<u>Supplementary Material:</u> Robotically-induced auditory-verbal hallucinations: combining self-monitoring and strong perceptual priors

Pavo Orepic[1], Fosco Bernasconi[1], Melissa Faggella[1], Nathan Faivre[2], Olaf Blanke[1,3]

[1] Laboratory of Cognitive Neuroscience, Neuro-X Institute & Brain Mind Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

[2] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

[3] Department of Clinical Neurociences, Faculty of Medicine, University of Geneva, Geneva, Switzerland

Corresponding author

Olaf Blanke

Bertarelli Chair in Cognitive Neuroprosthetics, Neuro-X Institute & Brain Mind Institute, School of Life Sciences, Campus Biotech, École Polytechnique Fédérale de Lausanne (EPFL), 1012 Geneva, Switzerland

E-mail: olaf.blanke@epfl.ch

Tel: +41 (0)21 693 69 21

Fax: +41 (0)21 693 69 22

Questionnaire ratings

Supplementary Table T1 reports means and standard deviations of all questionnaire ratings for both studies and Supplementary Figure S1 illustrates the ratings.

Table T1. Mean and standard deviation (SD) of questionnaire items assessed in sensorimotor blocks in both studies.

	Study 1				*Study 2*			
	Asynchronous		Synchronous		Asynchronous		Synchronous	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Self-touch	2.08	2.22	2.88	2.29	1.21	1.74	2.04	2.1
Somatic passivity	3.42	2.12	2.33	2.09	3.71	2.1	3.13	2.29
Presence Hallucination	1.75	1.92	1.25	1.7	2.54	2	1.88	2.29
Control	0.21	0.59	0.17	0.48	0.17	0.48	0.29	0.81

In Study 1, binomial mixed-effects analysis investigating the effects of questionnaire ratings on false alarm rates, with Response as dependent variable and Stimulation and Somatic Passivity as fixed effects, indicated a significant interaction between Somatic Passivity and Stimulation (estimate=-0.53, Z=-4.3, p<0.001). This effect, however, was not replicated in Study 2 (estimate=-0.04, Z=-0.35, p=0.724). Similarly, in Study 1, there was a tendency for an interaction between Presence Hallucination and Stimulation (estimate=-0.35, Z=-1.78, p=0.076) that was not replicated in Study 2 (estimate=-0.17, Z=-1.47, p=0.143). These interactions in Study 1 indicated

