Reply to the Recommender's / Reviewers' Comments – Round 4

We would like to thank the recommender and the reviewer once again for their time and comments on the revised manuscript. We have addressed each point in italics below, with additional text indicated with underline.

MAJOR

1. Missing focus on "finger/hand" resizing (intro)

• Para. 1 (l. 40-54) appears too general, i.e. the manuscript's core theme "finger/hand illusion" gets lost. Focus perhaps on hand/finger resizing illusions in para. 1 and add 1 sentence stating that such resizing illusions apply to other body parts too? Similarly, in other places, it is not always clear when the reported evidence refers to "finger/hand illusions" and when not.

The core theme of the manuscript is clearly laid out in the final paragraph of the introduction, therefore we do not see a need to change the language of the first paragraph. We have however, updated references to illusory resizing throughout the introduction to reflect whether they refer to finger or hand based resizing illusions. Examples can be seen within paragraph 2:

"The augmented reality system used to deliver these resizing illusions presents real-time video capture of the hand, from the same position and perspective as if the hand were being viewed directly (Preston & Newport, 2011). This allows the experimenter to deliver tactile manipulations, such as gently pulling or pushing the hand/finger, whilst the participant views their hand/finger either stretching or shrinking in the augmented image. Newport, Pearce and Preston (2010) found strong embodiment using a synchronous multisensory visuotactile illusion, which was replicated in our pilot data using the same experimental set up as the current study, showing a greater illusory experience during synchronous visuotactile manipulations compared to asynchronous (mismatching visuotactile manipulation) control conditions (Appendix B) for illusory finger resizing."

2. References cited in text not listed in reference section

• There are unlisted references (l. 44, 54, 59, 368, 369, 371), such as Preston et al. (2020), Hansford et al. (2023), Newport, Pearce, and Preston (2010), Muller-Putz et al. (2001), and Breitweiser et al. (2016).

Thank you for spotting these, most were accidentally removed when removing the chronic pain narrative. All references have been cited correctly in text and added to the reference list if missing.

3. Reported difference for async. vs sync. condition although there is none (intro, pilot data)

• In the intro (l. 61-63) and Appendix B, it is stated that the pilot data show a greater illusion strength for a synchronous vs an asynchronous condition, although Appendix B reports no significant differences, creating great confusion. Moreover, the data are collapsed across healthy and chronic pain individuals, which does not fit in with the current focus of the manuscript.

The pilot data are being referred to as being numerically greater in the synchronous condition compared to the asynchronous condition, with no implications of significant differences – this has been made clearer with the additions to the intro as can be seen below:

"…our pilot data using the same experimental set up as the current study, showing, although not significant, a numerically greater illusory experience during synchronous visuotactile manipulations compared to asynchronous (mismatching visuotactile manipulation) control conditions (Appendix B) for illusory finger resizing."

Pilot data in Appendix B has also been updated to show healthy participants only:

"Pilot data regarding the experience of the illusion for healthy participants undergoing synchronous and asynchronous illusory resizing of the index finger can be seen in figure B1. 9 participants had either synchronous or asynchronous multimodal manipulations delivered first in a random order, and were then given the other condition, after which all participants were given an illusion scale. Findings showed that across all participants, no significant difference in illusion experience between the synchronous and asynchronous conditions, t(8) = 1.877, p = 0.097, however as can be seen in figure B1, despite the small sample size, illusion strength was seen to be greater in the synchronous condition compared to the asynchronous condition.

