Citation: Freeman, E. D. (2020). Hearing what you see: distinct excitatory and disinhibitory mechanisms contribute to visually-evoked auditory sensations. Cortex, 131, pp. 66-78. doi: 10.1016/j.cortex.2020.06.014

Hearing what you see: distinct excitatory and disinhibitory mechanisms contribute to visually-evoked auditory sensations.

Elliot Freeman

Cognitive Neuroscience Research Unit, Department of Psychology,
City, University of London, UK

Corresponding author:

Dr Elliot Freeman
City, University of London
Northampton Square
London EC1V 0HB
United Kingdom
Phone: +44 7739765561
Email: elliot.freeman@city.ac.uk

Keywords:

Synaesthesia; individual differences; Audiovisual perception

Highlights

Some people hear what they see. This research provides, for the first time:

(1) an objective diagnostic of visual-ear synaesthesia (vEAR), which relates their subjective experience of sounds evoked by a moving image to its physical ‘motion energy’, via a model of the physiology of motion processing in early visual cortex;

(2) independent support for the hypothesis that vEAR depends on increased cortical disinhibition, evidenced by an association between higher video ratings and weaker visual surround suppression;

(3) evidence for distinct excitatory and disinhibitory mechanisms that each independently predict the perceived intensity of vEAR, and also respectively predict susceptibility to different clusters of associated traits, such as migraine and pattern glare versus auditory-induced phosphenes and vEAR self-ratings.

Abstract

Visual motion or flashing lights can evoke auditory sensations in some people. This putative form of synaesthesia (visually-evoked auditory response, vEAR) may depend on cortical excitation and/or disinhibition of early visual and/or auditory brain areas.

To test this, the present large-scale internet study measured (1) subjective ratings of the vividness of auditory sensations evoked by a series of looping videos; (2) an independent psychophysical measure of visual contrast surround-suppression, though to index inhibition in the early visual system; (3) ratings for a range of traits that may be associated with cortical excitability and/or disinhibition.

As predicted by a disinhibition account of vEAR, video ratings were overall higher in individuals showing weaker surround-suppression. Independently, video ratings reliably depended on the amount of Motion Energy (ME) in each video, calculated using a popular computational model of motion processing in early visual cortex. The strength of this association for each individual provided an objective diagnostic measure of vEAR 'ME-sensitivity'. Interestingly, surround-suppression and ME-sensitivity did not correlate, suggesting that they each represent different mechanisms contributing to vEAR. Additionally, both surround-suppression and ME-sensitivity each independently predicted different clusters of trait measures: Surround-suppression was associated with vEAR self-ratings and auditory-evoked visual phosphenes, while ME-sensitivity was independently associated with ratings of other traits including susceptibility to migraine and pattern glare.

Altogether, these results suggest there are two independent mechanisms underlying vEAR and its associated traits, based putatively on disinhibition versus cortical excitability.

Introduction

For some people the sight of flashing shop displays, people walking, or any visual movement can evoke a phantom auditory sensation. This phenomenon was first described by Saenz & Koch (2008) in a small number of self-selected participants as a new form of synaesthesia. Our own recent investigations have uncovered more detail about this curious phenomenon, which we call the 'visually evoked auditory response' (vEAR or 'visual ear'). Findings to date suggest that vEAR may be substantially more prevalent than canonical varieties of synaesthesia (approximately 20%), and that it correlates with a broad range of perceptual traits, while the visually-evoked phantom sounds can interfere with detection of real faint auditory signals (Fassnidge, Cecconi Marcotti, & Freeman, 2017; Fassnidge & Freeman, 2018). We have previously proposed that vEAR may depend on individual differences in cortical excitability or disinhibition, and this is supported by our recent evidence of reduced competition between auditory and visual cortex (Fassnidge et al., 2019) as well as independent evidence of greater excitability in visual cortex (Rothen, Bartl, Franklin, & Ward, 2017). However, in contrast with other synaesthesias for which there are objective tests of genuineness (Simner et al., 2006), an objective diagnostic measure of vEAR has so far been lacking. The present study sought to validate an objective test of vEAR reports based on the relationship between ratings of the auditory-vividness of videos, and a physical measure of the amount of raw 'motion energy' contained in the videos. A second goal was to assess the hypothetical role of sensory disinhibition (Grossenbacher & Lovelace, 2001) in vEAR using an independent psychophysical measure of visual surround suppression of apparent contrast (Xing & Heeger, 2001). A third aim was to map the association of the above objective measures with a broad array of perceptual traits, taking precautions to avoid response biases.

In canonical synaesthesias, objective tests of genuineness have been devised to establish whether subjective reports of synaesthetic experiences relate to genuine sensory experiences (Eagleman, Kagan, Nelson, Sagaram, & Sarma, 2007). For example, in grapheme-colour synaesthesia such a test assesses the reliability with which an individual associates specific letters with specific colours, over repeated testing sessions, or whether speeded identification of a letter is implicitly aided or disrupted by displaying it in colours that are congruent versus incongruent with the individual's synaesthetic colour. While statistical regularities may exist in some cases (Bor, Rothen, Schwartzman, Clayton, & Seth, 2014; Witthoft, Winawer, & Eagleman, 2015), it can be challenging to identify lawful relationships between the physical properties of a given inducer and its concurrent sensation. In contrast, the intensity of vEAR does seem to depend lawfully on the intensity of raw motion energy in the visual stimulus, as we established tentatively in an earlier study (Fassnidge & Freeman, 2018). We now seek further validation for an objective measure of vEAR genuineness based on this psychophysical relationship, which might provide insight into the underlying mechanisms.

In our previous study (Fassnidge & Freeman, 2018), participants were asked to rate the intensity of any auditory sensations evoked by live-action videos. We quantified the motion energy (ME) in each video using a simple model of spatiotemporal sensitivity of cells in early visual cortex to moving patterns (Adelson & Bergen, 1985), and used ME estimates to predict ratings for each video. The strength of the correlation per individual provides, in principle, an objective and psychophysiological underpinning for subjective ratings of vEAR, which could serve as a measure of genuineness. However, some videos depicted collisions, bouncing, and vocalisation, and responses appeared to be strongly dominated by expectations of the associated sounds. Although some forms of vEAR may validly depend on semantic associations as do other forms of synaesthesia (Mattingley, Rich, Yelland, & Bradshaw, 2001; Myles, Dixon, Smilek, & Merikle, 2003; Smilek, Dixon, Cudahy, & Merikle, 2001), our present goal is to seek a purer measure of vEAR based on the relationship between raw visual motion and ratings of the intensity of auditory sensations. To avoid bias from associations, the present study selected videos composed of computer-generated abstract motion rather than live action.