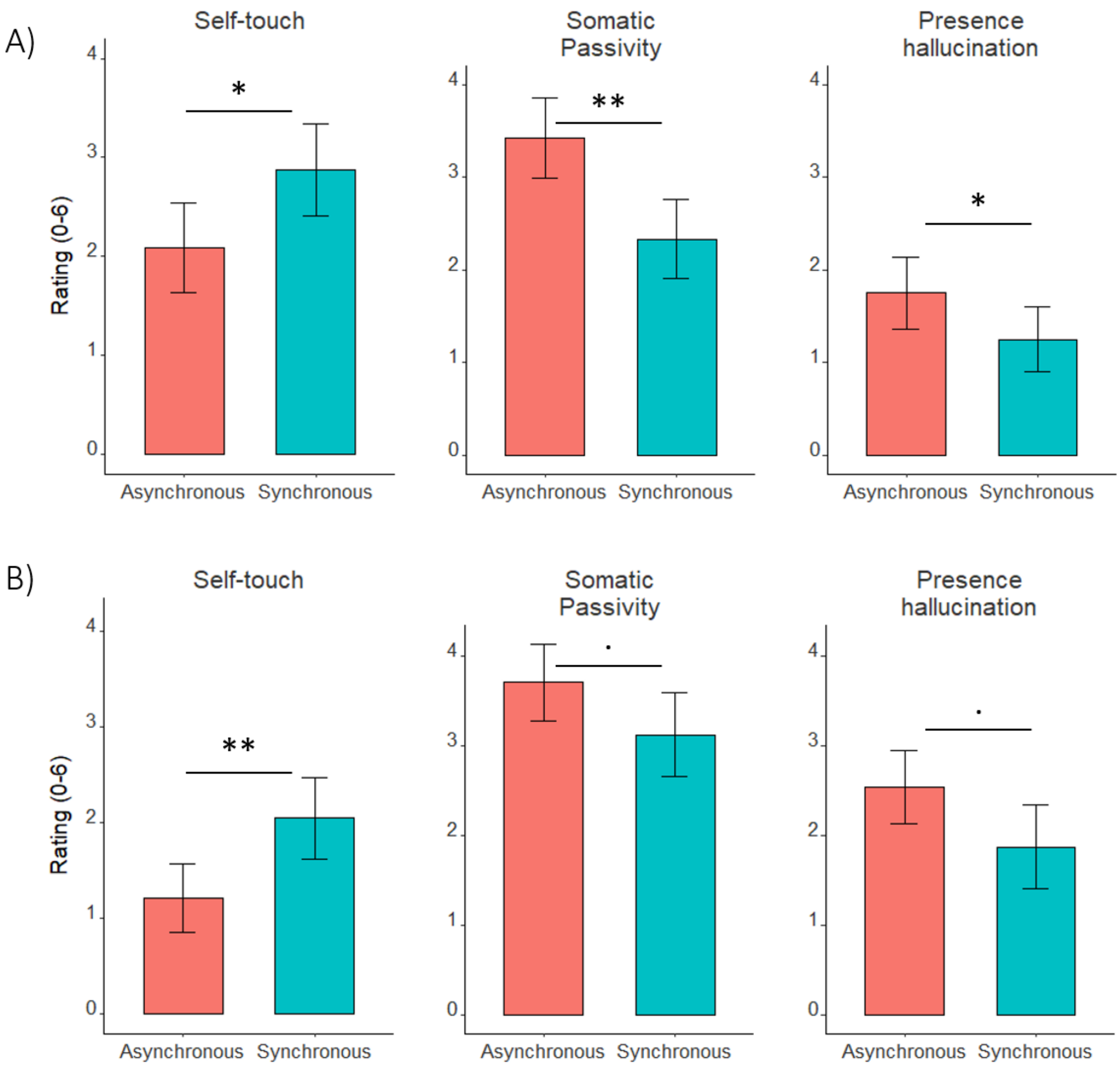

Figure S1. Illusory effects assessed in sensorimotor blocks Study 1 (A) and Study 2 (B). Height of bar plots indicates mean rating, and error bars 95% confidence intervals. In both studies, self-touch was higher during synchronous, whereas somatic passivity and presence hallucination during asynchronous stimulation. **: $p<0.01$, *:$p<0.05$, .:$p<0.1$

a stronger relationship between Somatic Passivity and Presence Hallucination with false alarms in Asynchronous, compared to Synchronous stimulation. Self-touch did not have any effects on false alarms in either study (all $p > 0.05$).

Signal detection theory

In neither study did we observe significant effects on d' (Supplementary Table T2) and criterion (Supplementary Table T3). However, as in both studies there were tendencies for a main effect of Stimulation only in other-voice trials (Study 1 Self: $F(1, 23)=0.17$, $p=0.685$; Study 1 Other: $F(1, 23)=3.6$, $p=0.071$; Study 2 Self: $F(1, 23)=0.73$, $p=0.403$; Study 2 Other: $F(1, 23)=2.95$, $p=0.099$;), indicating a lower criterion for asynchronous compared to synchronous stimulation (Supplementary Figure S2), we additionally ran exploratory analysis of d' and criterion by merging the data from both studies. This analysis revealed a significant interaction between Stimulation and Voice for the criterion value ($F(1, 47)=4.11$, $p=0.048$). Further investigation of this interaction revealed that criterion was reduced in the asynchronous (mean = 0.44, 95% CI = [0.19, 0.69]) compared to synchronous condition (mean = 0.57, 95% CI = [0.35, 0.79]) only for other-voice blocks ($F(1, 47 =6.36$, $p=0.015$), with no differences between the two stimulations for self-voice blocks ($F(1, 47)=0.79$, $p=0.38$; asynchronous: mean = 0.50, 95% CI = [0.25, 0.75]; synchronous: mean = 0.43, 95% CI = [0.16, 0.70]). There were no significant main effects of Stimulation ($F(1, 47)=0.35$, $p=0.558$) and Voice ($F(1, 47)=0.81$, $p=0.372$) on the criterion value with merged studies, as well as no significant effects on d' (Condition: ($F(1, 47)=0.09$, $p=0.761$; Voice: ($F(1, 47)=0.23$, $p=0.632$; Condition * Voice: ($F(1, 47)=1.18$, $p=0.283$).

