Reply to the Recommender's / Reviewers' Comments – Round 3

We would like to thank the recommender and the reviewers again for their time and comments, and have addressed each point in italics below, with additional text indicated with underline (please note: all excerpts of text are from the final version of the document, and therefore will include revisions relating to all reviewers' comments, not just the recommended revisions from a single reviewer):

Recommender:

Your Stage 1 manuscript has now been reviewed by two of the original reviewers. While they both appreciate your hard work addressing some of their concerns, a number of fundamental issues remain. There is consensus between reviewers that the proposed experiment is very complex and that the introduction is too long and could be more focussed. I concur with both these points.

After conversation with the recommender via email, we have decided to reduce complexity in the proposed study by removing the chronic pain group of participants, and we now just plan to test a healthy group of participants to provide a basis for understanding resizing illusions in healthy participants, to inform later work with chronic pain participants. This change has also reduced the introduction substantially as the narrative surrounding the rationale for including a chronic pain population and the underlying theories of analgesia from resizing illusions has been removed.

One reviewer previously suggested that the experiment could be split into two projects. I would consider that option, but at the very least the rationale should be clarified and number of hypotheses could be reduced for clarity. As discussed in previous rounds, many of the hypotheses are actually main effects irrespective of group. In fact, the only group comparison is the final hypothesis and seems to be about the baseline condition - this leaves unclear why having a control group is actually necessary.

In line with the comments from the reviewer, we have decided to split this study into two project, one with healthy participants and another, which will be completed with chronic pain participants. The present study proposes a plan for healthy participants. This has aided in clarifying the rationale for the current study, which is to assess the basis of somatosensory cortex changes during resizing illusions in healthy participants, to use as a basis for later exploration with chronic pain samples. The removal of the chronic pain group has also reduced the number of hypotheses substantially. Please see below for instances of rationale and hypothesis change:

Rationale change:

"Several studies have investigated the analgesic effect of these resizing illusions, as they have been shown to reduce chronic pain in conditions such as osteoarthritis (Preston & Newport, 2011; Preston et al., 2020; Stanton et al., 2018), chronic back pain (Diers et al., 2013), and complex regional pain syndrome (Moseley, Parsons & Spence, 2008). However, the understanding of how these illusions reduce pain is still undetermined. It has been suggested that there are cortical misrepresentations of the size of the affected body part, however, it is unknown if resizing illusions affect this cortical misrepresentation, and if this is therefore what causes the reduction in pain. No study has yet used neuroimaging with a chronic pain population to determine the cortical activity correlated with this illusory analgesia. However, importantly, there has also not been research conducted using SSEPs in healthy participants, to understand what the cortical representations of these resizing illusions are like without the impact of a chronic pain condition. Therefore, the aim of this study is to examine potential changes in the somatosensory cortex during illusory finger resizing in healthy participants,

using vibrotactile SSSEPs, to use as a basis for later investigations in a sample of chronic pain participants."

Hypotheses changes:

"Using different sensory manipulations of finger resizing illusions, in addition to using an electromagnetic solenoid stimulator, this study aims to investigate subjective illusory experience and SSEP responses in healthy participants, to better understand the experience of body ownership illusions from subjective experience and cortical representation perspectives. To test this, different 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. In line with previous findings regarding effective UV conditions (Hansford et al., 2022), subjective questionnaire data will be used to identify individuals who experience an effective UV condition, and these participant's SSEP data will then be analysed. 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 non-illusion 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 (2a) MS visuotactile illusory resizing to NI conditions, and when comparing (2b) effective UV illusory resizing to NI conditions."

Alexandra Mitchell:

It is clear that Hansford and colleagues have put a lot of work in addressing reviewers comments and I am generally content with the manuscript as it stands. I especially appreciate the addition of pilot data. However, there is one prevailing concern and that is the authors have still not really justified why there are so few between group comparisons. Not only that, but they actually expect the same pattern of results in illusory experience and SSSEPs between patients and controls. Therefore, it is not convincing why this approach should be presented in individuals with chronic pain (especially the neuroimaging). Clearer justification for this, and their specific hypotheses, and also why only one group comparison is being made, is needed.

We appreciate your concern regarding the lack of group comparisons and have removed the need for these by changing the proposal to one group of healthy participants as opposed to 2 groups (one of which had chronic pain). We have also updated the hypotheses to reflect this change, removing those which concerned chronic pain samples, as can be seen in the paragraph of text in the section above.

I also have a few minor comments:

1. **The introduction remains too long and should be condensed**

The introduction has been substantially condensed by removing all narrative regarding chronic pain, and the theories underpinning this in relation to resizing illusions, apart from that which is needed for rationale for the current study, as can be seen in the rationale change section in the previous page.

2. **I like the addition of figure 2 but the current version is a little weird and needs a lot more to make it understandable. Does each vertical panel represent time? If so, some measure of time (and time passed) is needed. Direction of stretching should be included in all conditions that include a virtual finger stretch. What happens during the habituation stage**

(i.e. why are there two panels here when nothing changes from panel 3-4?). It should be made clear that these images represent the virtual feedback.