Our previous findings (Fassnidge et al., 2019; Fassnidge & Freeman, 2018) have led us to hypothesise that vEAR depends primarily on normally-occurring systemic variations in cortical excitability or disinhibition within sensory brain areas. This may result from disinhibition of feedback from higher areas to unimodal sensory areas (Grossenbacher & Lovelace, 2001; Neufeld et al., 2012), or more directly via disinhibition within the sensory areas themselves (Lalwani & Brang, 2019). This ‘disinhibition’ hypothesis contrasts with the ‘cross-activation’ hypothesis assuming idiosyncratically, rather than systemically, abnormal patterns of unusual neural connectivity (Bargary & Mitchell, 2008; Baron-Cohen, 1996; Hubbard & Ramachandran, 2005; Tomson et al., 2011). To further test this disinhibition hypothesis here, we have included a test of visual surround suppression, where the apparent contrast of a central test patch appears lower when surrounded by a similar high-contrast context (Chubb, Sperling, & Solomon, 1989; Xing & Heeger, 2001). This phenomenon is thought to depend on inhibitory gain control mechanisms in early visual cortex (Heeger, 1992). If vEAR depends on generally reduced cortical inhibition, this predicts that surround suppression should be reduced in vEAR. Surround suppression has been found to be reduced in patients diagnosed with schizophrenia (Dakin, Carlin, & Hemsley, 2005), which is associated with a deficit in the inhibitory neurotransmitter GABA (Tibber et al., 2013; Yoon et al., 2009). However evidence for a similar mechanism in synaesthesia is limited (Terhune, Song, Duta, & Cohen Kadosh, 2014).

The hypothesised role of systemic variables such as disinhibition received support from our previous finding that video ratings correlated with a variety of nominally unrelated perceptual traits (Fassnidge & Freeman, 2018), including the tendency to experience musical imagery, tinnitus, and also the little-known phenomenon of auditory-evoked visual phosphenes (Jacobs, Karpik, Bozian, & Gøthgen, 1981; Lessell & Cohen, 1979; Nair & Brang, 2019), which is an example of auditory-to-visual cross-talk. However, it is possible that some of these associations were subject to an acquiescence bias (‘yea-saying’), where some participants might have tended to respond generally more positively than others to all questionnaire items. In an attempt to neutralise such a bias, the present study introduced a reverse-coded alternative for each trait question (see **Table 1**). It was randomly determined for each participant whether a given question had a positively or negative wording. We also extended the trait questionnaire, adding questions about musicality given our previous finding of higher vEAR prevalence in musicians (Fassnidge et al., 2019), and a question about the effects of background noise on speech comprehension, given its strong dependence on visual cues such as lipmovements (Ipser et al., 2017; Sumby & Pollack, 1954); we also included questions about traits that may be associated with sensory excitability or disinhibition, in order to test whether this might provide an explanation for vEAR. For example we asked about susceptibility to migraine aura (Palmer, Chronicle, Rolan, & Mulleners, 2000; Tibber, Kelly, Jansari, Dakin, & Shepherd, 2014), pattern glare (Wilkins et al., 1984), insomnia (Van Der Werf et al., 2010) and photic sneezing (Langer, Beeli, & Jäncke, 2010).

Methods

Participants

The study protocol was approved by the Psychology Ethics committee at City, University of London. Participants were recruited anonymously via web-links appearing in popular media publications such as New Scientist (Firth, 2018) reporting on a previous publication about vEAR (Fassnidge & Freeman, 2018). Location data showed world-wide participation, concentrated mostly in western and central Europe, north America, and Brazil. Participants were not offered payment. Of the 21201 participants who began the questionnaire, 9232 completed it (44%). Incomplete records were excluded from analysis. Of the completing participants, 5170 identified as female, and 3804 as male. Mean age of participants under 65 was 34 years (SD 11); there were 152 participants over 65.

Materials and procedure

The survey was administered using Qualtrics, in English, and can be accessed from tinyurl.com/vEARsurveyNS (see also Figure 1). The survey contained sections probing a variety of traits, then eliciting ratings of video, and finally testing surround suppression.

The first section provided an explanation of the purpose of the study, *'An examination of the types of visual motion which evoke an internal auditory sensation'*, and information about anonymity and confidentiality. After informed consent, participants were prompted to identify their gender and select their age from following categories: 18-24, 25-34, 35-44, 45-54, 55-64, >65. The following introductory text and question about prior awareness of vEAR was then displayed:

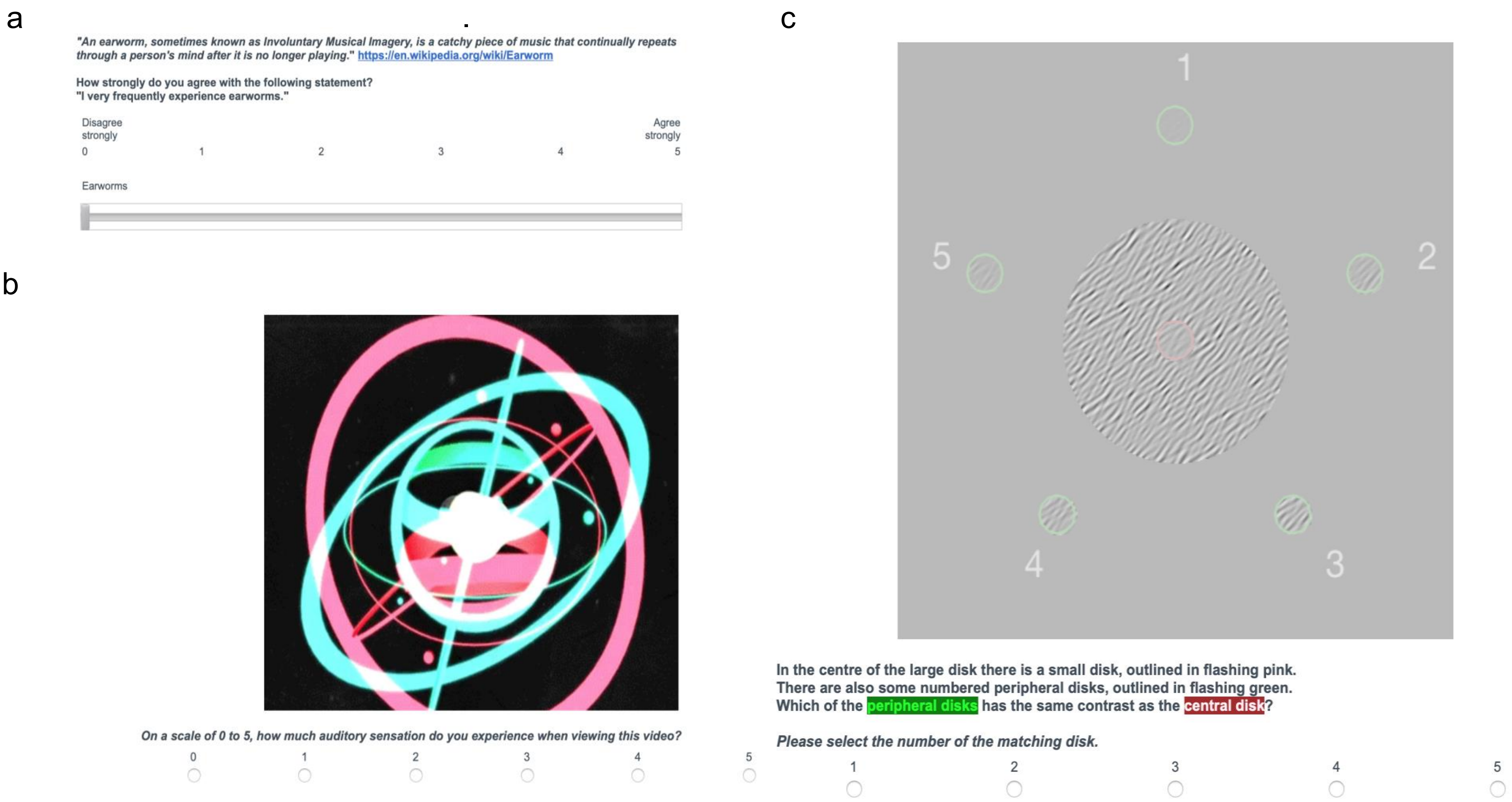
> *'We are interested in whether different types of visual motion evoke an imaginary sound, although in reality no such sound exists. The sound may be experienced within your head rather than in the outside environment. This may be perceived in a number of different ways. You may experience it as if you are vividly imagining the sound, or it may sound like a ringing in your ears, or it might resemble the experience of 'hearing' phrases of a popular song in your mind's ear, or the voices of people on television when watched with the volume off. Alternatively it may be an abstract experience, but closer to being an auditory experience than a visual experience. Some people describe it as imaginary white noise. To avoid confusion we will from now on refer to any such experience as 'auditory sensation' rather than hearing. What is important is that the auditory sensation occurs in time with visual change over time, caused by motion or sudden flashes. It is typically involuntary (i.e. it happens automatically rather than as a result of*