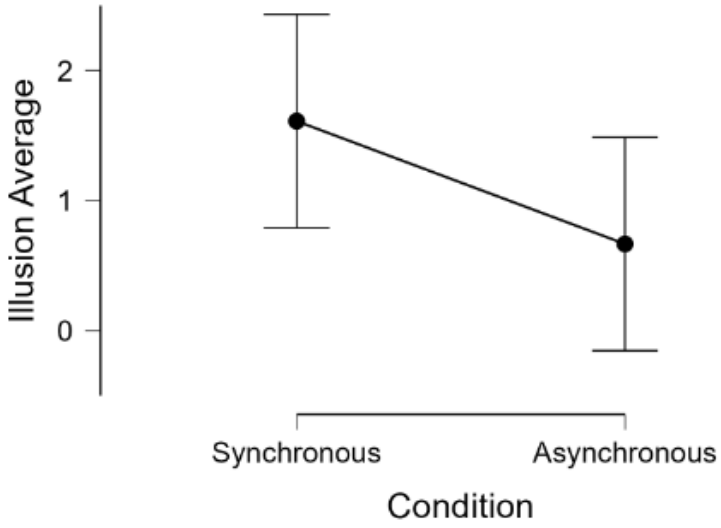

Figure B1. Pilot data from Healthy Participants Undergoing Synchronous and Asynchronous Illusory Finger Resizing."

4. Circular data analysis not explicitly flagged as such (intro)

• In para 2 (l. 66-70), it needs to be outlined much more clearly that this subset analysis is circular and that the found difference cannot be distinguished from a statistical artifact. Moreover, a replication per se is of course not helpful; it is only helpful if this replication omits circularity.

We have already addressed this issue in the previous round of revisions and clearly state that the analysis is from a subset, that this was of a small sample size, and was selected based on measures already analysed.

5. Missing rationale for inclusion of 2 control conditions (intro)

• In para. 6 (l. 126-139), it remains unclear why it is important to have 2 control conditions. Something along these lines is already mentioned in the methods section (l. 227-229), which I think should be moved to the intro. Similarly, it might be sensible to mention comparisons to such control conditions already in para. 2 (l. 55-73), not least because such comparisons are reiterated in l. 316 (where references are missing).

The rational for both control conditions has been moved from the method section to the introduction as can be seen below:

"To test this, different finger resizing illusions consisting of multisensory (visuotactile) stretching (MS), unimodal-visual stretching (UV), a non-illusion control condition without tactile input (NI), and a non-illusion control condition with tactile input (NIT) will be used to assess alternate aspects of illusory resizing manipulations and their related effects on SSEP response. The inclusion of two control conditions (NI, NIT) is to assess whether localisation of cortical representations arise from resizing manipulations to the finger, or from tactile input given to the finger."

A note regarding comparisons to NI conditions in previous work has been added to paragraph 2:

"When comparing multisensory visuotactile resizing illusions to unimodal visual resizing illusions, our recent work (Hansford et al., 2023a) shows that multisensory illusions elicit significantly greater illusory experience compared to non-illusion and unimodal visual illusion conditions in healthy participants."

References have also been added to the planned analyses section for Hypothesis 1:

"The subjective illusory experience questionnaire will be used as a positive control for the current study. Previous research has shown significantly greater illusion strength for MS conditions compared to NI conditions (Carey et al., 2019; Hansford et al., 2023a), which we will attempt to replicate."

6. Confusing usage of term "NI conditions"

• There are 2 control conditions, NI and NIT. As such, there is only a single NI condition. The whole manuscript including Table A1 repeatedly refers to NI conditions (plural). Moreover, when stating hypothesis 1 (l. 312- 313) singular is used ("NI condition"), although plural seems to be intended. To prevent confusion, the text and Table A1 need to refer to the NI and NIT condition whenever both these conditions are of relevance.

The text regarding the non-illusion conditions has been updated to reflect when either on or both conditions are implied, and hypothesis 2 has been updated to reflect an expectation of a significant difference when comparing all 4 conditions within a repeated measures one way ANOVA, with hypotheses 2a-2c reflecting predictions about differences in post hoc tests:

Intro final paragraph: "The first hypothesis, acting as a positive control (1), is that there will be a greater illusory experience, measured via a subjective illusory experience questionnaire, in the MS condition compared to the NI and NIT conditions. The main experimental hypothesis for this study is that (2) there will be a significant difference in SSEP response across the electrodes of interest (F1 & FC1, see section 3. Pilot Data) when comparing across all conditions. Subsequent hypotheses are that there will be significant differences in SSEP response when comparing (2a) MS visuotactile illusory resizing to the NI condition, when comparing (2b) the UV illusory resizing to the NI condition, and (2c) that there will be no significant difference when comparing the NIT condition to the NI condition.