Updates to Figure 2 have been made in line with suggestions from both reviewer comments, which include the addition of notes showing the manipulations in each condition, addition of the stimulator to the finger in all conditions, removal of the second image in the habituation stage, and updating of the caption. The temporal element of the figure can be seen in the caption, showing 2.4 seconds for the manipulation phase and another 2.4 seconds for the habituation phase. The direction of the stretching included in the MS condition, is actually the direction of the researcher manipulation, as the change in finger length shows the direction of the stretching in all other conditions. This is also now reflected in the caption to aid interpretation:

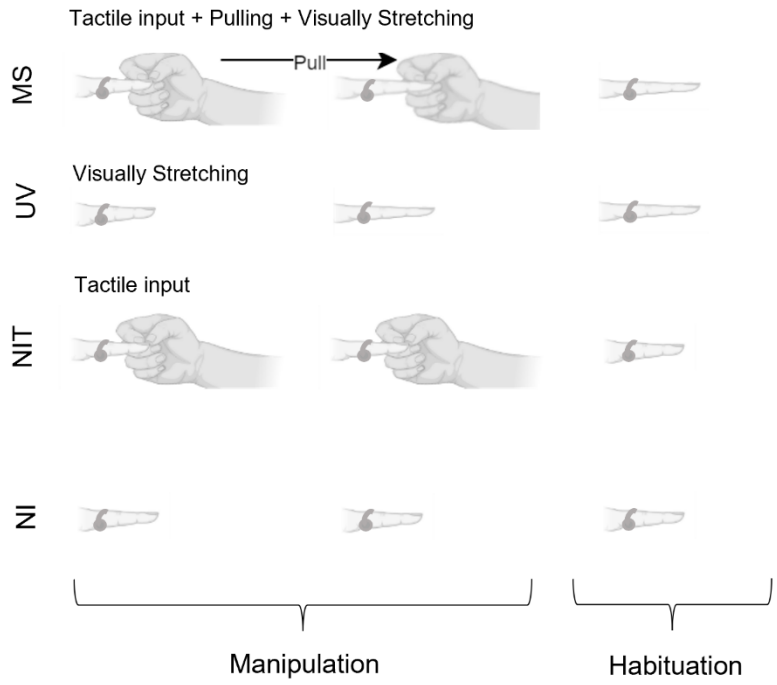

Figure 2. Infographic of Experimental Conditions. MS = Multisensory Stretching, UV = Unimodal Visual Stretching, NIT = Non-Illusion Tactile, NI = Non-Illusion. During the manipulation phase (2.4 seconds) the visual image of the finger is stretched in the MS and UV conditions, and/or the experimenter provides tactile input in the MS (pulling) and NIT (touch) conditions. 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.

3. **Although the authors have clearly justified their use for a paired sample t-test an pre-registered, exploratory ANOVA would add value here as interaction effects might be worth exploring**

Both hypotheses are now using ANOVAs to compare the differences between experimental conditions and (1) subjective illusory experience and (2) SSEP response. This can be seen in section 2.4.2 Planned analysis:

"2.4.2 Planned analyses

2.4.2.1 Hypothesis 1 (Positive Control)

(1 – Positive Control) There will be a greater illusory experience, measured via a subjective illusory experience questionnaire, in the MS condition compared to the NI condition.

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, which we will attempt to replicate. Questionnaire data will be analysed using JASP (JASP Team, 2022). A one-way ANOVA will be run to compare the dependent variable of median illusion score from each independent condition. Given significant findings, post-hoc tests will be run, with Bonferroni correction for 4 comparisons (MS / NI conditions, UV / NI conditions) at an initial alpha of 0.05. Subjective data will also be used to identify participants who effectively experience the unimodal visual condition where participants will be included in further EEG analysis if their median illusion scores on the subjective illusory questionnaire scale for the unimodal-visual condition are greater than 50, in line with previous research using mean subjective embodiment scales (Carey et al., 2019), which will indicate experience of the illusion.

Interpretations for hypothesis 1 can be found in the design table (Appendix A).

2.4.2.2 Hypothesis 2

There will be a significant difference in SSEP response across the electrodes of interest (F1 & FC1) when comparing (2a) multisensory visuotactile illusory resizing to non-illusion, and when comparing (2b) effective unimodal visual illusory resizing to non-illusion conditions.

After pre-processing steps as mentioned in section 2.4.1 are taken, analysis of EEG data will first involve importing the waveforms from MATLAB into R, and then using R to take a Fourier transform for each waveform across all remaining electrodes, to obtain individual results per participant. These will then be averaged across all participants to give overall results, before running a repeated measures within factors 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. No additional filtering or denoising steps will be applied to the EEG data, in line with Figueira et al.'s (2022) report that only a Fourier transform is typically needed for this type of EEG data. Based on the pilot data in Figure 3, we would expect to see activation most pronounced over mid-frontal distributions, covering F1 and FC1 electrodes.

Interpretations for hypothesis 2 can be found in the design table (Appendix A)."