> *conscious effort) and it happens consistently.*
> *Have you previously been aware of experiencing this type of auditory sensation when viewing visual movement? [Yes / No / Not sure]'*

The next section included 11 compulsory questions about traits (**Table 1,** and see **Figure 1**a for an example display). Some items included brief introductory definitions and relevant weblinks, e.g. '*An earworm, sometimes known as Involuntary Musical Imagery, is a catchy piece of music that continually repeats through a person's mind after it is no longer playing. https://en.wikipedia.org/wiki/Earworm".* Each question item had two alternative wordings (**Table 1**). One version was positively worded (e.g. *'I suffer from tinnitus'*) and the other negatively (*'I do not suffer from tinnitus'*), and participants were required to indicate their agreement with the statement on a scale from 0 ('disagree strongly') to 5 ('agree strongly'). One version of each pair was randomly allocated to each participant. In addition, in some questions the absence of the trait was phrased positively (e.g. '*I can walk in bright sunshine without experiencing the urge to sneeze'*), and the presence was phrased negatively (e.g. '*I cannot...').* These features were intended to neutralise any acquiescence bias towards generally agreeing with statements, and to counteract any potential biases from social-desirability and demand characteristics present in the statements. In analysis, responses were reverse-coded as appropriate so that a higher rating indicated the presence of a trait. All questions were compulsory, however for one question about experiencing aura with migraine a 'not applicable' option was included for participants who do not suffer migraine.

The video rating section contained 20 colour videos sequences (**Figure 2b**), which looped continuously. Sequence duration varied between 0.4 and 8 seconds. Image size was 495x495 pixels. 16 royalty-free videos were downloaded from giphy.com with using the search term 'abstract'. Sequences were selected with the criterion that they should contain motion across a range of velocities and accelerations, be graphically abstract and avoid depicting natural scenes, or events that might naturally be associated with sounds, such as collisions, frictions or explosions. Two further videos depicted a grating undergoing periods of oscillatory motion at a range of frequencies. Two additional videos showed the much-publicised skipping-pylons sequence (credit: HappyToast). These typically obtained high ratings but were not included in the analysis, which focused on abstract rather than live-action depictive imagery. Initial instructions asked participants to *'rate the clips from 0 (no auditory sensation at all) to 5 (very vivid and definite auditory sensation)'*. For each video participants were then asked *'on a scale from 0 to 5, how much auditory sensation do you* *experience when watching this video?'*. Order of items was randomised for each participant.

The final section tested for surround suppression (**Figure 2c**). There were 14 gray level images, displayed in 8-bit colour depth, with dimensions 500x500 pixels (dimensions are given in pixels rather than visual angle because viewing distance could not be controlled). Gamma correction was applied using exponent of 2.2 for a typical computer display. Each stimulus comprised a central 'target' disk of 30x30 pixels, in the centre of a circular 'surround' disk measuring 200x200 pixels. The Michelson contrast of the target was always 30%, while the surround was displayed at 100% contrast. These stimuli comprised bandpass filtered noise centred on a wavelength of 6 pixels, and with a bandwidth of one octave. The filter was oriented at 45° or 135° ±20°. The central target orientation was either co-oriented or orthogonal to the orientation of the surround. Seven examples of each were tested. Each display also included five 'sample' disks, with the same dimensions and as the target patch, with the following contrasts 5%, 17.5%, 30%, 42.5%, 55%. These were arranged in random order in a circle of radius 180 pixels around the main centre-surround display and numbered 1 to 5. Coloured rings indicated the locations of the central target and peripheral samples, flashing briefly for 200ms every 5 seconds. Instructions under each display asked participants '*Which of the peripheral disks has the same contrast as the central disk?'.* Compulsory responses were entered using radio-buttons numbered 1 to 5.



"An earworm, sometimes known as Involuntary Musical Imagery, is a catchy piece of music that continually repeats through a person's mind after it is no longer playing." https://en.wikipedia.org/wiki/Earworm

How strongly do you agree with the following statement?
"I very frequently experience earworms."

Disagree strongly					Agree strongly
0	1	2	3	4	5

Earworms

On a scale of 0 to 5, how much auditory sensation do you experience when viewing this video?

0	1	2	3	4	5
○	○	○	○	○	○

In the centre of the large disk there is a small disk, outlined in flashing pink.
There are also some numbered peripheral disks, outlined in flashing green.
Which of the peripheral disks has the same contrast as the central disk?

Please select the number of the matching disk.

