Increased excitation enhances the sound-induced flash illusion by impairing multisensory causal inference in the schizophrenia spectrum - Supplementary Information

1 Participants

We modelled the data collected in the study conducted by Ferri et al. (2018). One hundred and ninety-six adult volunteers were screened for schizotypal traits using the Schizotypal Personality Questionnaire (SPQ) (Raine, 1991). The questionnaire was administered online through Qualtrics. Participants were categorized into quintiles based on their SPQ scores. 64 participants participants falling within the first and fifth quintile were selected: 32 with low schizotypy scores (range: 5-16) and 32 with high schizotypy scores (range: 36-68). All participants had normal or corrected vision and hearing and no history of substance abuse or other psychiatric disorders. Age and gender were matched between the two groups. Demographic details are presented in Table 1.

Three participants in the low schizotypy group and four in the high schizotypy group were excluded from the analysis due to unreliable responses. This resulted in a final sample of 57 participants: 29 in the low schizotypy group and 28 in the high schizotypy group. All participants provided written informed consent before participating in the study.

Trait	L-SPQ	H-SPQ
SPQ Score Range	5-16	36-68
No. of Females	25	25
Mean Age (SD)	29 (9)	26 (7)
British White	41%	44%
Other White	50%	50%
Asian/Indian	9%	6%

Table 1: Demographics

2 Temporal Window of Illusion

Following Ferri et al. (2018), to determine the Temporal Window of Illusion (TWI), we analysed the proportion of trials in which participants reported perceiving the DFI (i.e., ”two” responses) at each Stimulus Onset Asynchrony (SOA). For each participant, we plotted the proportion of DFI reports against the SOA and fitted a psychometric sigmoid function to these data points:

$$y = a + \frac{b}{1 + \exp\left(-\frac{(x-c)}{d}\right)} \tag{1}$$

where:

$$\begin{aligned} y &: \text{Proportion of DFI reports} \\ x &: \text{SOA} \\ a &: \text{Upper asymptote} \\ b &: \text{Lower asymptote} \\ c &: \text{Inflection point (TWI)} \\ d &: \text{Slope} \end{aligned}$$

The estimated SOA (ms) corresponding to the inflection point (c) of the sigmoid function was defined as the individual's TWI, representing the temporal window within which the illusion was maximally experienced.

3 Multisensory causal inference network model

The network model used in this study is an extension of previous models developed by Cuppini and his collaborators (Cuppini et al., 2014, 2017). The model consists of three layers: two encode auditory and visual stimuli, separately, and connect to a multisensory layer where causal inference is computed. Each of these layers consists of 30 neurons arranged topologically to encode a 30° space. Hence, each neuron encodes 1° of space and neurons close to each other encode close spatial positions.

Each neuron will be noted with a superscript c indicating a specific cortical area (a, v or m for the auditory, visual or multisensory area respectively). Similarly, each neuron will hold a subscript j refereed to its spatial position within a given area.

Neurons in each layer have a sigmoid activation function and first-order dynamics:

$$\tau^c \frac{dy_j^c(t)}{dt} = -y_j^c(t) + F\left(u_j^c(t)\right), \qquad c = a, v, m \tag{2}$$

Here, $u(t)$ and $y(t)$ are used to represent the net input and output of a given neuron at time t. τ^c denotes the time constant of neurons belonging to a given area c. $F(u)$ represents the sigmoidal relationship:

$$F\left(u_j^c\right) = \frac{1}{1 + exp^{-s\left(u_j^c - \theta\right)}} \tag{3}$$

Here, s and θ denote the slope and the central position of the sigmoidal relationship respectively. Neurons in all regions differ only in their time constants, chosen to mimic faster sensory processing for stimuli in the auditory region compared to visual stimuli.

The net input of a neuron is the sum of an inside (i.e. within region) component (l_j^c) and an outside (i.e. extra-area) component ($o_j^c(t)$):

$$u_j^c\left(t\right) = l_j^c\left(t\right) + o_j^c\left(t\right) \tag{4}$$

The within region component l_j^c is defined as:

$$l_j^c\left(t\right) = \sum_k L_{jk}^c \cdot y_{jk}^c\left(t\right) \tag{5}$$

Here L_{jk}^c represents the strength of the lateral synapse from a presynaptic neuron at position k to a postsynaptic neuron at position j in the region c. y_{jk}^c is the activity of the presynaptic neuron at position k.