4. **The justification for the minimum effect size of interest for hypothesis 2 (d = .50) outlined in the reviewer response is odd. Why is this the minimum effect size of interest for patient studies over and above what is reported in Lakens? I will not press the point, as I think the study will have enough participants to at least develop a clear idea of any differences as a result of the illusion, but the justification for this chosen effect size should be clear in the manuscript (and more logically explained than it was to me in the previous round of reviews).**

Lakens does not specify a particular effect size to use, but rather advocates for researchers choosing the smallest effect size that they would be interested in detecting. For the purposes of the current

study, this effect size is a Cohen's d of 0.5, since this is a medium effect size of interest. We have updated the power analysis section for Hypothesis 2 to reflect this choice:

"This is the first study to investigate illusory finger stretching using SSEPs, so appropriate effect size estimates are not available. We therefore conducted power calculations based on a smallest effect size of interest, in line with the recommendation of Lakens (2014). Here, we have chosen an effect size of d = 0.5 (a medium effect, see Cohen, 1988), since this is the smallest effect size we are interested in detecting, which we have converted to a Cohen's f of 0.25 for power analyses."

Susanne Stoll:

I would like to thank the authors for their revision/hard work. I went through the response letter and revised clean manuscript in detail. I did not check the inserted text snippets in the response letter due to inconsistencies with the revised clean manuscript. I have several major concerns and a plethora of minor points to report, amounting to another major point.

MAJOR

1. Complexity

• The study involves 4 conditions (MS, UV, NI, NIT), 3 major dependent variables (illusory experience, SSSEPs, pain), 2 groups (healthy, chronic pain), as well as pre and post assessments of pain, rendering the study complex/the manuscript hard to follow. This complexity results in:

- **too many hypotheses (~11) and significance tests (~14 + planned exploratory analyses),**
- **a lengthy intro containing too much info and lacking a clear rationale**
- **somewhat unclear writing (see also minor points)**
- **inconsistencies btw. 2. Methods and Table B1 and within 2. Methods (see minor points)**
- **and a (seemingly) large number of mistakes (see minor points).**

As such, I think the study needs to be simplified considerably.

Suggestion 1: Reduce the number of experimental conditions

• The study aims to reveal potential changes in SSSEPs during illusory finger stretching in healthy and chronic pain individuals. Given that this is the first study of its kind, it seems sensible to focus on the strongest possible contrast btw. conditions (MS vs NI) and drop the remaining conditions (UV, NIT) entirely. This has several advantages:

- **Hypothesis 1a/b: The complex ANOVA + many follow-up tests can be reduced to 2 tests, one testing for MS vs NI in the healthy group and one for MS vs NI in the patient group.**
- **Hypothesis 2b/2d/3b/3d plus significance tests can be dropped entirely (which might also solve issues related to multiple comparisons that remain currently unaccounted for).**

Suggestion 2: Drop secondary hypotheses/tests

• Hypothesis 2e (differences in SSSEPs for healthy vs chronic pain individuals in condition NI)?

To reduce complexity in the study, we have removed the chronic pain group from the proposed plan. This in turn reduces our number of hypotheses considerably and reduces the number of significance tests to be run. With removing the chronic pain group, we also remove a large section of the introduction and have re-established the rationale for the current study, being that we are assessing SSEP changes in a healthy sample to use as a basis for future investigations in a chronic pain sample. Whilst we thank you for the suggestion of removing some experimental conditions, we think that

including the UV condition is very important, as this will allow us to establish if different presentations of the illusion results in different cortical representations. This point is important since the use of the UV condition is the most assessible version of the illusion, as it does not require a large, augmented reality system or the presence of a researcher to deliver the illusion. We have incorporated your suggestion of dropping hypothesis 2e regarding the difference between healthy participants and chronic pain participants in the NI condition, through removing the chronic pain sample.

Suggestion 3: Rewrite and cut the intro to establish a clear rationale

• Establish a clear "red thread". For instance, after para. 1 on the need to find alternative treatment options (l.39-48), one expects that para. 2 outlines directly that resizing illusions/illusory finger stretching are one such alternative.

The introduction has been re-written to remove the narrative regarding chronic pain and to clarify the rationale for the present study in healthy participants. The "red thread" is no longer needed between paragraph 1 and paragraph 2 as both have been removed.

• Drop secondary aspects entirely. Currently, the following represents a distraction to me: predictive coding, central sensitization, rubber hand illusion, the blurring and magnifying hypothesis, and your work involving comparisons btw. asynchronous and synchronous conditions including Appendix C (as not directly relevant).

Regarding the confusion surrounding the predictive coding, central sensitisation, blurring and magnifying theories, these have also all been removed. The rubber hand illusion is still included as this is the basis for the resizing illusions and therefore needs narrative to enhance reader comprehension of the illusions. The section referring to the pilot data in Appendix B (previously C) has been reworded to show relevance as can be seen below:

"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)."

• Increase coherency within a given paragraph. One example is para. 3 (l.69-82). It starts off with illusory finger stretching, then talks about the rubber hand illusion, then about resizing illusions more generally, and then about multisensory resizing, leading to a lack of coherency.