1	2	3	4	5
○	○	○	○	○

Figure 1 Examples items from vEAR survey: (a) trait question, (b) video rating, (c) surround suppression (with coloured placeholders which were only intermittently displayed).

Results

Analysis of the surround suppression data showed a significant effect of surround orientation on perceived contrast of the central test patch [t(9231) = 92.83, p<0.001, Cohen's D = 0.91]: co-oriented surrounds resulted in significantly lower matching contrast (Mean: 21.2% Michelson contrast, Standard Error 0.07%) compared to orthogonally-oriented surrounds (Mean: 27.32%, SE 0.068%). Surround suppression scores (SS) were derived for each individual by subtracting matching contrast for co-oriented surrounds from orthogonal. Positive SS values indicate greater suppression from co-oriented surrounds.

Internal consistency of video ratings was high [Cronbach's α= 0.96]. To derive a measure of individual sensitivity to Motion Energy, each individual's data were analysed in a linear regression, predicting ratings for each video from the motion energy of each video. The slope of the fitted model provided an individual measure of sensitivity to motion energy (ME-sensitivity). There was no significant correlation between SS and ME-sensitivity [r(9230) = .01, ns]. Histograms for each measure are shown in **Figure 2a and b,** showing smooth distributions. The spike in the ME-sensitivity distribution at zero represents participants who gave all videos the same rating (typically zero).

When asked whether they had previously experienced phenomena that matched our written description of vEAR, 40% of participants answered 'Yes', 31% were 'Not Sure' and 29% answered 'No'. The proportion of 'Yes' responses was substantially higher than previously found (Fassnidge & Freeman, 2018), but this may be due to the publicity that vEAR has been continuing to receive. Responses to this question were used to group participants and compare measures of SS and ME-sensitivity. SS scores were significantly lower in participants who responded 'Yes' to the question about Previous Awareness of vEAR compared to 'Not Sure' or 'No' respondents [F(2,9229) = 7.39, p<.0006] (see **Figure 2c**). Likewise, ME-sensitivity scores were significantly higher in participants who responded 'Yes' to the question about Previous Awareness of vEAR compared to 'Not Sure' or 'No' respondents [F(2,9229) = 45.71, p<.0001] (see **Figure 2d**).

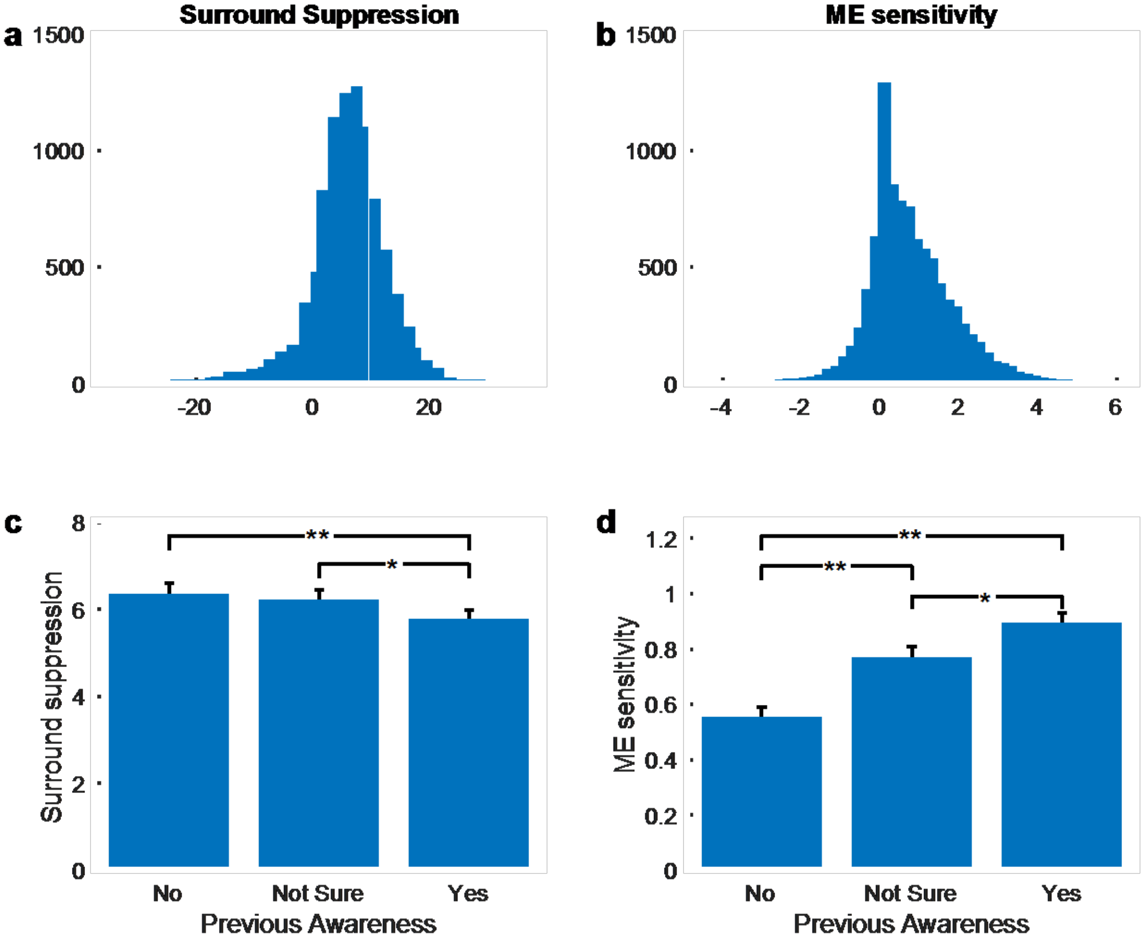


Figure 2 Histograms for (a) Surround suppression and (b) ME-sensitivity. Mean scores for (c) surround suppression and (d) ME-sensitivity, split by responses to a preliminary question about 'previous awareness' of vEAR. SE error bars; * p<.05; ** p<.01.