Such synapses are symmetrical and arranged according to a "Mexican hat" pattern (i.e. a central excitatory area surrounded by an inhibitory ring for each neuron, so that the entire layer generates excitation for spatially close stimuli and inhibition for distant stimuli):

$$L_{jk}^c = \begin{cases} L_{0ex}^c \cdot exp\left(-\frac{(D_{jk})^2}{2\cdot(\sigma_{ex}^c)^2}\right) - L_{0in}^c \cdot exp\left(-\frac{(D_{jk})^2}{2\cdot\left(\sigma_{in}^c\right)^2}\right), & D_{jk} \neq 0 \\ 0, & D_{jk} = 0 \end{cases} \tag{6}$$

L_{0ex}^c and L_{0in}^c denote the highest level of excitatory and inhibitory synaptic efficacy in the region c, respectively. D_{jk} indicate the distance between the pre-synaptic neuron and the post-synaptic neurons within a given area:

$$D_{jk} = \begin{cases} |j-k|, & |j-k| \leq N/2 \\ N - |j-k|, & |j-k| > N/2 \end{cases} \tag{7}$$

This defines a circular structure where each neuron receives the same number of lateral connections.

Importantly, the extra-area input is defined differently for unisensory and multisensory areas. The extra-area input for the unisensory areas includes a stimulus from the external world ($e_j^c(t)$), a cross-modal component coming from the other unisensory area ($c_j^c(t)$), a feedback component coming from the multisensory area ($b_j^c(t)$) and a noise component (n_j^c).

The stimulus from the external world is simulated as a 1-D Gaussian function to represent the uncertainty in the detection of external stimuli:

$$e_j^c\left(t\right) = E_0^c \cdot exp\left(-\frac{\left(d_j^c\right)^2}{2\left(\sigma^c\right)^2}\right) \tag{8}$$

Here, E_0^c denotes the strength of the stimulus, d_j^c the distance between neuron at position j and the stimulus at position p^c, and σ^c the degree of uncertainty in sensory detection. The distance d_j^c is defined as:

$$d_j^c = \begin{cases} \left|j - p^c\right|, & \left|j - p^c\right| \leq N/2 \\ N - \left|j - p^c\right|, & \left|j - p^c\right| > N/2 \end{cases} \tag{9}$$

Furthermore, the cross-modal input is defined as:

$$\begin{aligned} c_j^a\left(t\right) &= \textstyle\sum_{k=1}^{N} W_{jk}^{av} \cdot y_k^v\left(t - \Delta t_{cross}\right) \\ c_j^v\left(t\right) &= \textstyle\sum_{k=1}^{N} W_{jk}^{va} \cdot y_k^a\left(t - \Delta t_{cross}\right) \end{aligned} \tag{10}$$

Here Δt_{cross} represents the latency of cross-modal inputs between two unisensory regions. The synaptic weights are symmetrically defined ($W_0^{av} = W_0^{va}$ and $\sigma^{av} = \sigma^{va}$) by the Gaussian function:

$$W_{jk}^{cd} = W_0^{cd} \cdot exp\left(-\frac{\left(D_{jk}\right)^2}{2\left(\sigma^{cd}\right)^2}\right), cd = av\ or\ va \tag{11}$$

W_0 denotes the highest level of synaptic efficacy and D_{jk} is the distance between neuron at position j in the post-synaptic unisensory region and the neuron at position k in the pre-synaptic unisensory region. σ^{cd} defines the width of the cross-modal synapses.

Furthermore, the feedback input is defined as:

$$b_j^c\left(t\right) = \sum_{k=1}^{N} B_{jk}^{cm} \cdot y_k^c\left(t - \Delta t_{feed}\right) \tag{12}$$

Here Δt_{feed} represents the latency of feedback inputs between the multisensory and unisensory regions. The feedback synaptic weights are also symmetrically ($B_0^{am} = B_0^{vm}$ and $\sigma^{am} = \sigma^{vm}$) defined:

$$B_{jk}^{cm} = B_0^{cm} \cdot exp\left(-\frac{\left(D_{jk}\right)^2}{2\left(\sigma^{cm}\right)^2}\right), cm = am\ or\ vm \tag{13}$$

Here, B_0^{cm} denotes the highest level of synaptic efficacy and D_{jk} is the distance between neuron at position j in the post-synaptic unisensory region and the neuron at position k in the pre-synaptic multisensory region. σ^{cd} defines the width of the feedback synapses.

The noise component n_j^c is extracted from a standard uniform distribution on the interval $[n_{max} + n_{max}]$. Here n_{max} is defined as the 40% of the strength of the external stimulus for each modality.