This paragraph has been reformatted to address coherency issues as can be seen below:

"Illusory finger stretching is a form of multisensory illusion, specifically a resizing illusion, which alters the subjective perceptual experience of the size of one's finger. Resizing illusions, through changing the way in which a body part is perceived, exploit principles of multisensory integration to elicit modulations in the perceived size and shape of the body (Preston & Newport 2011; Preston et al., 2020; Stanton et al. 2018). Resizing illusions are based on the rubber hand illusion, in which touch is delivered to a visible fake hand at the same time and in the same place that touch is delivered to the hidden real hand. This manipulation elicits feelings of ownership over the fake hand through the integration of multisensory (tactile and visual) inputs highlighting the apparent malleability of bodily self (Botvinick & Cohen, 1998). Multisensory resizing illusions typically involve both tactile and visual inputs to the participant and can be delivered via an augmented reality system or through magnifying optics. Recent studies have also shown resizing illusions to be effectively administered through visual only, and visuo-auditory manipulations. However, multisensory visuotactile

manipulations are reported as the most effective at inducing a strong experience of the illusion within an augmented reality system (Hansford et al., 2023)."

• Communicate scientific findings more clearly. One example is para. 6 (l.133-135) talking about a direct impact of illusions on the neural representation of the body without letting the reader know what this direct impact consists of. Another example is para. 7 (l.152-153) talking about enhanced responses for "within-modality stimulation" and that this is "in contrast to previous findings" without clarifying what this within-modality stimulation consists of, rendering it unclear why it is different from previous findings. Yet another example is para. 4 (l.88-91) talking about pilot data showing a trend towards greater illusory experience in a synchronous vs asynchronous condition. The pilot data in Appendix C do not suggest such a trend; rather there was no difference.

The section regarding direct impact on neural representations from illusions has been changed to include information about what was used as an index of direct impact as can be seen below:

"Specifically looking at stretching multisensory visuotactile illusions, which as mentioned are those that elicit the greatest illusion strength in a majority of participants, recent research suggests that these illusions directly impact the neural representations of the body and reflect early-stage multimodal stimulus integration through modulation of gamma band activity (Kanayama et al., 2021)."

Regarding the section about enhanced responses for within modality stimulation compared to previous findings, the text has been changed to reflect the difference between these findings as can be see below:

*"*This paradigm has been used with other sensory modalities to better understand the neural mechanisms underlying multisensory integration, with findings showing that presentation of temporally congruent auditory and visual stimuli significantly enhances the magnitude and inter-trial phase coherence of auditory and visual steady-state responses (Nozaradan et al., 2012). However, 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), in contrast to Nozaradan et al.'s findings regarding visuo-auditory combination."

The wording for the text about the pilot data has been changed to remove the comment about the trend towards significance as can be seen below:

"Newport, Pearce and Preston (2010) found strong embodiment using this 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)."

2. Recurring elements of circularity/"double-dipping"

• l.91-96 (Hansford et al., prior work): A subset of participants has been selected based on their illusory experience in a unimodal visual condition, and this subset has then been used for further analysis, rendering this analysis circular.

The work referred to has now been published so the reference has been updated, and the text has been changed to reflect the exploratory nature of the subset analyses and need for replication:

"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 unimodal visual illusions in healthy participants, whilst also showing

in exploratory analysis that a subset of participants who experienced an illusion in the unimodal visual condition reported a stronger illusory experience in this condition than in an incongruent control condition. This subset analysis, however, was of a small sample size, and was selected based on one of the measures analysed thus should be taken with caution, meaning further replication of the findings are needed."

Additionally, the issue of double-dipping has been removed from the plan for the current study by using the whole sample for EEG analysis in hypotheses 2a and 2b, not using the illusory experience data to determine a subsample as mentioned in previous versions. If we find support for both hypothesis 2a and 2b, then the analysis will end here. However, if we do not find support for one or both hypotheses, we will run exploratory correlation analyses between EEG effect and Illusion effect for whichever condition's hypothesis is not supported. For example, if we do not find support for a difference between UV and NI conditions, as per hypothesis 2b, then we will run an exploratory correlation between EEG effect and illusory effect in the UV condition to assess if the greater the illusion strength, the greater the SSEP response. This will address the issue of some participants not experiencing the illusion in the UV condition, as they will simply fall at the lower end of this correlation (if we find a correlation), with small illusion strength and likely small SSEP response. This exploratory analysis has not been mentioned in the manuscript as we will not know if we should run it or not until we have collected our data. If we do run this, it will be clearly stated as exploratory, and it will be highlighted that replications of this are needed in subsequent studies.

• l.440-443 (Hansford et al., prior work): An effect size for participants with an "effective unimodal visual illusion" has been reported, again suggesting that the same data has been used for selection and selective analysis, rendering this analysis circular.

The effect size used for the proposed study comes from the lower end of the effect sizes mentioned, which is the effect size from the full sample of analysis from Hansford et al., (2023). The additional mention of the effect size from the MS and UV comparison in the subsample has now been removed to avoid confusion, as can be seen in the text below:

"Effect sizes are determined by research from Hansford et al (2023a) using the subjective illusory experience questionnaire and comparing MS and UV finger-based resizing illusions using the same finger stretching illusions and the same equipment (n = 48), which show an effect size of $n^2 = .33$ (converted to a Cohen's f = .73). Additional effect size information comes from a visual capture study (n = 80) using a subjective embodiment questionnaire and visual and tactile manipulations to a mannequin body (Carey et al., 2019), showing an effect size of r = .64 (converted to a Cohen's f = .83). An effect size of f = .73 was used for hypothesis 1 to adhere to the lower end of previous effect sizes."