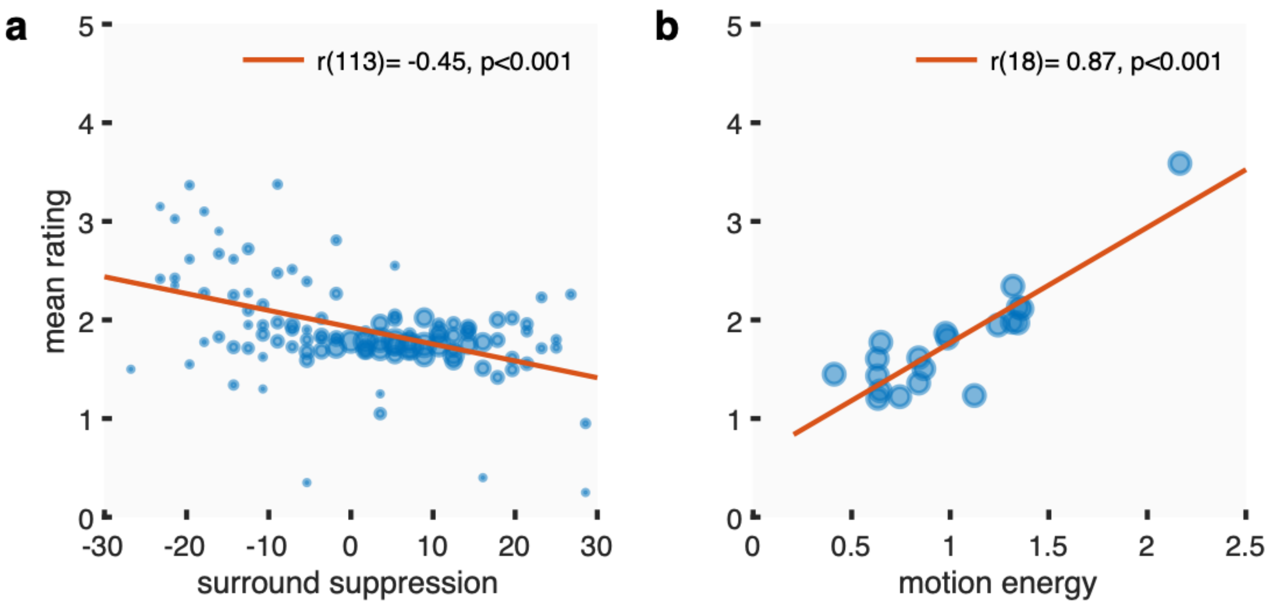


Figure 3 Scatterplots of mean video ratings against (a) surround suppression scores; (b) motion energy for different videos. Dot area represents number of averaged observations at each x-value. Least-squared line of best fit is shown in red, with Pearson's correlations coefficients in the legend.

In a set of linear mixed effects analyses, models were compared predicting individual video ratings from each individuals' Surround Suppression score (SS), and/or each video's ME, including subjects as a random variable. Log-likelihood of a model including both SS and ME predictors was significantly higher than a model including ME only [$\chi^2(1) = 8.17$, $p = 0.004$]. Adding the interaction term (SS x ME) made no significant difference to the log-likelihood estimate [$\chi^2(1) = 1.73$, ns]. The final model predicted 56.7% of the ratings variance with the following equation: Ratings = -0.0057*SS + 1.17*ME + ε [SS: $t = -2.86$, $p = 0.004$; ME: $t = 177.3$, $p < .0001$] (**Figure 3**). In summary, weaker surround suppression was associated with overall higher video ratings, but this effect worked additively rather than modulating the dependence of video ratings on motion energy.

All traits correlated significantly and positively with video ratings, regardless of whether the analysis included only positively-worded or negatively-worded questions [p<.003] (see **Figure 4** and **Table 2**). After Bonferroni correction, all correlations except Earworms (negatively-worded) remained significant [p<.05]. Response bias cannot easily account for the similar correlations with negatively-worded items. The following trait analyses combined data across positively-worded and negatively-worded versions.

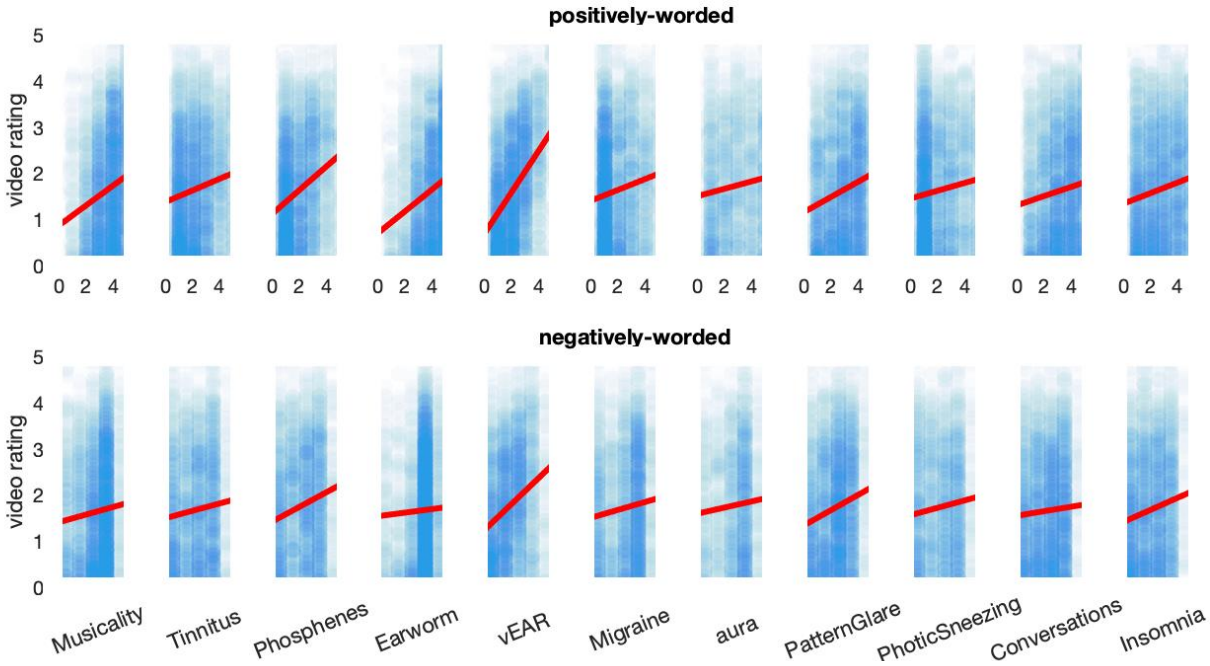


Figure 4 Scatterplots illustrating correlations between video ratings and ratings for different traits; separate rows for positively-worded and negatively-worded trait questions. Colour saturation represents frequency of responses.

A set of separate regression analyses next assessed whether ME-sensitivity and SS scores could predict each of the traits questions (**Table 3**). ME-sensitivity predicted all but two traits (‘photic sneezing’ and ‘speech-in-noise impairment’), all with slopes significantly greater than zero (p<.05, after Bonferroni correction, see yellow bars in **Figure 5a**). Weaker surround suppression independently predicted higher ratings for Musicality, Phosphenes, and vEAR, showing significantly negative slopes (**Figure 5b**). Although often highly significant, these predictors only accounted for a very small fraction of the overall variance in trait ratings (**Table 3**). Repeated analysis including an interaction term showed no significant interactions between ME-sensitivity and SS, and a negligible improvement in model fits (mean .02%, SE .007).