Lower criterion in asynchronous condition suggests a liberal observer, i.e. proneness to respond 'yes'. Thus, these results suggest that in other-voice blocks, participants were more likely to report hearing a voice during asynchronous, compared to synchronous stimulation Although this analysis is exploratory and indicates weak results, these findings are in accordance with the results reported in the main text.

Table T2. A two-way ANOVA assessing d' in both studies.

	Study 1			*Study 2*		
	df	F	p	df	F	p
Stimulation	1, 23	0.1	0.759	1, 23	0	0.95
Voice	1, 23	1.27	0.272	1, 23	3.97	0.058
*Stimulation * Voice*	1, 23	0.2	0.66	1, 23	1.45	0.241

Table T3. A two-way ANOVA assessing criterion in both studies.

	Study 1			*Study 2*		
	df	F	p	df	F	p
Stimulation	1, 23	0.1	0.759	1, 23	0	0.95
Voice	1, 23	1.27	0.272	1, 23	3.97	0.058
*Stimulation * Voice*	1, 23	0.2	0.66	1, 23	1.45	0.241

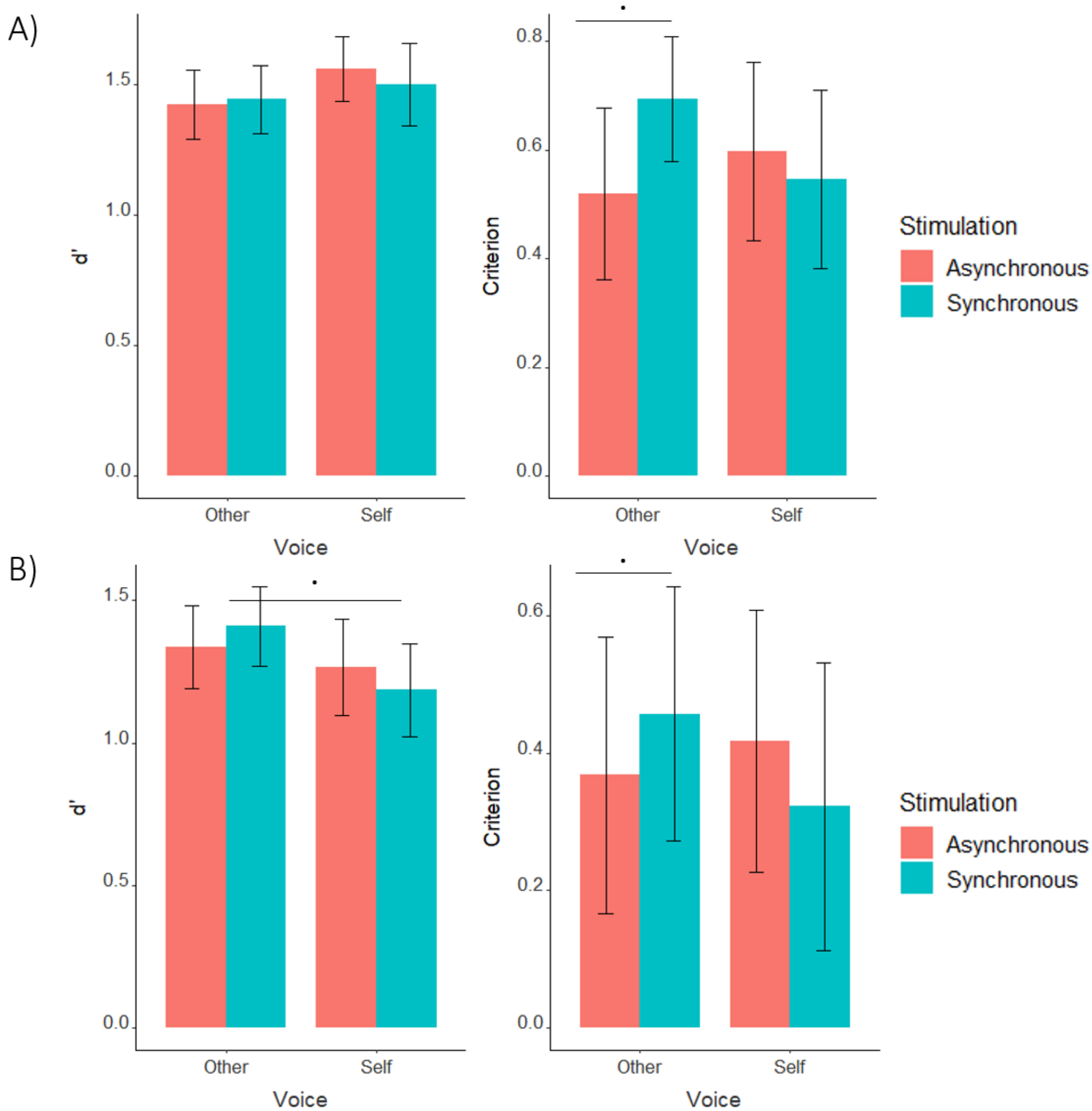

Figure S2. Criterion and d' values in both Study 1 (A) and Study 2 (B). In both studies, there were tendencies for a reduced criterion in Asynchronous compared to Synchronous stimulation, but only in Other-voice blocks. In Study 2, there was a tendency for a reduced d' in Self-voice compared to Other-voice blocks. .:p<0.1

PDI subcategories

Besides the general PDI score, PDI questionnaire contains three subcategories – distress, preoccupation and conviction (Peters *et al.*, 2004). Previous work has indicated effects of beliefs on perception that were associated specifically with the conviction subscore, in addition to the general PDI score (Schmack *et al.