3. Assessment of subjective illusory experience without continuous tactile stimulation

• As the response letter says, no continuous tactile stimulation will be applied when assessing illusory experience. This, however, seems necessary to recreate the experimental scenario. Assessing illusory experience throughout the experiment based on a single trial after the experiment already comes with quite a few assumptions, namely that participants' perception is somewhat stable and/or that they can generate an "internal average".

The wording of the section has been changed to show that the subjective illusory experience questionnaire will be delivered at the end of each block to assess subjective illusory experience, rather than at the end of all trials for the experiment, which means that there will no longer be a repeat of the illusory conditions, participants will instead be asked to recall their experiences from the blaock of trials they have just experienced.

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

4. Redundancies in illusory experience questionnaire

• The questionnaire consists of 6 questions (2 for illusory experience, 2 for disownership, 2 for compliance). Three questions seem to assess the same aspect: "I felt like the finger I saw was part of my body" (illusory experience), "I felt like the finger I saw no longer belonged to me" (disownership); "I felt like the finger I saw was no longer part of my body" (disownership). Why?

The rationale for using the questionnaire with the current 6 items is because this is a questionnaire that has been used several times in research previously and therefore allows for comparison between experiments. The specific questions noted in the above comment might seem to assesses the same concept, but the first regarding illusory experience assesses embodiment of the newly sized finger, whilst the second assess disownership of the newly sized finger, assessing if the change in size no longer makes it feel their own, whilst the third question assesses bodily ownership of resized limbs, as the resized limb could feel like it still belonged to the individual, but not a part of their body, for example if they have embodiment over the limb, but do not see it as part of their concept of their body any longer.

5. Uncertainty in sample size

• To ensure the desired sample size, it needs to be clarified that a data set will be replaced if more than 50% of electrodes need removal (l.369-370) and if electrode F1 and FC1 need to be removed, as these are the electrodes of interest (l.421-422). Similarly, the selection of participants experiencing an "effective unimodal visual illusion" leads to a reduction in sample size that remains currently unaccounted for (l.392-396).

It has been clarified as can be seen below that a dataset will be replaced if more than 50% of electrodes need removal, or if either electrode F1 or FC1 needs removal:

"Data collection will be terminated when the full sample of participants have been tested. If a participant completes <100% of the experiment or if over 50% of electrodes need removal, or if either electrode F1 or FC1 needs removal, then their data will not be included, and additional participants will be recruited to replace any lost data."

If either electrode F1 or FC1 are available then we will keep the dataset.

The reduction in sample size for the subsample experiencing an effective unimodal visual condition is no longer mentioned as we are not using this subset for analysis and are instead using the whole sample of participants for EEG analysis.

"Overall, based on the power analyses in section 2.5, a total sample size of 30 participants will be recruited. This sample size adheres to the higher end of sample size estimates (Hypothesis 2 (2.5.2) showing 30 participants needed)."

6. Minor points

As mentioned, I think the study needs to be simplified. Nonetheless, I will outline some minor points more generally.

Referencing of prior work

• „Hansford et al. (2022)“ → “Hansford et al. (2023)”?

Corrected.

Inclusion/exclusion criteria

• l.229/237: Why are age criteria mentioned as part of both inclusion and expulsion criteria? What about people who are exactly 18 or 75 years old?

“Inclusion Criteria: Right-handed, 18 years of age or over, no older than 75 years of age (include those aged 75 years).” – Corrected.

• l.231: “hand-based” → “finger-based”?

Removed.

• l.249-250: It needs to be added that “50% of electrodes needing removal” applies to the data set of a single participant. It would be also good to point out here in which section more info about data removal can be found.

“Less than 100% of the experiment completed by a participant, more than 50% of electrodes for a single participant needing removal from EEG data, or if either electrode F1 or FC1 (electrodes of interest) need removal. More information about data removal can be found in section 2.4.1 Preprocessing Steps.” – Corrected.

• l.359-361: “recruited” → “tested”?

Corrected.

Experimental procedure

Questionnaire – handedness and pain

• l.337-339: It needs to be stated that pain levels will be assessed before and after each experimental condition.

No longer needed as chronic pain sample removed.

• l.262-265/284-285: It is not clear why participants have to rate the level of pain for the most painful finger and then again the level of pain for their hand on the same day, and will then be asked again what their most painful finger is. Some of this seems to refer to “recruitment” and “testing”. This needs to be signposted better.

No longer needed as chronic pain sample removed.

• l.285-286: Why not select a finger randomly if multiple fingers are equally painful to omit participant bias?

No longer needed as chronic pain sample removed.

Digit manipulation

• l.298: “hand” → “finger”?

This is correct as their hand will be placed into the system, with their finger outstretched, not just their finger placed into the system.