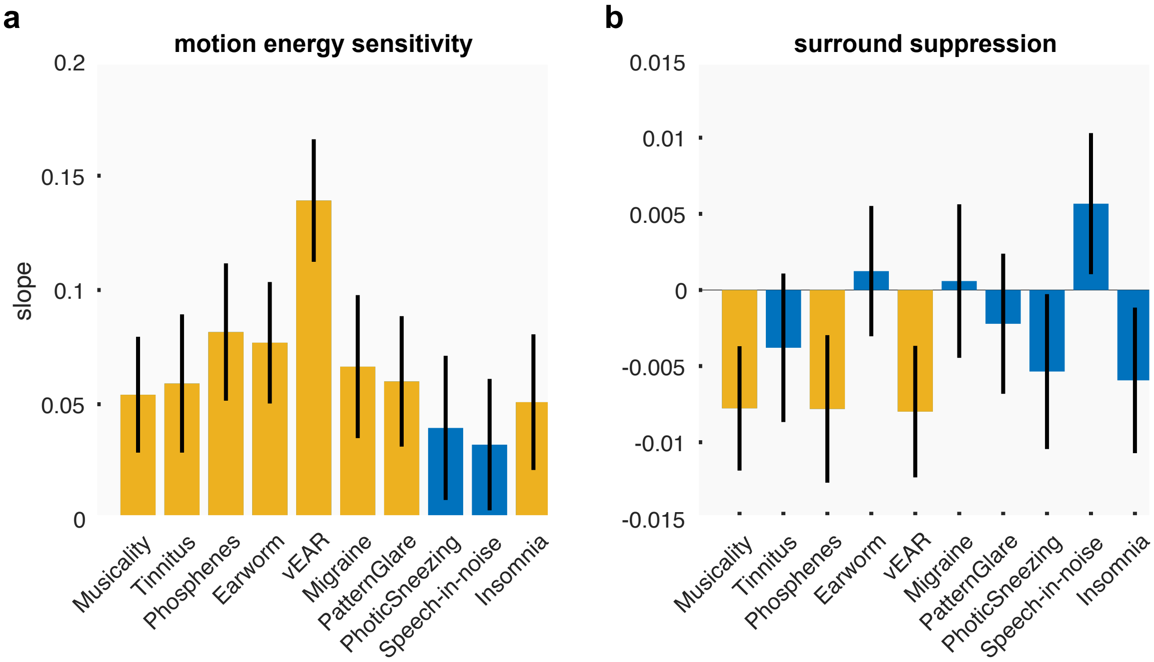


Figure 5 Results of linear regression analyses, predicting ratings for each trait from (a) ME-sensitivity scores and (b) surround suppression scores (i.e. Rating = ME + SS + ε). Bars in each graph show the slope relating each predictor score to the rating. Yellow bars indicate where slopes are significantly different from zero (p<.05, Bonferroni corrected). Errorbars show uncorrected 95% confidence intervals.

A final factor analysis tested whether ME-sensitivity and surround associations are each respectively associated with distinct clusters of traits. The analysis was limited to two factors. Traits loading the most strongly and uniquely onto the Factor 1 included Migraine, Pattern Glare, Insomnia, and difficulty with Speech-in-Noise comprehension (factor loadings >.3); Factor 2 uniquely included vEAR self-ratings, and auditory-evoked Phosphenes (**Table 4**). Factor loadings are shown in **Figure 6a**. As shown in **Figure 6b**, ME-sensitivity correlated significantly and positively with both Factor 1 and Factor 2 scores, albeit weakly; SS correlated significantly but negatively only with Factor 2 scores, again weakly [p<.005 corrected].

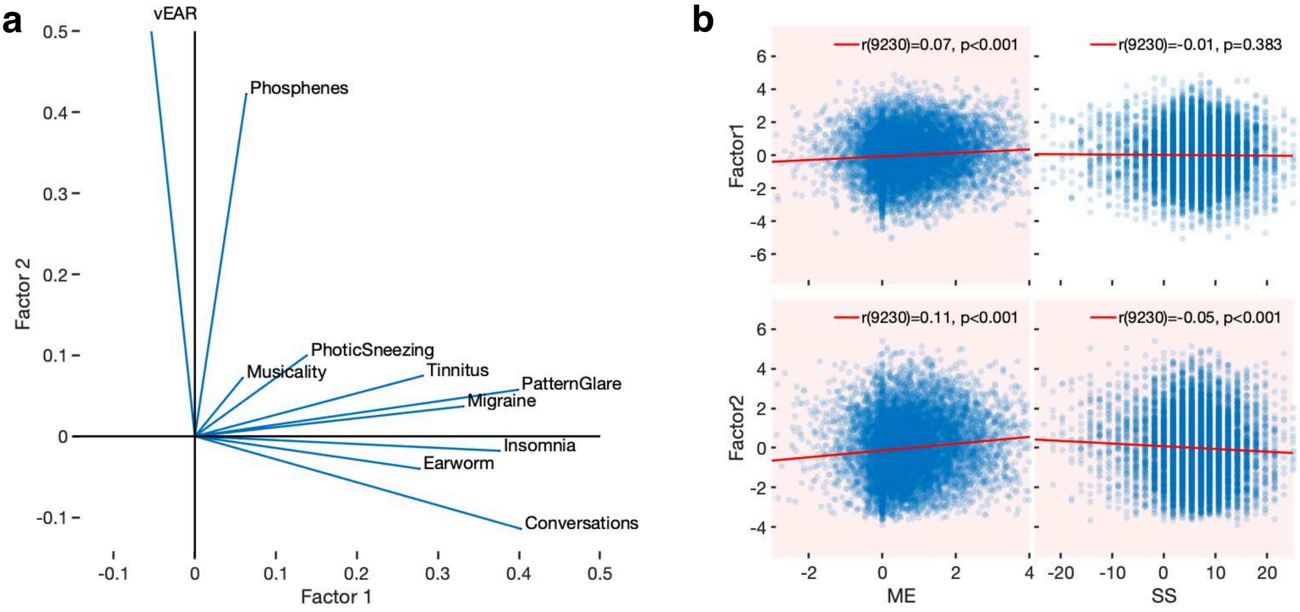


Figure 6 Results of factor analysis of traits: (a) biplot of the loadings for each trait on two factors with Promax rotation; (b) Scatterplots of ME-sensitivity (left column) and surround suppression (right column) against the factor scores for Factor 1 (top row) and for Factor 2 (bottom row). Legends show Pearson correlation statistics; red shading indicates Bonferroni-corrected statistical significance.