*, 2013). Similarly, we found a significant interaction between the fixed effects Conviction and Stimulation – indicating a stronger positive relationship between false alarm rate and conviction score in asynchronous, compared to synchronous stimulation (Supplementary Figure S3, details in the main text). In Study 1, we observed equivalent interactions with the effect of Stimulation for both effects of Distress (estimate=-0.1, Z=-2.77, p=0.006) and Preoccupation (estimate=-0.13, Z=-3.61, p<0.001), again indicating a stronger relationship between PDI sub-scores and false alarms during Asynchronous stimulation. These interactions were, however, not replicated in Study 2 (distress: estimate=-0.03, Z=-0.99, p=0.325; preoccupation: estimate=-0.03, Z=-1.64, p=0.1). Study 1 also indicated tendencies for the main effects of all 3 PDI subcategories (distress: estimate=0.08, Z=1.85, p=0.065; preoccupation: estimate=0.08, Z=1.92, p=0.054; conviction: estimate=0.07, Z=1.65, p=0.099), indicating a general increase of false alarms with the increase of the scores. In Study 2, only the main effect of distress proved to be significant (distress: estimate=0.11, Z=2.35, p=0.019; preoccupation: estimate=0.08, Z=1.52, p=0.129; conviction: estimate=0.04, Z=0.85, p=0.395).

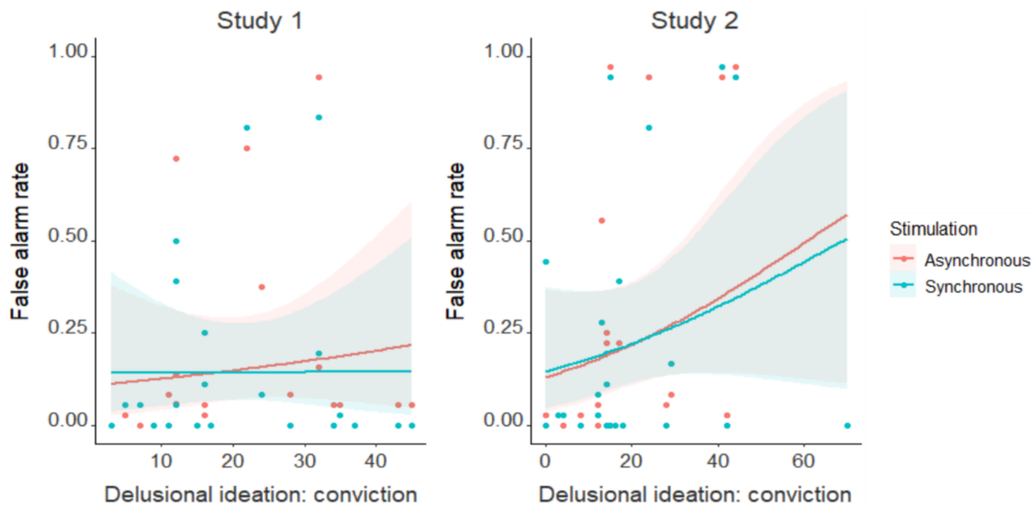

Figure S3. In both studies, increase in PDI conviction score was more strongly related to an increase in vocal false alarms rate during asynchronous, compared to synchronous stimulation. Shaded areas around each curve represent 95% confidence intervals.