• l.301: “augmented” → “(augmented)”? There is only augmentation in the illusory conditions.

Corrected.

• There are issues with pluralisation. For instance (l.302/304): "conditions" → "condition"

Corrected.

• l.306/310: "second NI control condition"/"the second control condition" → "the NIT condition"

Corrected.

• l.328-330: Why not repeat an erroneous trial instead of removing it (given that mistakes by the experimenter are rare, as indicated in the response letter)? This would guarantee the desired number of trials per condition block.

The trials are randomly generated by a MATLAB script which will keep running until all trials are completed and then the script will run the final trials for the subjective illusory experience questionnaire. Therefore, it is not possible to get up in the middle of a session and stop the script to run an additional trial, and then start the script again, as the randomisation will be incorrect. We are running a lot of trials per block, therefore in the rare instance that one is incorrectly delivered, removing one trial from the dataset will not cause issue.

• l.323-325: I think it should read "whether to pull or touch the finger or […]". Moreover, what color does the box have if there is no manipulation?

This has been corrected in text. The box does not need a colour if there is no manipulation, as there will still be tactile input, so the box will just be white. This can be seen in the text below:

"If the box is blue, this will indicate a need to pull the finger, if it is white it will indicate a need to touch the finger, if there is no box displayed then this indicates no tactile manipulation from the experimenter."

Augmented reality system

• l.279-284: For resolutions, add "pixel". Is it correct that the screen is 56 cm above the felt base? The mirror is 26 cm from the felt base and the screen 26 cm from the mirror, resulting in 54 cm. Also, what's the height of the screen?

"pixel" has been added to the text regarding the resolution and the screen is 54cms from the felt base as mentioned in the text below, and the height of the screen has been added:

"26cms above the felt base of this central area, there is a mirror, which is placed 26cms below a 1920 x 1200 resolution screen, with a width of 52cms and a height of 32cms. The distance between the central area with the felt base and the area from the mirror to the screen is 2cms. This screen is 54cms from the base of the system, and the base of the system is 82cms from the ground."

Experimental conditions

• I think Figure 2 needs to be clearer:

- **Why are there 2 images for each condition in the habituation phase?**

  The second image was to show the duration of the habituation phase but has been removed to improve understanding.

- **MS/UV: In the habituation phase, the visually stretched fingers seem longer than the visually stretched finger in the manipulation phase. Why?**

This has been changed so that the fingers are now the correct length.

- **NIT: In the manipulation phase, the fingers appear to be visually stretched, although I think they should be as short as the fingers in the habituation phase.**

This has been changed so that the fingers are now the correct length.

- **To increase clarity, above the right image for each condition in the manipulation phase, it might be worth adding some text [e.g., MS: Touched + Pulled + visually stretched; UV: Visually stretched; NIT: Touched]. Similarly, it would be good to add the stimulator to the finger as this is an integral part of the experimental setup.**

Text has been added to demonstrate the manipulation for each condition and a stimulator has been added to the fingers in each condition.

- **Caption: The statement about manipulation phase seems incorrect, as there is no illusion in condition NIT or NI.**

Caption has been updated to reflect the addition of the stimulator and to correct the statement about the manipulations as can be seen below:

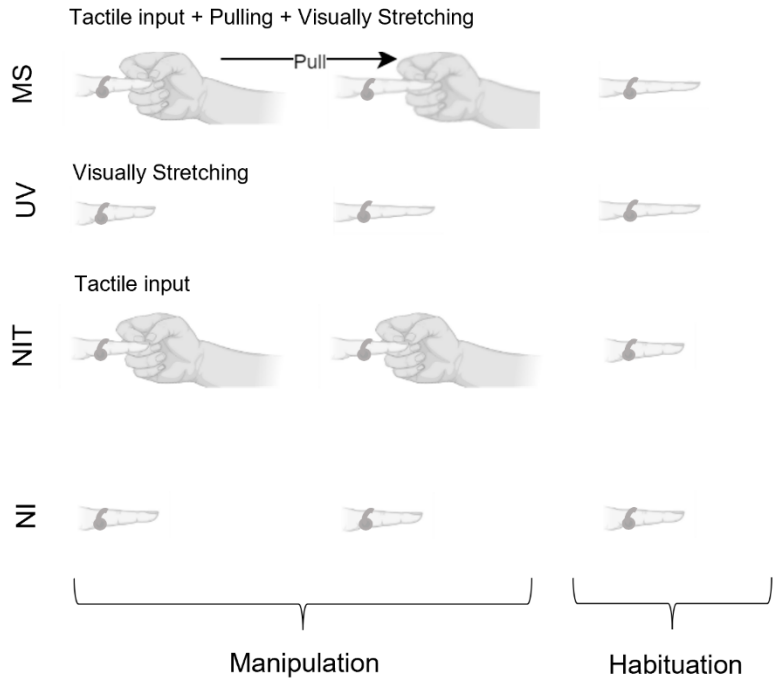

Figure 2. Infographic of Experimental Conditions. MS = Multisensory Stretching, UV = Unimodal Visual Stretching, NIT = Non-Illusion Tactile, NI = Non-Illusion. During the manipulation phase (2.4 seconds) the visual image of the finger is stretched in the MS and UV conditions, and/or the experimenter provides tactile input in the MS (pulling) and NIT (touch) conditions. 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.