Planned analysis hypothesis 1: ". Given significant findings, post-hoc tests will be run, with Bonferroni correction for 2 comparisons (MS Vs NI, MS Vs NIT) at an initial alpha of 0.05.*"*

Planned analysis hypothesis 2: "Given significant findings in the ANOVA, post hoc comparisons of every condition will be conducted at a new alpha of .01 (corrected for 6 comparisons (MS Vs NI, MS Vs NIT, MS Vs UV, UV Vs NI, UV Vs NIT, NI Vs NIT))."

Design planner interpretations:

Hypothesis 1:

"If Hypothesis 1 is unsupported

(No significant difference comparing MS to both NI and NIT): Indicates that the augmented reality manipulations are not inducing effective illusions, and therefore the findings regarding hypotheses 2 will be called into question.

(No significant difference comparing MS to NI): Indicates that tactile input is needed for an effective non illusion condition, calling EEG analyses with NI condition into question.

(No significant difference comparing MS to NIT): Indicates that tactile input removes effect of NIT condition, calling EEG analyses with NIT condition into question."

Hypothesis 2:

"If Hypothesis 2 is supported: Indicates that there are significant differences in SSEP response when comparing across all conditions in healthy participants.

If Hypothesis 2 is unsupported: Indicates that there is no evidence of a significant difference in SSEP response when comparing across all conditions in healthy participants. This will result in hypotheses 2a – 2c being unsupported.

Hypothesis 2a:

"If Hypothesis 2a is supported: Indicates that there are significant differences between MS and the NI condition in healthy participants.

If Hypothesis 2a is unsupported: Indicates that there is no evidence of a difference between MS and the NI condition in a healthy population."

Hypothesis 2b:

"If Hypothesis 2b is supported: Indicates that there are significant differences between UV and the NI condition in healthy participants.

If Hypothesis 2b is unsupported: Indicates that there is no evidence of a difference between UV and the NI condition in a healthy population."

Hypothesis 2c:

"If Hypothesis 2c is supported: Indicates that there are no significant differences between NI and NIT conditions in healthy participants.

If Hypothesis 2c is unsupported: Indicates that there is evidence of a difference between NI and NIT conditions in a healthy population."

7. Inconsistencies b/w hypotheses, sampling plan, and analyses in text and Table A1 + lack of clarity

Stated hypotheses

• Just like for hypothesis 1 and Table A1, condition acronyms should be used for hypothesis 2 (l. 327-329).

Corrected: "There will be a significant difference in SSEP response across the electrodes of interest (F1 & FC1) when comparing (2a) the MS condition to the NI condition, when comparing (2b) the UV condition to the NI condition, and (2c) there will be no significant difference when comparing the NIT condition to the NI condition."

Sampling plan (power analysis) - hypothesis 1

• Text mentions 4 measurements (i.e. conditions), 5 participants, and a power of 90% (l. 352) and Table A1 3 measurements, 4 participants, and a power of 80%.

*Text has been corrected in the table: "A priori power analysis using G*Power shows that for a repeated measures, within factors ANOVA, with an effect size (f) of 0.73, alpha of 0.05, power at 90% and 1 group with four measurements, 5 participants are needed."*

Sampling plan (power analysis) - hypothesis 2

• If "1 group" is indicated for hypothesis 1 (l. 352), it should also be indicated for hypothesis 2 in the text (l. 360-362). This information should also be added to Table A1.

*Corrected: "A priori power analysis using G*Power shows that for a repeated measures, within factors one way ANOVA, with an effect size (f) of 0.25, alpha of 0.05, power at 90%, and 1 group with four measurements, a total sample size of 30 participants is needed." – also corrected in table.*

• Text mentions a power of 90% (l. 361) and Table A1 80%.