Gender effects

Previous work has indicated that females are more likely to experience PH (Alderson-Day *et al.*, 2022) and personified voices (Alderson-Day *et al.*, 2021), compared to males. In Study 1, we found a tendency for the corresponding effect of Gender – false alarms were more often reported by female participants (estimate=-2.05, Z=-1.88, p=0.06). However, this effect was not replicated in Study 2 (estimate=-1, Z=-0.71, p=0.479), and could have arisen from the fact that in Study 1 there were more female participants. There were no significant effects of gender on questionnaire ratings. When pooling the data from both studies together (see below), we observed the same effect of gender as in Study 1.

Merging both studies

Since experimental design was equivalent in both studies, as exploratory analysis, we analyzed both studies together – i.e. we pooled all participants from both studies in the same mixed-effects binomial regression and added an additional factor Study (with levels 1 and 2). The effects reported in the main text were replicated – there were more FAs during asynchronous stimulation and this was more prominent in other-voice blocks. The gender effect was significant – there were more false alarms in female participants – and there was no significant effect of Study. The model is summarized in Supplementary table T4. Running the analysis separately for each level Voice (self, other), we observed a significant effect of Stimulation in other-voice blocks (estimate=-0.6, Z=-3, p=0.003), again indicating more false alarms during asynchronous stimulation. As in both studies separately, in self-voice blocks, we observed the opposite effect – more false alarms during synchronous stimulation (estimate=0.53, Z=2.66, p=0.008).

Table T4. Binomial mixed-effects regression for merged studies 1 and 2.

	estimate	*Z*	*p*
Stimulation	-0.76	-2.10	0.035
Voice	-0.18	-0.89	0.374
*Stimulation * Voice*	1.08	3.79	< 0.001
Gender	-1.83	-2.04	0.041
Study	0.52	0.56	0.576

False alarms rate of 0 and 1

The scenarios with very low and very high false alarm rate are an important contributor to our results, because they significantly differ across conditions. However, our results persist even when we remove participants who exhibited such behavior in our experiments.

21/48 participants had no false alarms in synchronous, and 16/48 in asynchronous condition. Considering that false alarm rate of 0 is very common in signal detection studies, it is important to indicate that it occurred significantly less during asynchronous stimulation. No participants had false alarm rate of 1, but 3/48 participants had false alarm rate higher than 0.9 in synchronous and 5/21 in asynchronous condition. Similarly, false alarm rate higher than 0.9 is very rare in signal detection studies, and yet in our studies it was higher during asynchronous condition.

Removing all participants that fit into one of these 4 categories (with false alarm of 0 or higher than 0.9 in either condition) leaves us with 18/48 participants (10 in Study 1 and 8 in Study 2). Our main effects persist even without those participants. In both studies, there was a significant interaction between Stimulation and Voice (Study 1: estimate = 1.67, $Z = 0.45$, $p < 0.001$; Study 2: estimate = 1.09, $Z = 2.41$, $p = 0.16$, Supplementary figure S4).

The relationship between false alarm rate and PDI score was not significant when analyzing studies separately (Study 1: estimate = -0.164, $Z = -1.85$, $p = 0.064$; Study 2: estimate = -0.08, $Z = -1.24$, $p = 0.214$; Supplementary figure S5). This could be simply due to a small sample size ($N1=10$ and $N2 = 8$, compared to previous $N1=N2=24$). However, it was significant when pooling these 18 participants together – the interaction between PDI and Stimulation was significant (estimate = -0.1, $Z = -2.06$, $p = 0.039$) with the relationship between false alarm rate and PDI scores being stronger during asynchronous stimulation (Supplementary Figure S6 B).

