0480
	Criterion *c*	4	84	0.4939	0.7402	0.8075
	Real image accuracy	4	104	3.8119	0.0062	0.0480
	Scrambled image accuracy	4	108	0.4955	0.7390	0.8075
5%	Hit rate	4	103	2.1152	0.0842	0.2527
	FA rate	4	103	1.8738	0.1207	0.2896
	Sensitivity *d′*	4	76	0.7877	0.5367	0.8075
	Criterion *c*	4	75	3.6633	0.0088	0.0480
	Real image accuracy	4	102	0.7263	0.5760	0.8075
	Scrambled image accuracy	4	104	0.1733	0.9516	0.9516

Notes: significances (FDR-corrected $p < 0.05$) are highlighted as red.

For sensitivity *d′* and real image accuracy change at 70% DC, the effects remained statistically significant in both baseline conditions ($p_{FDR} < 0.05$). For criterion *c* at 5% DC, the effect was significant in Baseline-2 and marginally significant in Baseline-1 (uncorrected $p = 0.0215$). These findings confirm that our primary results are not dependent on baseline averaging and are not artifacts of reduced variance.

We also note that our original motivation for including Baseline-2 was to evaluate the potential carryover or long-lasting effects of LIFU. As we found no such effects (Supplementary Fig. 3),

averaging the two baselines offers practical and statistical advantages while preserving the integrity of the analysis.

We have now included these control analyses in the supplementary materials to reassure readers that our conclusions are robust to baseline definition.

To ensure that these effects were not artifacts of baseline averaging, we conducted control analyses using each baseline condition separately, showing that the main findings remained robust across both individual baselines (Supplementary Table 3). (Page 8, Line 234)

Question 8:

I don't think the conclusion follows from either the analysis or the results. To explain, the authors want to conclude that there is a "unique role for the matrix-rich transmodal region of thalamus" and that this research "provides a valuable insight into the unimodal transmodal functional organization of thalamus and its relationship with human conscious perception". Later in the discussion, the authors conclude by saying that "these results provide insights into the unimodal transmodal functional organization of the thalamus and its relationship to conscious perception." Yet, that's what the results show, and I'm not even sure that the analysis they did allows for this conclusion. To explain:

With respect to the former, it is hard to draw the conclusion that what matters is matrix rich nuclei that have transmodal connectivity when DA has no effect on visual awareness whatsoever (in fact, it is as different from VA as is the baseline) despite having a similar pattern of cross-modal connectivity to VA and despite also being matrix rich and consciousness-sensitive (as shown in their prior work; reference #20). The only conclusion one can draw from the region specific analysis is that VA only might have an effect on the sensitivity of detection. This, of course, does not match the much more encompassing speculations in the conclusion about cytoarchitectonics and connectivity (which, again, does not fit with the result of DA).