Table has been corrected to power at 90% and the sample size needed updated to 30 participants in the table:

"A priori power analysis using G*Power shows that for a repeated measures, within factors one way ANOVA, with an effect size of f = .25, alpha of 0.05, power at 90%, and 1 group with four measurements, a total sample size of 30 participants is needed."

• Text mentions 4 measurements (l. 361), whereas the number of measurements is not indicated in Table A1. More importantly, Table A1 lists 2 ANOVAs involving 3 conditions each and thus 3 measurements.

Text and table have been updated to show that hypothesis 2 regards a repeated measures ANOVA with 4 measurements, and subsequent hypotheses are now post hoc comparisons between conditions, resulting in 1 ANOVA overall:

See Table A1.

• Text mentions 30 participants (l. 143, 145, 362) and Table A1 24.

30 participants is correct and this has been updated In table A1.

• Just like the text (l. 327), Table A1 should refer to electrodes of interest for hypothesis 2a.

Electrodes of interest have been added to all needed sections in Table A1:

Example from hypothesis 2a: "(2a) There will be a significant difference in SSEP response across the electrodes of interest (F1 & FC1) when comparing the MS illusory resizing condition to the NI condition."

Analyses

• If 2 ANOVAs will be performed for hypothesis 2a and 2b, this needs to be indicated in the main text too.

This has been corrected to one ANOVA being run for hypothesis 2, and post hoc comparisons being run for hypotheses 2a – 2c:

Text update: "These amplitudes will then be averaged across trials to give overall results for each participant, before running a repeated measures one way ANOVA comparing SSEP response from each experimental condition. The dependent variable will be SSSEP amplitude in µV, whilst the independent variable will be the different manipulations given in each comparison condition. Given significant findings in the ANOVA, post hoc comparisons of every condition will be conducted at a new alpha of .01 (corrected for 6 comparisons (MS Vs NI, MS Vs NIT, MS Vs UV, UV Vs NI, UV Vs NIT, NI Vs NIT))." – also reflected in Table A1.

8. Multiple comparisons issues not accounted for + missing rationale for two ANOVAs (hypothesis 2a/b)

Multiple comparisons have now been accounted for as can be seen in the text above, and we are no longer planning on using two ANOVAs – just one for hypothesis 2 and then subsequent post hoc comparisons for hypotheses 2a – 2c given significant findings for hypothesis 2.

• Table A1 lists 2 ANOVAs, one for hypothesis 2a and one for hypothesis 2b. Why are two separate ANOVAs needed? Due to the inclusion of the NI and NIT condition in both ANOVAs, they are not independent. As such, issues of multiple comparisons arise that remain currently unaccounted for.

1 ANOVA is now planned for hypothesis 2 and post hoc tests will be used for hypotheses 2a – 2c as can be seen in the comments above and in the revised Table A1.

9. Incorrectly converted effect size (hypothesis 1)?

• For hypothesis 1 (l. 345), a Cohen's f of 0.73 is reported. However, using the reported η 2 of 0.33 and the following conversion formula f = sqrt(η 2 /(1- η 2)), I obtain a Cohen's f of 0.70.

The effect size has been updated to be a cohen's f of 0.70:

*"A priori power analysis using G*Power for the smallest effect size of interest (f = .70) shows that for a repeated measures, within factors one way ANOVA, with an effect size (f) of 0.70, alpha of 0.05, power at 90% and 1 group with four measurements, 6 participants are needed."*

10. Selective analysis or no selective analysis?

• The response letter states that no selective analysis based on condition UV will be performed for hypothesis 2b. It is thus very confusing that the text still talks about "effective" UV illusory resizing (l. 138, 329) and states that such an analysis will be performed (l. 320-324).

These are fossils from the previous round of revisions and have therefore been removed. No subsample analyses will be run for the UV condition based on subjective illusion score data, we will

still assess correlations between conditions in hypotheses 2a – 2b if we do not find a significant difference between the conditions, as mentioned in the last round of revisions, but as mentioned before, this is not referred to in the manuscript at this stage.