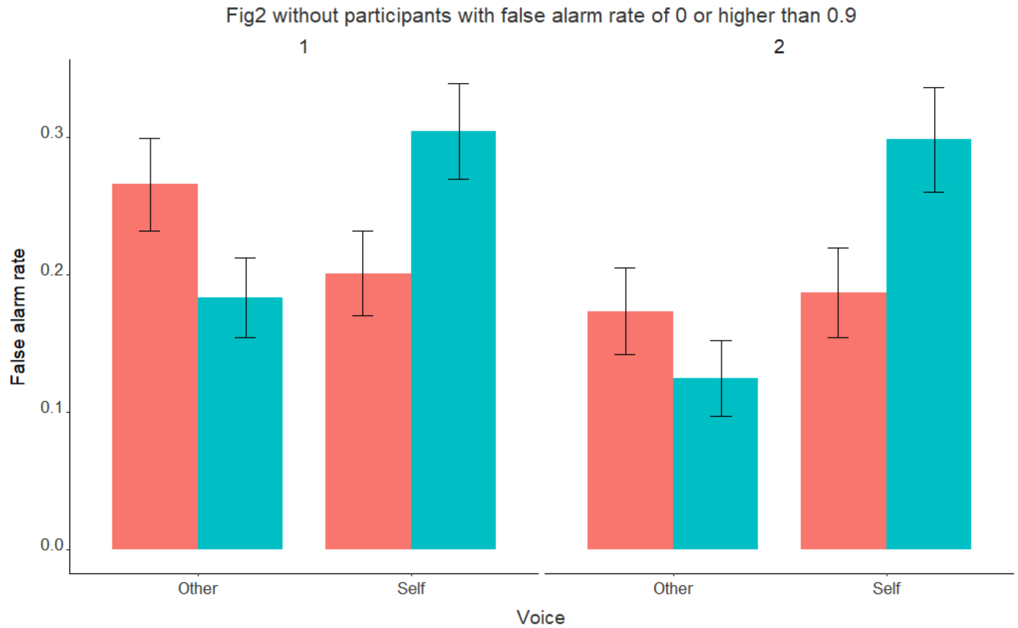

Figure S4. Figure 2 from the main text after removing participants with false alarm rate of 0 or higher than 0.9 in either condition (synchronous, asynchronous).

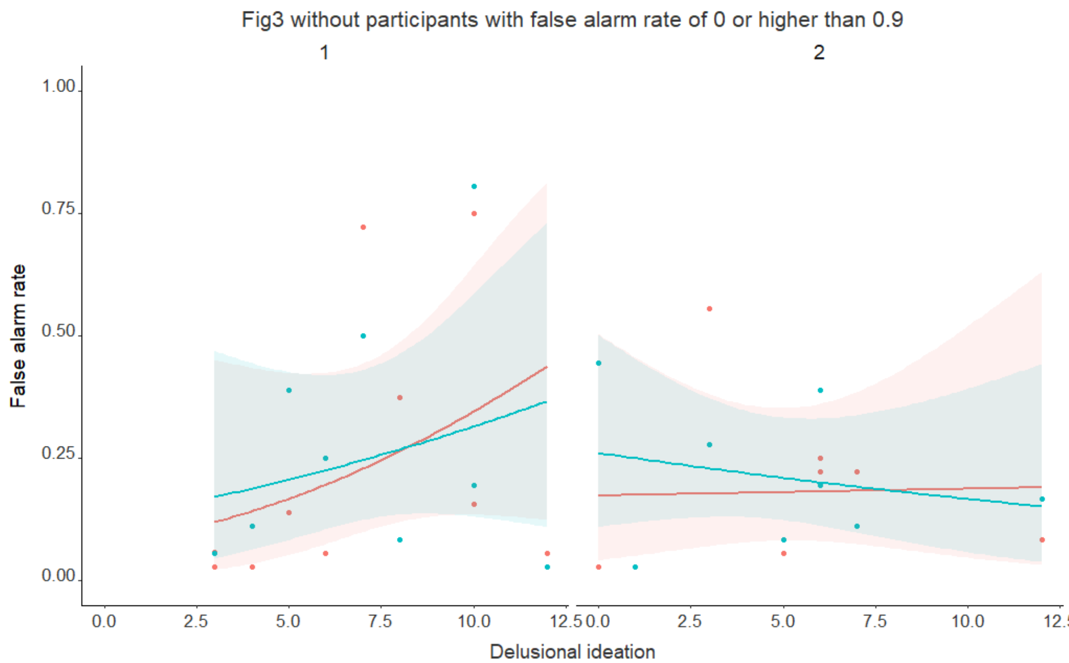

Figure S5. Figure 3 from the main text after removing participants with false alarm rate of 0 or higher than 0.9 in either condition (synchronous, asynchronous).