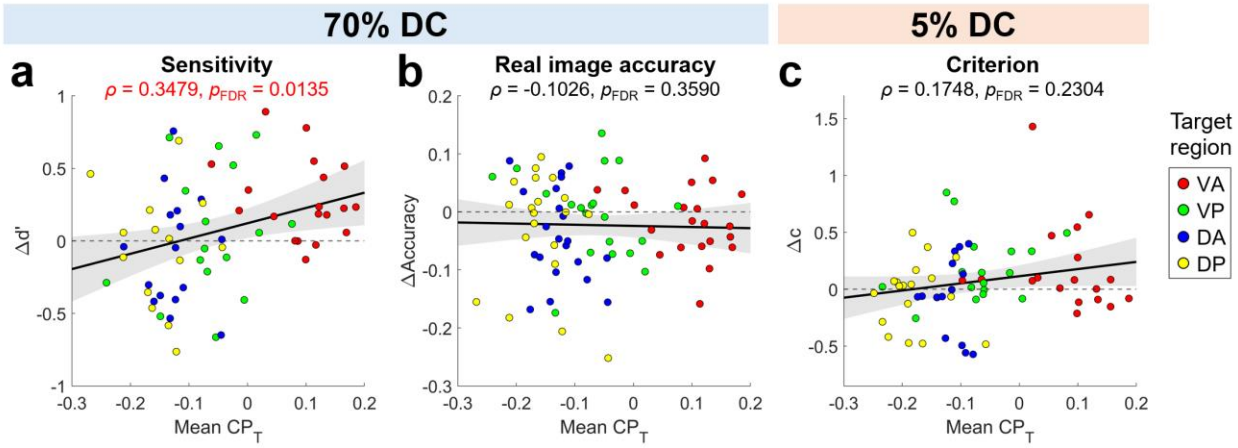
With respect to the latter, I think the whole region invariant analysis is contrary to the intended conclusion. A significant finding there implies the effect is nonspecific, particularly given that none of the findings in the region invariant analysis then have a correspective to show that there is some kind of specificity in the region specific analysis (meaning, none of the findings in the region invariant analysis are matched by any kind of difference between baseline and any of the nuclei, implying the result is likely just due to averaging and dispersion reduction). To be clear, it's perfectly fine to also have region invariant results, they just don't allow making the conclusion that there's anything specific or any dissociation between the nuclei, all the less that this has anything to do with being matrix rich or having transmodal connectivity. If the authors truly wanted to make this case, they could have looked at whether the average amount of matrix rich cells in each nucleus (from their prior work) or the amount of transmodal versus unimodal connectivity assessed here is the thing that correlates with observing an effect of lifu on the task, or something of the sort. In other words, while I believe that VA has had a differential effect (pending the reanalysis with the two baselines), the link from there to the story about cytoarchitectonics and connectivity is not supported by these data.

Response:

We thank the reviewer for this critical and thought-provoking comment regarding the alignment between our results, interpretations, and conclusions. We acknowledge that our original framing may have overextended the implications of our findings, particularly in connecting perceptual outcomes to thalamic cytoarchitectonics and connectivity without sufficient empirical support.

In response to the reviewer's suggestion, we conducted a follow-up correlation analysis to directly test whether the core-matrix cell composition in stimulated regions predicts the perceptual outcomes. For thalamic cytoarchitecture data, we utilized publicly available data of voxel-level relative mRNA expression levels of Calbindin1 (CALB1) and Parvalbumin (PVALB), denoted as CP_T. Since matrix and core cells preferentially expresses CALB1 and PVALB, CP_T can be used as a proxy for matrix-richness at certain thalamic voxel. For each LIFU-ON block, we calculated the mean CP_T value within the beam-present voxels (> 50% relative beam intensity), and correlated this with baseline-subtracted behavioral changes for the three outcomes that survived omnibus ANOVA (i.e., sensitivity and real image accuracy at 70% DC and criterion at 5% DC).

As shown in the Updated Fig. 4 (see below), we observed a significant positive correlation between CP_T and change in sensitivity d' at 70% DC (ρ = 0.3479, p_{FDR} = 0.0135), but no significant correlations for the other two outcomes (p_{FDR} > 0.05). This finding provides the direct empirical link between matrix cell composition and the perceptual sensitivity enhancement observed following thalamic sonication, leading data-driven support to our cytoarchitectural interpretation.



Updated Fig. 4: Correlation between matrix cell richness at beam-present regions (*x*-axis) and behavioral outcome changes (*y*-axis) across LIFU-ON blocks. CP_T reflects the relative difference between Calbindin (CALB1) and Parvalbumin (PVALB) expressions. Positive and negative CP_T values indicate matrix cell-richness and core cell-richness at the beam center (i.e., voxels exceeding 50% relative intensity), respectively. Spearman's ρ and FDR-corrected p-values are provided. Circle colors indicate target regions. Solid lines and shades show linear fits and 95% confidence intervals. VA: ventral anterior thalamus; VP: ventral posterior thalamus; DA: dorsal anterior thalamus; DP: dorsal posterior thalamus.

