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)?

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.
- 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).
- 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.
- 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.

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.
- 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.

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".

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?

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).

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)"?

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?
- l.231: "hand-based" → "finger-based"?
- 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.
- l.359-361: "recruited" → "tested"?

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

Digit manipulation

- l.298: "hand" → "finger"?
- l.301: "augmented" → "(augmented)"? There is only augmentation in the illusory conditions.
- There are issues with pluralisation. For instance (l.302/304): "conditions" → "condition"
- l.306/310: "second NI control condition"/"the second control condition" → "the NIT condition"
- 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.
- 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?

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?

Experimental conditions

- I think Figure 2 needs to be clearer:
  - Wy are there 2 images for each condition in the habituation phase?
  - MS/UV: In the habituation phase, the visually stretched fingers seem longer than the visually stretched finger in the manipulation phase. Why?
  - 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.
  - 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.
  - Caption: The statement about manipulation phase seems incorrect, as there is no illusion in condition NIT or NI.

Preprocessing steps

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

- 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).
- 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".

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.
- 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).

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).
- 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 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.
- 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.

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?)
- 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).
- 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).

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).

Abbreviations

- After/before introducing abbreviations, they should be always used/spelled out (e.g., CRPS, SSSEP)
- It should not be necessary to introduce the acronyms for the experimental conditions twice
- Figure 4-caption: It is helpful to spell out the acronyms used in the figure

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.
- 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.
  - Column "Analysis plan"/"Hypothesis": It would be good to always use the condition acronyms when specifying the contrast of interest.
  - Column "Question": The question for hypothesis 3c/d is incorrect.
  - Column "Sampling plan": Statements about achieved power can be removed.