Preprocessing steps

• l.365-368: This could be clearer; maybe: "the 5% of electrodes showing the largest standard errors will be removed"?

The text has been clarified as can be seen below:

"Across all the standard errors, the 5% of electrodes showing the largest standard errors will be used to create a standard error threshold. Any electrode with a standard error above this threshold, or with a value of 0, will be removed from analysis."

• The preprocessing of the EEG data needs to be outlined more clearly. It it needs to be clearer how one gets from a continuously recorded EEG signal per electrode and participant to an SSSEP amplitude per electrode, condition, and participant that will then be used for the t-tests. For that, the continuously recorded EEG signal needs to be segmented, the data need to be averaged across the 24 trials of a condition etc. Some of this info is mentioned in l.408-411/417- 419/483-487, which should be moved to 2.4.1 Preprocessing steps. If data will be collapsed across the electrodes of interest (F1, FC1), this should be indicated too.

The preprocessing section relating to the EEG data has been updated to reflect the missing components as can be seen below:

"Data will first be converted using MATLAB and EEGlab from the ANT EEprobe .cnt format to EEGlab .set format. All subsequent analysis will then be conducted using the MNE-Python toolbox (Gramfort et al., 2013). A 50Hz notch filter will first be applied to the raw EEG data for all electrodes, followed by calculation of the standard error across time for each electrode for each participant (Luck et al., 2021). Across all the standard errors, the 5% of electrodes showing the largest standard errors will be used to create a standard error threshold. Any electrode with a standard error above this threshold, or with a value of 0, will be removed from analysis. Where a participant has over 50% of their electrodes over the standard error threshold or with a value of 0, or if the electrodes requiring removal contain both electrodes F1 and FC1 (electrodes of interest), then their data will be removed. 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. Statistical comparisons will then be performed on the Fourier amplitudes at the stimulation frequency (26Hz), across conditions and participants. No additional filtering or denoising steps will be applied to the EEG data, in line with Figueira et al.'s (2022) report that only a Fourier transform is typically needed for this type of EEG data."

• l.371-376: If I am not mistaken the illusion/disownership indices will have values ranging from -100 to 100. This needs to be described in 2.4.1 Preprocessing steps along with an interpretation of what these indices (-100, -50, 0, 50, 100) mean exactly. Some of this is mentioned in l.504-506, which should be moved to 2.4.1 Preprocessing steps.

The range will be from 0 – 100, which has been clarified in the preprocessing section as can be seen below:

"Regarding questionnaire data, all data will be collected from a range of 0 – 100, with scores below 50 being indicative of disagreement to the statement, whilst a score of 50 is a neutral option

regarding the statement, and scores above 50 are indicative of agreement with the statement. 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)."

- **2.4.2 Planned analyses (l.390/394) refers to a median illusion score and Table B1 to mean illusion scores being used as dependent variable. Similarly, 3. Pilot Data mentions an average illusion score (l.500). This is somewhat inconsistent and incorrect as 2.4.1. Preprocessing steps states that an index will be calculated (l.373).**

Table B1 has been updated to show median scores being used. The text on the pilot data has also been updated to reflect the control index use as can be seen below. The data do not change for the figure as in this dataset all participants reported scores of 0 for the control questions.

"Pilot data was also collected using the vibrotactile stimulator at 26Hz to make sure that the illusory experience is not removed due to the addition of this vibrotactile input. Pilot data was collected from 4 additional healthy participants, who underwent the same experimental protocol as mentioned in the "Experimental Procedure" section, simply without EEG caps fitted. Illusory experience was calculated using the median of both illusion scores for each participant minus their median control scores, as per the preprocessing steps regarding the control index, and then the data were averaged over participants to give the results seen in Figure 4."

- **Whereas 2.4.1 Preprocessing steps states that mean pain scores will be calculated (l.376-377), 2.4.2 Planned analyses talks about median pain scores (l.432) and Table B1 about mean pain scores. It should be always "median".**

No longer needed as chronic pain sample removed.

Stated hypotheses

- **Hypothesis 1a/1b: Given that it says "non-illusion condition" (singular) for hypothesis 1a, but "non-illusion conditions" (plural) for hypothesis 1b, it is not clear what these hypotheses are about (l.381-387). I would assume this is a mistake and hypothesis 1a is MS > NI in the healthy group and hypothesis 1b is MS > NI in the patient group.**

This is updated to "conditions" throughout.

- **When stating the hypotheses in the text and Table B1, it might be better/clearer to always use the condition acronyms (e.g. MS condition or NI condition).**

Acronyms are now used throughout.