We agree with the reviewer that target-invariant effects, such as the observed decrease in real image accuracy at 70% DC, should not be used to support claims of target-specific contributions or interpretations involving cytoarchitectonics or transmodal connectivity. In fact, the target-invariant effects were not part of our primary hypothesis, which focused on the unique role of VA based on its matrix-rich profile and transmodal connectivity. These general effects likely reflect a different mechanism, such as interference with early visual cortical input that is shared across multiple thalamic pathways. We have revised the Discussion section accordingly and clearly distinguish target-specific effects (e.g., VA-related enhancement in sensitivity *d'*) from target-invariant effects (may reflect broader disruptions in visual categorization unrelated to thalamic microarchitecture).

We have revised the final paragraph of the Discussion to temper our original claims:

> Unlike the increase in perceptual sensitivity with VA thalamus stimulation, categorization accuracy was reduced with all four thalamic targets at 70% DC (Fig. 6b). This target-invariant effect did not correlate with core-matrix cell composition, suggesting a distinct underlying mechanism. Given the excitatory nature of high-DC LIFU and the shared thalamovisual connectivity among the targeted regions (Fig. 5c), it is plausible that 70% DC sonication consistently co-activated the visual cortex during the LIFU-ON blocks[44]. Such co-activation may have broadly disrupted thalamocortical visual processing, affecting visual cortical networks that integrate inputs from multiple thalamic regions. This interpretation aligns with previous work demonstrating a negative correlation between the pre-stimulus activity of the visual network and categorization accuracy[12]. (Page 14, Line 423)

Question 9:

Similarly, later on in the discussion, the authors say that the nonspecific decrease in category accuracy might be due to the fact that the 70% duty cycle stimulation might have led to consistent co-activation of the visual cortex doing the stimulation. Yet, positive stimulation of DP, which is uniquely connected primarily to visual cortex, didn't have any such effect. In light of this, the interpretation doesn't seem to follow logically from the results. Doesn't the absent finding in DP sonication contradict the point being made here?

Response:

We thank the reviewer for this astute observation. The reviewer is correct that our earlier explanation about nonspecific co-activation of visual cortex was not well-supported by our data, particularly given the absence of effects for DP vs. Baseline comparison in our original analysis.

However, our revised statistical analysis now reveals a more coherent pattern that supports the reviewer's interpretation. In our updated post-hoc results for real image accuracy at 70% DC, we find that all four thalamic regions, including DP, show significant decreases relative to baseline (Table R4). Notably, DP shows the strongest effect among all targets (F = 19.3209).

Table R4: Post-hoc pairwise comparisons against baseline for 70% DC real image accuracy.

Comparison	*F*-value	*p*-value	FDR-corrected *p*-value	Direction
Baseline vs. VA	9.6224	0.0024	0.0119	Baseline > VA
Baseline vs. VP	12.2472	0.0007	0.0049	Baseline > VP
Baseline vs. DA	16.6370	0.0001	0.0012	Baseline > DA
Baseline vs. DP	19.3209	0.0000	0.0007	Baseline > DP

Notes: direction of change is noted for uncorrected $p < 0.05$. Red indicates FDR-corrected $p < 0.05$.

This updated result directly addresses and resolves the logical contradiction pointed out by the reviewer. The absence of effect in DP that previously appeared to undermine our mechanistic interpretation is no longer present in the revised analysis. In fact, DP now provides the strongest evidence supporting the hypothesis that high-duty-cycle LIFU may lead to a generalized disruption of thalamocortical visual processing, particularly through pathways that are more directly connected to visual cortex. Thus, rather than contradicting our interpretation, the DP effect now reinforces the plausibility of the mechanism we proposed.

We are grateful to the reviewer for prompting this careful re-examination, which significantly strengthened both the rigor and internal consistency of our interpretation.

Reviewer #2 (Remarks to the Author):

The authors made a strong effort to address most major points raised in the reviews which is much appreciated. Of particular note, the authors thoroughly addressed the statistical rigor concerns and made substantial revisions that appear to resolve most of the issues. The authors adequately addressed issues including the justification and description of the duty cycles and the integration of the connectivity data.

