I would like to thank the authors for their extensive revision. The manuscript has become much more digestible. I am sure this distilling process required a great deal of effort.

I went through the response letter and clean manuscript. I did not check the text snippets in the response letter because of inconsistencies with the clean manuscript. I have several major points to report as well as a series of minor points. A subset of these points I have raised before. Many of the points cover inconsistencies and mistakes – a recurring problem of the manuscript.

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.

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).

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.

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.

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).

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.

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).

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%.

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.
- Text mentions a power of 90% (l. 361) and Table A1 80%.
- 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 mentions 30 participants (l. 143, 145, 362) and Table A1 24.
- Just like the text (l. 327), Table A1 should refer to electrodes of interest for hypothesis 2a.

Analyses

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

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

- 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.

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 $f = \text{sqrt}(\eta^2/(1- \eta^2))$, I obtain a Cohen's f of 0.70.

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).

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).

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?

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?

MINOR

Intro

- l. 68: What is meant by "incongruent"?
- l. 78-80: The part about the peripheral stimulus raises more questions than it answers. Simply drop?
- 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?
- 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.
- 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"?
- 1. 103-104: "[…] greater increases in steady-state response magnitude when this corresponds with the amplitude modulation rate […]." Unclear; what correspond with what?
- l. 114: Add reference?
- l. 121: Acronym SSSEPs not introduced.

Other

- 2.3 Experimental Procedure: It needs to be explicitly mentioned that right fingers will be used.
- l. 144: "recruited" → "tested"?
- 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".
- Figure 2/l.235: Caption could be clearer as "tactile input" in the figure means "just touching".
- l. 249, 251: "pulling"/"pull"/"pulls"→ "pulling/pull/pulls or just touching/touch/touches"?
- l. 265: "that condition" → "condition presented in a given block"?
- l. 267: "trial" → "trials"?
- l. 289-297: Re preprocessing steps, a few things could probably be clearer:
  - It would be good to add that "across all standard errors" refer to all standard errors across participants.
  - 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.
  - 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.
- l. 304: "per trial" → "per condition"?
- 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).
- l. 319: "MS / NI" etc. → "MS vs NI" etc.?
- 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).
- l. 344: n^2 → η^2 (eta squared)?
- 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?
- Figure 3-Scalp topography: Add saturation bar or explain in caption what saturation levels mean.