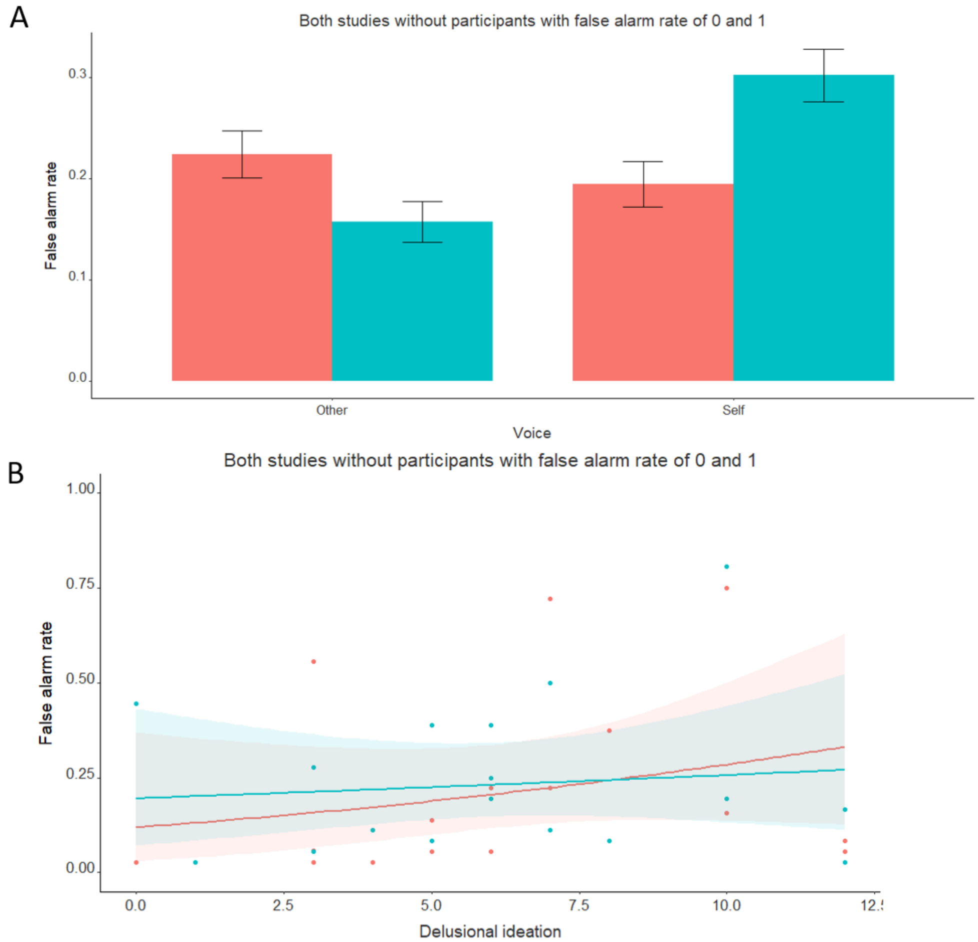

Figure S6. Merging participants from both studies after removing those with false alarm rate of 0 or higher than 0.9 in either condition (synchronous, asynchronous) and replicating Figures 2 (A) and 3 (B) from the main text.

Time evolution of false alarms

Supplementary figure S7 illustrates the interaction between the effects of Stimulation and Trial described in the main text.