Response:

We sincerely thank the reviewer for their valuable feedback and for acknowledging our efforts in addressing the important and critical points raised by the reviewer.

Question 1:

The authors somewhat addressed the potential issue of carry-over effects but did not perform more rigorous assessments such as sonication order effect per region which could still be important given potential accumulating or bleeding over effects per region given their small size and close proximity.

Response:

We thank the reviewer for highlighting the importance of systematically evaluating sonication order effects. First of all, we would like to let the reviewer know that we substantially revised our previous statistical framework. In particular, we now implemented a linear mixed-effects model that addresses multiple comparison issues while properly controlling covariates. We fit models using the following specification:

$$Outcome \sim Target + 6\ Covariates + (1|Subject)$$

where *Target* includes baseline (i.e., LIFU-OFF) and four thalamic regions. Therefore, now we could conduct a more comprehensive analysis to assess whether the temporal order of sonication influenced our outcomes.

Since sonication order is included as a covariate in all models within our new statistical framework, it is possible to directly assess its effect while controlling for regional differences. We found no significant sonication order effects across any outcome measures after FDR correction (Table R5), suggesting that sonication order did not systematically influence the results.

Table R5: Omnibus ANOVA testing the effects of the sonication order.

DC	Outcome	df_1	df_2	*F*-value	*p*-value	FDR-corrected *p*-value
70%	Hit rate	1	105	0.0775	0.7813	0.9375
	FA rate	1	106	0.8118	0.3696	0.7858
	Sensitivity *d'*	1	86	1.1433	0.2879	0.7858
	Criterion *c*	1	85	0.3907	0.5336	0.7858

	Real image accuracy	1	106	3.1010	0.0811	0.7858
	Scrambled image accuracy	1	106	0.0050	0.9438	0.9438
5%	Hit rate	1	103	1.6302	0.2045	0.7858
	FA rate	1	103	0.4912	0.4849	0.7858
	Sensitivity *d'*	1	78	0.5356	0.4664	0.7858
	Criterion *c*	1	77	0.2938	0.5893	0.7858
	Real image accuracy	1	103	1.3528	0.2475	0.7858
	Scrambled image accuracy	1	105	0.0201	0.8875	0.9438

As a complementary analysis, we also conducted a separate testing, treating sonication order as the primary independent variable, using the model:

$$Outcome \sim Sonication\ order + Other\ 5\ Covariates + (1|Subject)$$

ensuring that information about the target region was not included in the model. This approach tested systematic linear trends across the stimulation sequence, independent of regional identity. Again, we observed no significant effects after multiple comparison correction, reinforcing that sonication order did not meaningfully influence task performance (Table R6).

Table R6: Omnibus ANOVA with sonication order as primary predictor.

DC	Outcome	df_1	df_2	*F*-value	*p*-value	FDR-corrected *p*-value
70%	Hit rate	1	110	0.0192	0.8900	0.8900
	FA rate	1	114	0.7736	0.3810	0.8764
	Sensitivity *d'*	1	93	0.3015	0.5843	0.8764
	Criterion *c*	1	93	0.3984	0.5294	0.8764
	Real image accuracy	1	109	1.0594	0.3056	0.8764
	Scrambled image accuracy	1	116	0.1114	0.7391	0.8900
5%	Hit rate	1	104	6.7001	0.0110	0.1322
	FA rate	1	104	0.6560	0.4198	0.8764
	Sensitivity *d'*	1	79	1.0444	0.3099	0.8764
	Criterion *c*	1	78	0.3140	0.5769	0.8764
	Real image accuracy	1	104	0.0659	0.7979	0.8900
	Scrambled image accuracy	1	106	0.0267	0.8704	0.8900

These comprehensive analyses provide reassuring evidence that our findings are robust to potential temporal confounds and that the observed perceptual effects reflect genuine regional differences in thalamic function rather than artifacts of stimulation sequence. We hope these results address the reviewer's concerns about order effects and reinforce confidence in the specificity of our findings.