Planned analyses

- **Hypothesis 1a (MS > NI in the healthy group)/1b (MS > NI in the patient group): It is not clear why an ANOVA should be performed instead of 2 direct tests. Moreover, in both 2.4.2 Planned Analyses (l.390) and Table B1, it is unclear what type of post-hoc test will be conducted. Moreover, whereas 4 comparisons are mentioned in 2.4.2 Planned Analyses (l.391), Table B1 mentions 3, and "3 measurements" are mentioned in 2.5 Power Analysis (l.451) and also Table**

B1. Moreover, given that the suggested ANOVA seems to have the study groups as a factor, it is a "mixed ANOVA" and not a "within ANOVA" (l.450).

An ANOVA will be used to compare between all conditions, and the hypotheses relate to findings of a specific comparison. The type of post hoc test is not mentioned as until the data are collected it will not be clear if they meet the assumptions for an ANOVA or if a non-parametric equivalent will be needed, hence why this has been left unspecified. The 3 comparisons have been corrected to four comparisons and the power analyses redone and table and text updated. Since we have removed the chronic pain group, the ANOVA is correct as a within factors ANOVA.

- **Hypothesis 2: It needs to be stated more clearly that electrode F1 and FC1 are the electrodes of interest (l.421-422). This is because the number of electrodes of interest seems to determine the number of statistical tests. Moreover, it needs to be clarified whether a statistical test will be performed for each of these electrodes. The caption in Figure 3 suggest that data might be collapsed across F1 and FC1.**

Hypothesis 2 has been updated to reflect the electrodes of interest as can be seen below:

"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 (2a) MS visuotactile illusory resizing to NI conditions, and when comparing (2b) effective UV illusory resizing to NI conditions."

However, we plan to average across electrodes F1 and FC1, so there is no need for additional statistical tests to be run.

- **Hypothesis 2e: The text (l.414-415) and Table B1 state that a dependent samples t-test will be run, which is seems incorrect because study groups are being compared.**

This hypothesis has been removed due to the chronic pain group being removed.

- **Hypothesis 3: A two-tailed test is mentioned in 2.5 Power Analysis (l.466-467), a one-tailed test in 2.4.2 Planned analyses (l.431) and also in Table B1. Hypothesis 3a/b suggest a one-tailed test. Moreover, I think hypothesis 3c/d would require an (additional) Bayesian analysis to quantify evidence for the null hypothesis.**

This hypothesis has been removed due to the chronic pain group being removed.

Effect sizes and power analyses

- **Hypothesis 1a/1b: It is unclear why effect sizes for MS vs (effective) UV are being used (l.439-448), as hypothesis 1a seems to involve MS > NI in the healthy group and hypothesis 1b MS > NI in the patient group. It is also unclear what the effects size from Carey et al. refers to (l.445; what has been contrasted here?)**

Effect sizes from previous work are from a Friedman test incorporating all conditions within each sample, therefore these are used for the current study which will also compare all conditions, including an MS and UV condition as mentioned in the manuscript. The comparison made in the Carey et al paper has also now been included as can be seen below:

"Additional effect size information comes from a visual capture study using a subjective embodiment questionnaire and visual and tactile manipulations to a mannequin body (Carey et al., 2019), showing an effect size of r = .64 (converted to a Cohen's f = .83) when comparing embodiment scores from the questionnaire against control scores."

• Hypothesis 2e: A comparison btw. healthy controls and patients requires to determine a sample size for an unpaired (and not a paired) test (Table B1 and l.457).

This hypothesis has been removed due to the chronic pain group being removed.

• Hypothesis 3: Two effect sizes are mentioned, although only one is reported. Pilot data are mentioned, which do not seem to be shown in the manuscript (l. 462-465).

This hypothesis has been removed due to the chronic pain group being removed.

Pilot data

• Ideally, the pilot data in Figure 3 should show that this works for each condition (and not just across all conditions as collapsing across conditions should be more powerful).

The purpose of the pilot was to assess if we could obtain a reliable steady state signal across the conditions, which we show in Figure 3. We appreciate the comment about showing the data for each condition, however, since we only have data from 3 participants, we do not have enough power to look at condition specific effects within this pilot sample. In the proposed study this is exactly what we will be investigating, so the question of if there are different presentations of SSEP response within each condition, is what the study is seeking to find.

Abbreviations

• After/before introducing abbreviations, they should be always used/spelled out (e.g., CRPS, SSSEP)

Corrected.

• It should not be necessary to introduce the acronyms for the experimental conditions twice

Currently the acronyms are introduced once in the introduction and then used as a reference in the experimental procedure to aid readers in using the acronyms, which are then not described again.

• Figure 4-caption: It is helpful to spell out the acronyms used in the figure

Corrected.

Appendices

• Appendix A (Timeline): Given that Appendix A is not mentioned in the text and not necessary, it can be removed to keep the manuscript clean.

Removed.

• Appendix B (Table B1):

- **For reasons of clarity, I think it would be good to list each subhypothesis separately, which is currently not the case for hypothesis 1 and 3.**

The subhypotheses have been removed through removing the chronic pain sample.

- **Column "Analysis plan"/"Hypothesis": It would be good to always use the condition acronyms when specifying the contrast of interest.**

Corrected.

- **Column "Question": The question for hypothesis 3c/d is incorrect.**

Removed due to chronic pain group being removed.

- **Column “Sampling plan”: Statements about achieved power can be removed.**

Removed.