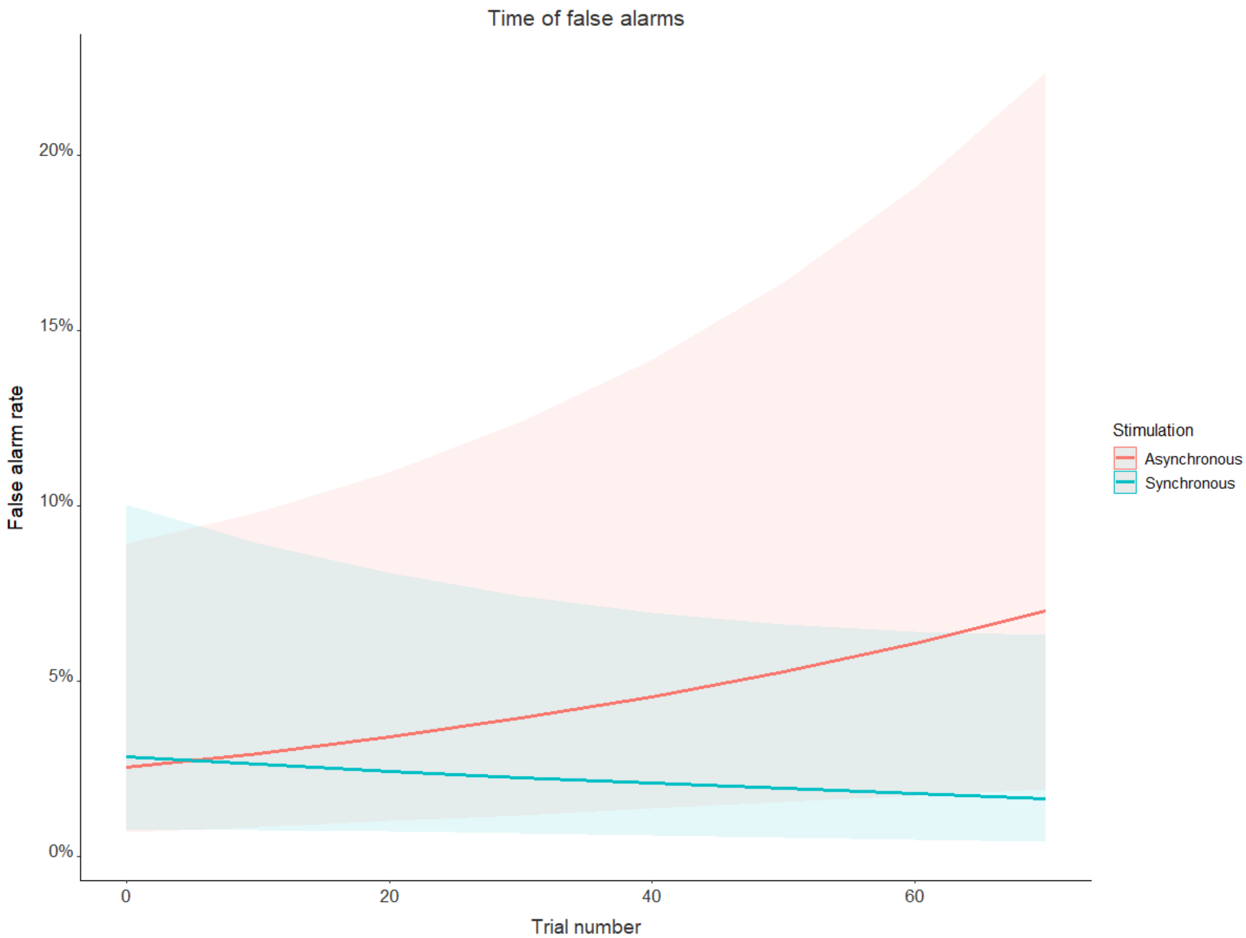

Figure S7. False alarm rate increased with time during blocks with asynchronous stimulation.

To investigate whether the occurrence of hits changed as a function of time, we ran equivalent binomial mixed-effects regressions with hit rate as dependent variable and an additional factor Trial (with values 1-63, indicating a trial within a block) together with a three-way interaction between Trial, Stimulation, and Voice. There were no significant effects (Supplementary Table T5).

Table T5. Binomial mixed-effects regression for the effects of Trial on hit rate.

	estimate	*Z*	*p*
Trial	0	-0.32	0.748
*Trial * Stimulation*	0	0.2	0.843
*Trial * Voice*	0	1.37	0.169
*Trial * Stimulation * Voice*	0.52	0.17	0.867
Study	0.31	1.35	0.178

Supplementary references

Alderson-Day B, Moseley P, Mitrenga K, Moffatt J, Lee R, Foxwell J, Hayes J, Smailes D, Fernyhough C (2022) Varieties of felt presence? Three surveys of presence phenomena and their relations to psychopathology. Psychol Med *Psychological medicine* 1–9.

Alderson-Day B, Woods A, Moseley P, Common S, Deamer F, Dodgson G, Fernyhough C (2021) Voice-Hearing and Personification: Characterizing Social Qualities of Auditory Verbal Hallucinations in Early Psychosis. Oxford University Press (OUP) *Schizophrenia Bulletin* **47**, 228–236.

Peters E, Joseph S, Day S, Qarety P (2004) Measuring delusional ideation: the 21-item Peters et al. Delusions Inventory (PDI). *Schizophrenia Bulletin* **30**, 1005–1022.

Schmack K, de Castro AGC, Rothkirch M, Sekutowicz M, Rössler H, Haynes JD, Heinz A, Petrovic P, Sterzer P (2013) Delusions and the role of beliefs in perceptual inference. *Journal of Neuroscience* **33**, 13701–13712.