Question 2:

The authors also partially addressed the targeting and intensity issues. While statistical regression of confounds is a good step there is still a lack of individualized skull modelling which remains a significant limitation of this study and should be explicitly acknowledged.

Some other small issues still remain including the manuscript still presents stylized beam data (and only free water data) and not individual specific data which weakens any precise claims; there is still no indication that individual anatomical MR were used which is a major limitation for thalamic targeting and there is still insufficient detail on whether the transducer offset or orientation was individually adjusted for each participants' anatomy. Overall, the authors did a good job to address most major points. However, the biggest unresolved limitation remains the lack of subject-specific beam modeling and skull compensation, which undermines the claim of precise region-specific effects especially for individual thalamic nuclei that are small, deep and in close proximity. This may not invalidate the results entirely, but it weakens causal claims about localization and mechanism.

Response:

We appreciate the reviewer's point regarding individualized skull modeling. While we addressed beam targeting variability through post-hoc simulation-based estimation of beam intensity and included this as a covariate in our statistical model, we acknowledge that the lack of subject-specific skull acoustic modeling is a significant limitation. This limitation may have introduced unmodeled variability in beam propagation and targeting precision, which could influence both the magnitude and localization of neuromodulatory effects.

We have now explicitly acknowledged this limitation in the Discussion section of the revised manuscript and noted that future studies should incorporate subject-specific skull modeling (e.g., via CT- or MRI-based acoustic simulations) to improve the precision and reproducibility of LIFU targeting in subcortical structures.

> Our study has limitations. First, although the template rescaling approach described above allowed us to effectively approximate each participant's thalamic coordinates, we did not create subject-specific acoustic models or apply skull-dependent phase corrections. The beam profiles therefore cannot capture the focal shift, peak-pressure drop, or off-axis energy leakage introduced by each participant's skull morphology. This lack of individualized skull modeling represents a limitation that may have introduced unmodeled variability in beam propagation and targeting precision, potentially affecting both the magnitude and localization of the observed neuromodulatory effects. Likewise, while the transducer was positioned with an optical neuronavigation system, its final offset and tilt were not iteratively refined based on native MR or computed tomography images. This limitation further compromises targeting precision and should be addressed in future investigations through individualized anatomical imaging or automated transducer positioning protocols. (Page 16, Line 483)

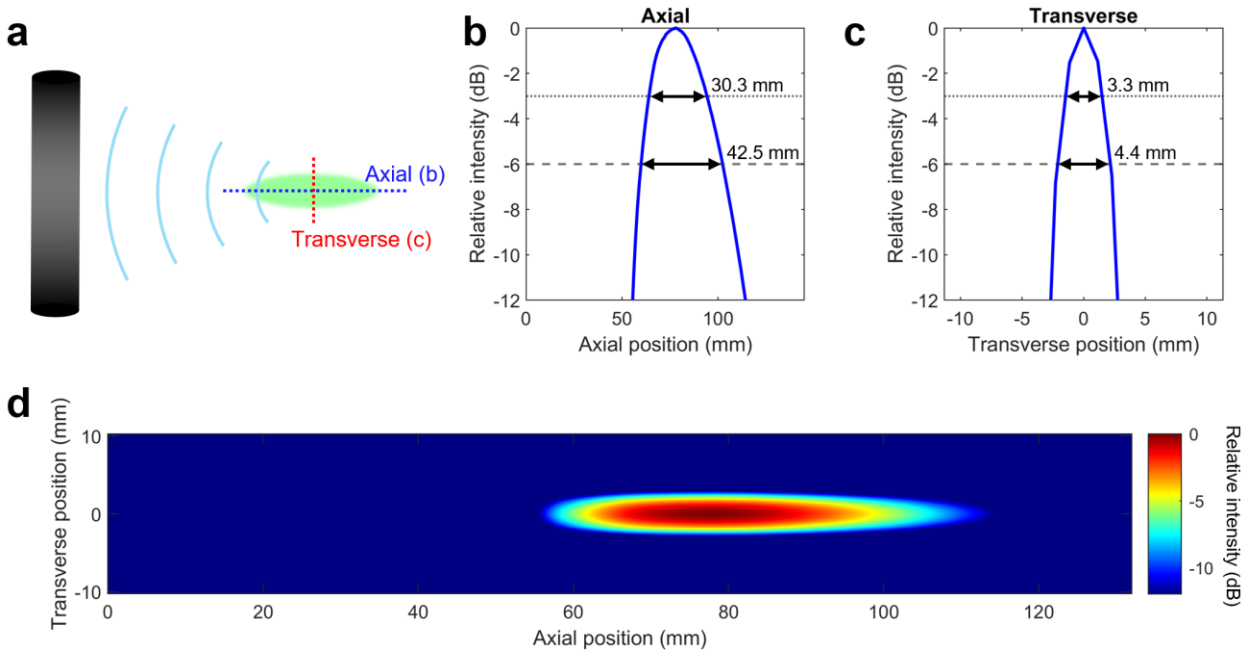
Reviewer #3 (Remarks to the Author):