11. Confusing usage of term "median illusion score" and "illusion index"

• It is stated that an illusion index will be calculated (l. 305-308). To this end, the median control score will be subtracted from the median illusion score. Will the illusion index be used as a dependent variable in the ANOVA? If so, it is very confusing that the manuscript states that median illusion scores will be used for the ANOVA (e.g., l..318 and Table A1).

The normalised (baseline corrected) data will be used in the ANOVA, this has been made clearer in text:

Preprocessing: "The median control scores will be used to create an index of the illusion and disownership scores by subtracting the median control score from the median illusion and median disownership scores, in line with previous research doing similarly (Matsumiya, 2021; Kilteni & Ehrsson, 2017; Kalckert & Ehrsson, 2012). The normalised (baseline corrected) data will be used for analyses."

Planned analyses: "A one-way ANOVA will be run to compare the dependent variable of normalised (baseline corrected) illusion score from each independent condition."

12. Mismatch b/w hypotheses and planned posthoc tests (hypothesis 1)

• Hypothesis 1 refers to MS vs NI and MS vs NIT (l. 312-313). It is thus unclear why 4 posthoc tests (l. 319) should be performed. Is there a hypothesis missing?

The text has been updated to reflect hypothesis 1 only concerning MS comparisons:

"Given significant findings, post-hoc tests will be run, with Bonferroni correction for 2 comparisons (MS Vs NI, MS Vs NIT) at an initial alpha of 0.05."

13. Sufficiently powered SSSEP responses (pilot data)?

• As far as my understanding of the preprocessing steps goes (l. 285-299, l. 330-33), the amplitude at a frequency of 26 Hz averaged across trials for a given condition will be used for the ANOVA, so 1 amplitude value per participant per condition. Wouldn't such an analysis presuppose a clear SSSEP response at a frequency of 26 Hz for each participant in each condition, as anything else would be just noise? If so, why does Figure 3 show data averaged across all conditions and 3 participants? Moreover, although averaging across 2 electrodes is planned, participants will be included if only data for one is available (l. 280). As such, a clear SSSEP response at 26 Hz for each participant in each condition for each electrode is needed, no?

The pilot experiment was only used to determine that the proposed stimulation frequency was feasible. We have confirmed that SSSEP signals are measurable for each condition and participant in our pilot data. However, our main purpose behind including the pilot data here is to demonstrate that there is overall a measurable signal, so we have retained the figure showing the overall mean.

MINOR

Intro

• l. 68: What is meant by "incongruent"?

Additional information has been provided: "incongruent (mismatching visual and tactile inputs) control condition."

• l. 78-80: The part about the peripheral stimulus raises more questions than it answers. Simply drop?

Removed.

• l. 82: "directly impacts the neural representation of the body". I continue to have great trouble imaging how how EEG can reveal such direct impacts. Tone down?

Reworded: "recent research suggests that these illusions impact the neural representations of the body and reflect early-stage multimodal stimulus integration through modulation of gamma band activity (Kanayama et al., 2021)."

• l. 84-87: Given that Kanayama et al. used EEG too, it is a little odd that the text refers to EEG only after describing Kanayama et al.'s findings.

Reworded: "We have recently also investigated this illusion in healthy participants using electroencephalography (EEG)"

• l. 100-102: I continue to have great trouble understanding the reported research finding by Giani et al. (2012). Re " […] within modality stimulation […]" → Was it temporal congruency that was manipulated here? Re " […] in contrast to Nozaradan et al. […]" → How can it be "in contrast to" if one study tackles sensory integration between and one within the senses; this seems rather "complementary"?

Reworded: "Research has also found evidence of enhanced steady-state responses for within-modality stimulation of auditory and visual stimuli in isolation (Giani et al., 2012), complementing Nozaradan et al.'s findings regarding visuo-auditory combination."

• 1. 103-104: "[…] greater increases in steady-state response magnitude when this corresponds with the amplitude modulation rate […]." Unclear; what correspond with what?