Discussion

This study has broadly achieved its three initial goals: (1) an objective measure of vEAR has been validated, which allows us to identify individuals whose ratings of the auditory vividness of silent abstract movies are reliably predicted by the low-level motion energy contained by the videos; (2) support has been obtaind for the hypotheses that vEAR depends on increased cortical disinhibition, showing consistent associations between video ratings and the reduction of visual surround suppression in an independent psychophysical test; (3) a detailed pattern of associations has been found between a broad set of perceptual traits, and each of our two objective measures of motion energy sensitivity and surround suppression, pointing to two factors that may independently govern the perceived intensity of vEAR via distinct mechanisms. These results reveal the lawful psychophysical relationship between objective visual motion and subjective auditory sensation, elucidating systemic neurophysiological variables that might account for vEAR along with a diverse variety of other associated perceptual traits.

The first finding is that average ratings of each video can be reliably predicted by the output of a popular model in which visual motion in a stimulus is detected by filters which possess spatiotemporal filtering characteristics resembling cells found in early visual cortex (Adelson & Bergen, 1985). This goes further than our previous study (Fassnidge & Freeman, 2018) which used live-action videos, where ratings appeared to be heavily dominated by the availability of learned associations between the depicted visual events and what they are expected to sound like. The present results imply that vEAR need not depend on prior processing of high-level semantic features as do some other forms of synaesthesia such as grapheme-colour (Mattingley et al., 2001; Myles et al., 2003; Smilek et al., 2001), because auditory sensations can be evoked by raw abstract motion which has no meaningful association with any specific sounds. We can now understand vEAR in terms of normal low-level motion sensitive mechanisms, presumably originating in early visual cortex, which in some individuals may feed into auditory and/or multisensory areas via routes that do not necessarily involve semantic analysis.

While the motion energy model predicted video ratings on average, individuals differed widely in terms of the extent to which motion energy influenced their ratings of auditory vividness. Using regression analyses, the slope of a function could be quantified for each individual relating the ratings of each video to their motion energy. This individual measure of motion energy sensitivity provides an objective diagnostic for vEAR, which is unbiased by semantic associations, and which is underpinned by objective measurement of stimulus characteristics. This measure was distributed smoothly, showing no dichotomous split between individuals who have high versus low motion energy sensitivity (**Figure 2a-b**). ME-sensitivity can therefore occur normally and frequently in a randomly sampled population, supporting our previous estimates of relatively high prevalence of vEAR self-reports (Fassnidge et al., 2019, 2017; Fassnidge & Freeman, 2018), and contrasting with other canonical synaesthesias which are typically more rare (Johnson, Allison, & Baron-Cohen, 2013; Simner et al., 2006). Furthermore, ME-sensitivity reliably predicted answers to our question about previous awareness and also self-ratings of vEAR characteristics, further reinforcing the link between our objective measure of ME-sensitivity and the subjective experience of vEAR.

As we have argued before (Fassnidge et al., 2019, 2017; Fassnidge & Freeman, 2018), the apparent normality and prevalence of vEAR, and also its broad pattern of trait associations, suggests that this form of synaesthesia may depend on systemic variables affecting cortical excitability or disinhibition (Grossenbacher & Lovelace, 2001; Lalwani & Brang, 2019), rather than depending exclusively on rare and specific patterns of anatomical cross-connectivity (Bargary & Mitchell, 2008; Baron-Cohen, 1996; Hubbard & Ramachandran, 2005; Tomson et al., 2011). Here we tested the role of inhibition using a psychophysical measure of surround suppression (Chubb et al., 1989; Xing & Heeger, 2001), finding that participants who made overall higher video ratings tended to show weaker surround suppression. Weaker surround suppression may result from a reduction of inhibitory gain control in early visual cortex (Heeger, 1992). Thus, the present association with vEAR suggests that auditory representations, and/or the connections between visual, auditory and multisensory areas may be disinhibited along with visual representations (Lalwani & Brang, 2019; Neufeld et al., 2012). A previous attempt to link grapheme-colour synaesthesia to reduced surround suppression was inconclusive (Terhune et al., 2014), however that form of synaesthesia might have weaker dependence on early visual representations, relying more on fusiform and parietal areas (Terhune et al., 2014; van Leeuwen, den Ouden, & Hagoort, 2011). The present research is the first, to our knowledge, to demonstrate that visual surround suppression can provide an independent behavioural marker for the reduction of sensory inhibition in at least one form of synaesthesia. It is interesting that surround suppression also tends to be significantly weaker in schizophrenia (Dakin et al., 2005), which may be associated with a deficit in the inhibitory neurotransmitter GABA (Keverne, 1999; Tibber et al., 2013; Yoon et al., 2009). It is tempting to speculate that the reduction of cortical inhibition (perhaps via GABA reduction) may be one common factor underlying the phenomenon of vEAR and the experience of auditory hallucinations which are a diagnostic symptom of schizophrenia.

Two aspect of our results may be particularly informative about the possible mechanisms underlying vEAR. Firstly, our regression analyses showed that that surround suppression and ME-sensitivity each predict video ratings independently, with no evidence of any interaction. Secondly, video ratings were associated to each factor in distinct ways mathematically: ME-sensitivity is a multiplicative factor, which determines the gradient of the slope relating ratings to the physical motion energy of the videos; in contrast, SS is a purely additive factor, associated with the overall magnitude of rating across all videos regardless of their motion energy, and not interacting with the effect of ME-sensitivity. We can attempt to explain these features of the results in the context of a minimalistic conceptual model, where motion energy signals are first transduced in visual cortex with varying gain and thresholds, then relayed to auditory cortex via interconnections of variable weights, before finally triggering a response in auditory cortex, again with varying gain and thresholds.

In this model, ME-sensitivity might relate to factors which could multiplicatively gate the transfer of ME signals between visual and auditory areas by weighting the mutual interconnections; alternatively, ME-sensitivity might modulate the output gain within these areas, thus amplifying or suppressing their response to incoming motion-energy signals. The independent effects of ME-sensitivity relative to SS suggest that ME-sensitivity is not directly related to expression of GABA, at least in the early visual system, given the putative role of GABA in surround suppression (Tibber et al., 2013; Yoon et al., 2009), for otherwise ME-sensitivity and SS effects should be found to interact. An alternative mechanism underlying ME-sensitivity might be an overexpression of the neurotransmitter glutamate, which appears to be elevated in the visual cortex of grapheme-colour synaesthetes and those susceptible to electromagnetically-induced visual phosphenes (Terhune et al., 2015). Independently, reduced SS in vEAR individuals might relate to a reduction of GABA-mediated inhibition that increases input gain, or lowers the threshold in auditory areas for responding to ME signals of visual origin, thus generally increasing the effective intensity of

ME signals. Such a shift might then appear to affect video ratings of ME signals of different intensities additively. An alternative possibility is that disinhibition in auditory cortex might increase spontaneous auditory activity, perhaps providing greater raw material for hallucination-like auditory sensations (Kumar et al., 2014; Northoff & Qin, 2011), or allowing neural activity to cross a threshold into awareness via stochastic resonance (Lalwani & Brang, 2019). A similar mechanism might also account for the converse phenomenon of auditory-induced phosphenes, where visual areas respond more sensitively to auditory signals (Bolognini, Senna, Maravita, Pascual-Leone, & Merabet, 2010).