Question 1:

While the free-field acoustic beam profile shown in Supplementary Fig. 5 provides useful baseline information, it would be more intuitive and informative to present it as a two-dimensional (2D) axial-lateral intensity map rather than as separate 1D line plots.

Response:

We have updated Supplementary Fig. 5 to present the beam profile as a 2D map, derived from the measured axial and transverse intensity data.



Updated Supplementary Fig. 5: Free-field acoustic profile of the transducer. a Schematic illustration of intensity measurements at planes relative to the transducer. **b, c** Relative beam intensity across **b** axial and **c** transverse directions. Dotted and dashed lines indicate -3 dB and -6 dB (equivalent to 50% and 25% of the initial intensity). **d** Reconstructed two-dimensional (2D) axial-transverse intensity map of the acoustic beam profile.

Question 2:

Additionally, when defining the acoustic focus, it is standard practice to use the full width at half maximum (FWHM) of the intensity distribution, corresponding to the -3 dB boundary where intensity exceeds 50% of the maximum value, rather than the -6 dB (25% intensity) threshold currently applied. According to the ITRUSST practical guide (Murphy et al., 2025, Clinical Neurophysiology), the FWHM of the intensity profile should be used to define focal dimensions in transcranial ultrasonic stimulation studies. Therefore, we suggest reanalyzing and presenting the

focal dimensions based on the FWHM criterion to align with established standards. Similarly, in Fig. R11, when visualizing beam-present regions, it would enhance technical rigor and clarity if the focus boundaries were defined and visualized according to the FWHM standard.

Response:

Thank you for drawing our attention to the current convention of using -3 dB to define the acoustic focus. Following your recommendation and citing the suggested reference, we have revised all analyses to -3 dB standard. Specifically, Fig. 2j-m has been fully recomputed with the -3 dB criterion. Accordingly, the Updated Supplementary Fig. 8-10 now displays the raw intensity profiles together with the new -3 dB contours for full transparency (see Fig. R1 below for an example). We gave additional change to Fig. 2f-i by showing the group-averaged relative intensity maps rather than simple binary overlap. The outcomes and overall conclusions are unchanged. We believe these revisions improve technical rigor and clarity while following the established standards.

To quantify this further, we computed a coverage ratio defined by how much of the “beam-present” volume (defined by > 50% intensity) belongs to the thalamic (or other non-thalamic) regions (Fig. 2j-m). (Page 6, Line 197)

The spatial coverage of each beam was determined as the percentage of “beam-present” voxels (defined as voxels exhibiting more than 50% of the relative intensity[ref: Murphey et al., 2025]) that fell within specific thalamic (or other non-thalamic) regions. (Page 19, Line 639)