All these external sources are filtered by a second order differential equation to mimic the temporal dynamics of the stimuli in a cortex:

$$\begin{cases} \frac{d}{dt} o_j^c(t) = \delta_j^c(t) \\ \frac{d}{dt} \delta_j^c(t) = \frac{G^c}{\tau^c} \cdot \left[e_j^c(t) + c_j^c(t) + b_j^c(t) + n_j^c \right] - \frac{2 \cdot \delta_j^c(t)}{\tau^c} - \frac{o_j^c(t)}{(\tau^c)^2} \end{cases}, c = a, v \quad (14)$$

Here, G^c represents gain and τ^c the time constants of the dynamics.

In contrast, the extra-area input for the multisensory area comes only from feedforward synapses from the two unisensory areas:

$$i_j^m(t) = \sum_{k=1}^{N} W_{jk}^{ma} \cdot y_k^a(t - \Delta t_{feed}) + W_{jk}^{mv} \cdot y_k^v(t - \Delta t_{feed}) \quad (15)$$

Here Δt_{feed} represents the latency of feedforward inputs between the unisensory and multisensory regions. W_{jk}^{ma} and W_{jk}^{mv} are the synapses connecting the pre-synaptic neuron at position k in a given unisensory area and the post-synaptic neuron at position j in the multisensory area.

The weights of these feedforward synapses are symmetrically defined as:

$$W_{jk}^{mc} = W_0^{mc} \cdot exp\left(-\frac{(D_{jk})^2}{2(\sigma^{mc})^2} \right), c = a, v \quad (16)$$

Here W_0^{mc} denotes the highest value of synaptic efficacy, D_{jk} the distance between the multisensory neuron at position j and the unisensory neuron at position k, and σ^{mc} the width of the feedforward synapses.

These external sources are also filtered by a second order differential equation:

$$\begin{cases} \frac{d}{dt} o_j^m(t) = \delta_j^m(t) \\ \frac{d}{dt} \delta_j^m(t) = \frac{G^m}{\tau^m} \cdot \left[i_j^m(t) \right] - \frac{2 \cdot \delta_j^m(t)}{\tau^m} - \frac{o_j^m(t)}{(\tau^m)^2} \end{cases} \quad (17)$$

Here, G^m represents gain and τ^m the time constants of the dynamics in the multisensory neurons.

4 Values of L-SPQ model parameters

	Neurons		
$N = 30$	$\theta = 20$	$s = 0.3$	$\tau = 1$
	External stimuli		
$E_0^a = 2.325$	$\sigma^a = 32^\circ$	$E_0^v = 1.45$	$\sigma^v = 4^\circ$
	Input dynamics		
$G^{a,v,m} = e^1$	$\tau^a = 6.56$	$\tau^v = 9.19$	τ^m=120
$\Delta t_{cross} = 16$ ms	$\Delta t_{feed} = 95$ ms		
	Lateral synapses		
$L_{0ex}^c = 0.5$	$L_{0in}^c = 0.4$	$\sigma_{ex}^c = 3^\circ$	$\sigma_{in}^c = 24^\circ$
	Inter-areal synapses		
$W_0^{mc} = 3.892$	$\sigma^{mc} = 0.5^\circ$	$B_0^{cm} = 0.623$	$\sigma^{cm} = 0.5^\circ$
$W_0^{av,va} = 0.001$	$\sigma^{av,va} = 5^\circ$		

Table 2: Values of the parameters in the L-SPQ model. This parameterisation remained fixed throughout the simulations computed to identify the H-SPQ model.

5 Parameter exploration

We systematically varied the parameters governing recurrent excitation (L_{0ex}), the strength of feedback (B_0^c) and cross-modal (W_0^{cd}) weights, and feedforward ($\rho_{W^{mc}}$) and cross-modal ($\rho_{W^{cd}}$) pruning to explore plausible mechanisms behind the changes in the temporal window of illusion (TWI) observed in individuals with H-SPQ (see Figure 3).

In this parameter exploration, we examined seven uniformly distributed values within the operational range of each aforementioned parameter. The operational range is characterized by the parameter variations for which the base model (L-SPQ) demonstrates responses that align with those observed in human participants during the behavioral experiment (Ferri et al., 2018). The bounds considered for the parameter exploration for L_{0ex}, B_0^c, W_0^{cd}, $\rho_{W^{mc}}$, and $\rho_{W^{cd}}$ were (0.5, 0.65), (0.24, 0.25), (0.01, 0.055), (0, 0.35), and (0, 0.013), respectively.