Reworded: "Research using vibrotactile stimulation has found increases in steady-state response magnitude corresponding with the amplitude modulation rate of stimulation (Colon et al., 2012; Rees et al., 1986) suggesting an entrainment of oscillatory activity to temporal features of sensory stimulation (Timora & Budd, 2018)."

• l. 114: Add reference?

Reference added: "It has been suggested chronic pain involves cortical misrepresentations of the size of the affected body part (Boesch et al., 2016)"

• l. 121: Acronym SSSEPs not introduced.

Added: "However, importantly, there has also not been research conducted using somatosensory steady state evoked potentials (SSEPs) in healthy participants"

Other

• 2.3 Experimental Procedure: It needs to be explicitly mentioned that right fingers will be used.

Added: ". Participants will be instructed to place either their right index or middle finger outstretched onto the felt"

• l. 144: "recruited" → "tested"?

Corrected.

• l. 198-201: Not clear what is meant by "central area" and "distance between the central area with the felt base and the area from the mirror to the screen".

*Reworded: "*Within the self-built system there is a 1920 x 1080-pixel Spedal Webcam Wide Angle Camera at the edge of the black felt on the side the participant sits, away from the participant's view."

"The thickness of section on which the mirror sits is 2cms."

• Figure 2/l.235: Caption could be clearer as "tactile input" in the figure means "just touching".

Reworded: "Infographic of Experimental Conditions. MS = Multisensory Stretching, UV = Unimodal Visual Stretching, NIT = Non-Illusion Tactile, NI = Non-Illusion. During the manipulation phase (2.4 seconds) the visual image of the finger is stretched in the MS and UV conditions, and/or the experimenter provides tactile input (touch) in the MS and NIT conditions. The tactile input in the MS condition is accompanied by pulling. During the habituation phase (2.4 seconds) participants are free to move their finger. The arrow denotes the direction of the experimenter's action. The vibrotactile stimulator is depicted on the finger in each phase of the experiment as vibrations are presented throughout."

• l. 249, 251: "pulling"/"pull"/"pulls"→ "pulling/pull/pulls or just touching/touch/touches"?

Reworded: "The experimenter will be seated opposite the participant, the other side of the augmented reality machine and will touch the digit during MS and NIT conditions by holding onto the distal interphalangeal joint and gently touching (NIT) or pulling (MS) the finger whilst the participant keeps their hand in place"

• l. 265: "that condition" → "condition presented in a given block"?

Reworded: "Finally, at the end of each block, the participant will be asked to complete the subjective illusory experience questionnaire regarding a condition presented in a given block using the Samsung Galaxy Tab A6 tablet via a questionnaire on Qualtrics (Qualtrics, Provo, UT)."

• l. 267: "trial" → "trials"?

Changed.

• l. 289-297: Re preprocessing steps, a few things could probably be clearer:

o It would be good to add that "across all standard errors" refer to all standard errors across participants.

Reworded: "Across the standard errors for all participants"

o It would be good to add that the averaging of the signal across the electrodes of interest and calculation of the Fourier transform for each trial is done for each participant separately.

Reworded: "Primary analysis of the remaining EEG data will then involve averaging the signal across the electrodes of interest (F1 and FC1), and calculating the Fourier transform for each trial per participant"

o It would be good to add that for each participant, the amplitude for each trial at the vibration frequency of 26 Hz is derived and then averaged across all trials of each condition. Some of this is outlined in l. 330- 333, which I think should be moved to 2.4.1 Preprocessing steps to have everything in one place.

Added: "These amplitudes will then be averaged across trials to give overall results for each participant. Statistical comparisons will then be performed on the Fourier amplitudes at the stimulation frequency (26Hz), across conditions and participants."

• l. 304: "per trial" → "per condition"?

Changed.

• l. 300-308: No need to reiterate the range of the scales for the illusion, disownership, and control questions. Instead, the range of the calculated indices needs to be explained (lb: 0-100=-100; ub: 100-0= +100).