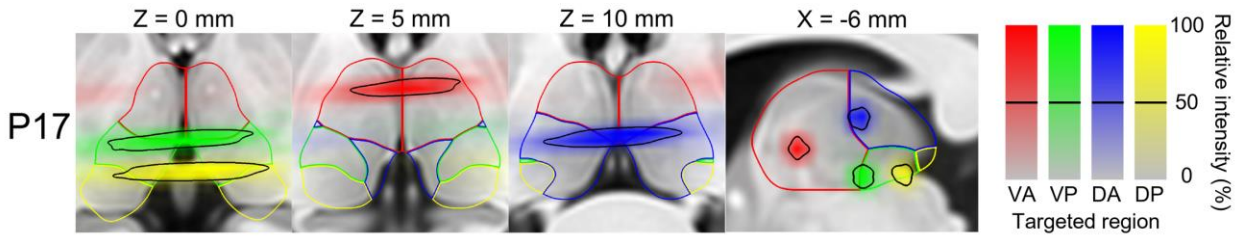


Fig. R1: An example of the updated individual mosaic views of the beam trajectories. Black contours indicate 50% (-3 dB) relative intensity. Color coding for the thalamic regions is as follows: VA (red), VP (green), DA (blue), and DP (yellow) thalamus.

RESPONSE TO REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

I sincerely thank the authors for having patiently gone along with my sometimes "detailed" feedback. This latest revised version is very strong methodologically and in terms of the connect between the specific statistical tests conducted and the conclusions drawn. I think this is work that the authors and the field will be very proud of.

Response:

We sincerely thank the reviewer for their thoughtful and detailed feedback throughout the revision process. We greatly appreciate the time and care invested in improving our manuscript, and we are grateful for the recognition of our work. Your insights have significantly strengthened the clarity and impact of our work.

Reviewer #3 (Remarks to the Author):

While the authors present a well-structured behavioral study, the methodological pipeline for targeting deep brain regions—particularly thalamic subnuclei—has inherent precision limitations that should be acknowledged more explicitly. Specifically, the use of pseudo-anatomical MR images derived from linearly rescaled MNI templates does not account for individual variability in skull morphology and deep brain anatomy. The scalp-surface–based neuronavigation registration (fiducial + ICP) introduces additional alignment errors that may accumulate at subcortical depths. Furthermore, acoustic beam modeling was performed on segmented pseudo-MRI using homogenized skull properties, without subject-specific CT or MRI-based modeling, thus failing to capture individual focal shifts and intensity variations. Given the close proximity (3–6 mm) of thalamic subnuclei and their deep location, these factors—combined with single-focus sonication without adaptive beam correction—can result in cumulative targeting errors of several millimeters. This limits the strength of claims regarding precise region-specific neuromodulatory effects at the individual level.

Response:

Thank you for this important comment. We agree that individualized targeting accuracy is a critical methodological consideration for deep brain neuromodulation studies. In response, we have expanded the discussion of these limitations in the revised manuscript. Specifically, we now acknowledge that the use of linearly rescaled MNI templates and scalp-based neuronavigation introduces subject-dependent variability in focal positioning, especially at subcortical depths. The revised text now clearly emphasizes that these factors can limit the precision and reinforce the need for individualized modeling and adaptive beam correction in future work.

Our study has limitations. First, although the template rescaling approach allowed us to approximate each participant's thalamic coordinates, it did not incorporate individualized skull modeling or subject-specific acoustic corrections. The beam profile was estimated on pseudo-anatomical MR images derived from

linearly rescaled MNI templates with homogenized skull properties, without accounting for interindividual variability in skull morphology, acoustic impedance, or deep brain anatomy. Therefore, the modeled beam cannot capture the focal shift, peak-pressure drop, or off-axis energy leakage introduced by skull heterogeneity. (Page 10, Line 390)

Scalp-contact-based neuronavigation introduces residual alignment errors that may accumulate with depth, especially when targeting deep brain areas. This limitation further compromises targeting precision and should be addressed in future investigations through individualized anatomical imaging or automated transducer positioning protocols. (Page 10, Line 401)