6 Fitting procedure

The probability of perceiving two flashes in a single trial $P(2f)$ was defined as the product of the two peak values above .15 read from visual neurons. We used a fitting procedure with the cost function defined by Equation 18.

$$Cost = \sum_{i=1}^{N} \left(\frac{P(2f)_i^{data} - P(2f)_i^{model}}{P(2f)_i^{data}} \right)^2 \tag{18}$$

Here, $P(2f)_i^{data}$ and $P(2f)_i^{model}$ denote the P (2f) measured in the ith stimulus onset asynchrony (SOA). N represents the number of SOAs measured (i.e. 15 in the empirical study).

This cost function was minimised by the implementation of the differential evolution algorithm (Storn and Price, 1997) available in the SciPy v1.11.4 library for the Python programming language (Virtanen et al., 2020).

$P(2f)^{model}$ is composed of the 15 $P(2f)$ generated by the model at each SOA defined in the experimental task. Similarly, $P(2f)^{data}$ is composed of 15 proportion values taken from the mean sigmoid fit obtained out of the experimental data of a given group. We assumed that the percentage of perceived second illusory flash reported in the original study Ferri et al. (2018) is equivalent to the probability of perceiving two flashes in each condition.

6.1 L-SPQ Fitting

We fitted the model to the L-SPQ group data. Here, W_0^{mc}, $W_0^{av,va}$, B_0^{cm}, τ^a, τ^v and τ^m (see Equations 16, 11, 13 14 and 17) varied freely to represent the group proneness to the illusion due to stable anatomical factors (i.e. immutable throughout the duration of the experiment). All other parameters (see Table 2) remained fixed and set to the values published by Cuppini et al. (2014, 2017). The bounds given to the optimisation algorithm to find τ^a, τ^v, τ^m, $W_0^{av,va}$, B_0^{cm} and W_0^{mc}, were (6, 20), (6.25, 60), (6, 120), (0.001, .05), (0.001, 1) and (0.001, 15) respectively. The results of the fit to the data of the L-SPQ group are presented in Table 2.

We also fitted an alternative L-SPQ model without the causal inference area. We maintained the same base parameterisation, and used the optimisation algorithm to find τ^a, τ^v and $W_0^{av,va}$. The bounds given were (6, 20), (6.25, 60), (6, 120) and (0.001, .05). This latter model failed to reproduce responses to SOAs beyond 60 ms, as shown in Figure 4A.

6.2 H-SPQ Fitting

We fitted the L-SPQ model to the H-SPQ data. The bounds given to the optimisation algorithm to find L_{0ex}^c and $W_0^{av,va}$ were (1.5, 1.7) and (0, 0.06) respectively. These limits were chosen because the response profile generated by values outside of this range ceases to resemble a sigmoid curve, which is characteristic of the average response profile found in the empirical study (Ferri et al., 2018).

It is observed that the model may not accurately reproduce the DFI responses observed at very short inter-beep intervals (below 75 ms). This discrepancy may be attributed, at least in part, to the fixed assumptions pertaining to the temporal delays of cross-modal and feedforward-feedback inputs incorporated within the model. Additionally, it is essential to acknowledge that the literature indicates substantial variability in DFI responses at very short inter-beep intervals across different individuals (Hirst et al., 2020). We recognize that further refinement of the model's parameters and architecture may be requisite to more precisely capture the complexities of DFI perception over the entire spectrum of inter-beep intervals.

References

Cuppini, C., Magosso, E., Bolognini, N., Vallar, G., Ursino, M., 2014. A neurocomputational analysis of the sound-induced flash illusion. Neuroimage 92, 248–266. doi:10.1016/j.neuroimage.2014.02.001.

Cuppini, C., Shams, L., Magosso, E., Ursino, M., 2017. A biologically inspired neurocomputational model for audiovisual integration and causal inference. Eur. J. Neurosci. 46, 2481–2498. doi:10.1111/ejn.13725.

Ferri, F., Venskus, A., Fotia, F., Cooke, J., Romei, V., 2018. Higher proneness to multisensory illusions is driven by reduced temporal sensitivity in people with high schizotypal traits. Conscious. Cogn. 65, 263–270. doi:10.1016/j.concog.2018.09.006.

Hirst, R.J., McGovern, D.P., Setti, A., Shams, L., Newell, F.N., 2020. What you see is what you hear: Twenty years of research using the Sound-Induced Flash Illusion. Neurosci. Biobehav. Rev. 118, 759–774. doi:10.1016/j.neubiorev.2020.09.006.

Raine, A., 1991. The spq: a scale for the assessment of schizotypal personality based on dsm-iii-r criteria. Schizophrenia bulletin 17, 555–564.

Storn, R., Price, K., 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization 11, 341–359.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al., 2020. Scipy 1.0: fundamental algorithms for scientific computing in python. Nat. Methods 17, 261–272.