Reworded: "Regarding questionnaire data, scores for both illusion experience questions will be combined to give median scores, along with both disownership questions and both control questions, resulting in 3 median scores per trial per participant. The median control scores will be used to create an index of the illusion and disownership scores by subtracting the median control score from the median illusion and median disownership scores, in line with previous research doing similarly (Matsumiya, 2021; Kilteni & Ehrsson, 2017; Kalckert & Ehrsson, 2012). The normalised (baseline corrected) data will be used for analyses, with a new scale from -100 to +100 with 100 indicating strongly agree, 50 indicating a neutral opinion, and scores below 0 indicating strongly disagree with the statements on the questionnaire."

• l. 319: "MS / NI" etc. → "MS vs NI" etc.?

Changed for both hypotheses post hoc tests.

• l. 342-343: The response letter states that the effect size calculation involved control conditions too. This should be added to the text, as one expects this in light of hypothesis 1 (MS vs NI, MS vs NIT).

Reworded: "Effect sizes are determined by research from Hansford et al. (2023a) using the subjective illusory experience questionnaire and comparing MS, UV, and incongruent finger-based resizing illusions to control conditions with no illusory resizing, using the same finger stretching illusions and the same equipment (n = 48), which show an effect size of η 2 = .33 (converted to a Cohen's f = .70)."

• l. 344: n 2 → η 2 (eta squared)?

Corrected.

• l. 390, 403: The text refers to illusory experience not being affected by vibrotactile input. This claim appears too strong, as it requires a comparison between vibrotactile vs no vibrotactile input. Tone down?

Reworded: "This therefore shows that the experience of illusory resizing is maintained when vibrotactile stimulation is added to the procedure"

• Figure 3-Scalp topography: Add saturation bar or explain in caption what saturation levels mean.

A saturation bas has been added to the figure along with notes regarding this within the caption:

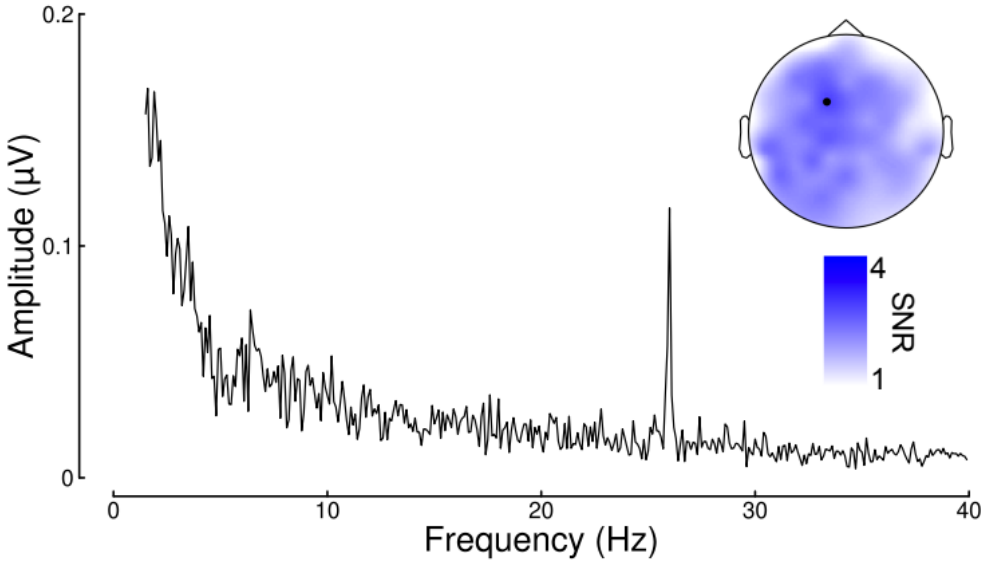

Figure 3. Averaged Pilot Data showing peak frequency at 26Hz, centred between electrodes F1 and FC1. The spectrum is derived from electrode FC1. Saturation bar represents signal to noise ratio (SNR).