As well as showing independent influences on video ratings, SS and ME-sensitivity each appeared to correlate with different clusters of traits. The use of reverse-coding of survey questions in this study helps to argue against the possibility that such correlations were caused by acquiescence bias, where some participants might have tended to respond generally more positively than others. ME-sensitivity was associated with most of the traits we probed (in decreasing order of the magnitude of the slope relating trait to ME-sensitivity: vEAR self-rating, auditory-evoked phosphenes, earworms, migraine, pattern glare, tinnitus, musicality, and insomnia), with the exception of two (photic sneezing and speech-in-noise comprehension). We can therefore consider the possibility that multiplicative gain-control is a common factor that explains not only ME-sensitivity but this broader set of associated traits, by modulating cortical excitability and sensory sensitivity. This would be consistent with previous research identifying a role for cortical excitability in migraine aura (Palmer et al., 2000; Tibber et al., 2014), pattern glare (Wilkins et al., 1984), insomnia (Van Der Werf et al., 2010), tinnitus (Kaltenbach, 2011), and visual phosphenes (Jacobs et al., 1981; Lessell & Cohen, 1979; Nair & Brang, 2019), as well as EEG evidence that vEAR is associated with stronger early responses to visual stimulation (Rothen et al., 2017). Our questions about musicality and earworms were also selected on the assumption that greater sensitivity to musical patterns, and experiences of involuntary musical imagery might result from greater excitability and spontaneous activity of areas involved in processing the patterns of sound (Griffiths, 2000; Kumar et al., 2014).

In contrast with ME-sensitivity, the traits associated with surround suppression appear much more selective, including only vEAR self-ratings, auditory-evoked phosphenes, and musicality. These traits all seem to involve crossmodal interactions, either from vision to audition (vEAR), and from audition to vision (phosphenes), while musicianship may generally require a high degree of audiovisual integration (Tsay, 2013). It is particularly interesting that in our previous brain-stimulation study (Fassnidge et al., 2019), our sample of highly-trained musicians showed evidence of reduced inhibition (or increased cooperation) between visual and auditory cortices, as well as overall a greater prevalence of vEAR experiences. In sum, this clustering of traits is consistent with our previous proposal that vEAR is characterised by a reduction of mutual inhibition between and within auditory and visual modalities, which may weaken surround suppression, but also reinforces richer crossmodal interactions and the propensity for experiencing synaesthesia-like sensory phenomena.

In conclusion, this study has validated an objective measure of vEAR, which relates the physical characteristics of visual stimuli to subjective reports of auditory sensations, underpinned by a model of motion energy transduction in early visual cortex. Furthermore, this research has contributed to understanding of the possible mechanisms underlying this phenomenon as well as other traits and sensory phenomena with which it is associated. In particular it has been proposed that vEAR may be characterised by two independent mechanisms: firstly higher sensory gain, which functions to multiplicatively amplify motion energy signals and/or their transfer across modalities; secondly reduced inhibition between modalities which may function to lower the threshold for experiencing visually-evoked auditory sensations, as well as auditory-evoked phosphenes. Finally, each of these routes to vEAR are associated with different clusters of traits including musicality, tinnitus, pattern glare, and migraine, pointing to a possible common basis for understanding a wide range of different perceptual phenomena and comorbidities. Most intriguingly, the reduction of surround suppression that we observed in vEAR may have parallels with similar reduction in suppression in schizophrenia, pointing to a possible common disinhibitory framework for understanding the neural basis for spontaneous auditory sensations in both healthy and pathological populations.

Acknowledgements

This research was supported by British Academy/Leverhulme grant SG151380.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America. A, Optics and Image Science*, *2*(2), 284–299. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3973762

Bargary, G., & Mitchell, K. J. (2008). Synaesthesia and cortical connectivity. *Trends in Neurosciences*. https://doi.org/10.1016/j.tins.2008.03.007

Baron-Cohen, S. (1996). Is there a normal phase of synaesthesia in development. *Psyche*, *2*(27), 223–228.

Bolognini, N., Senna, I., Maravita, A., Pascual-Leone, A., & Merabet, L. B. (2010). Auditory enhancement of visual phosphene perception: the effect of temporal and spatial factors and of stimulus intensity. *Neuroscience Letters*, *477*(3), 109–114. https://doi.org/10.1016/j.neulet.2010.04.044

Bor, D., Rothen, N., Schwartzman, D. J., Clayton, S., & Seth, A. K. (2014). Adults can be trained to acquire synesthetic experiences. *Scientific Reports*, *4*, 7089. https://doi.org/10.1038/srep07089

Chubb, C., Sperling, G., & Solomon, J. A. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Sciences*, *86*(23), 9631–9635. https://doi.org/10.1073/pnas.86.23.9631

Dakin, S., Carlin, P., & Hemsley, D. (2005). Weak suppression of visual context in chronic schizophrenia [2]. *Current Biology*, *15*(20), 822–824. https://doi.org/10.1016/j.cub.2005.10.015

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., & Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *Journal of Neuroscience Methods*, *159*(1), 139–145. https://doi.org/10.1016/j.jneumeth.2006.07.012

Fassnidge, C. J., Ball, D., Kazaz, Z., Knudsen, S., Spicer, A., Tipple, A., & Freeman, E. (2019). Hearing through your eyes: Neural basis of audiovisual cross-activation, revealed by transcranial alternating current stimulation. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn_a_01395

Fassnidge, C. J., Cecconi Marcotti, C., & Freeman, E. D. (2017). A deafening flash ! Visual interference of auditory signal detection. *Consciousness and Cognition*, *49*, 15–24. https://doi.org/10.1016/j.concog.2016.12.009

Fassnidge, C. J., & Freeman, E. D. (2018). Sounds from seeing silent motion: Who hears them, and what looks loudest? *Cortex*, *103*, 130–141. https://doi.org/10.1016/j.cortex.2018.02.019

Firth, N. (2018). A fifth of people hear sounds when watching silent GIFs. Do you? *New Scientist*. https://doi.org/https://www.newscientist.com/article/2164086

Griffiths, T. D. (2000). Musical hallucinosis in acquired deafness: Phenomenology and brain substrate. *Brain*, *123*(10), 2065–2076. https://doi.org/10.1093/brain/123.10.2065

Grossenbacher, P. G., & Lovelace, C. T. (2001). Mechanisms of synesthesia: cognitive and physiological constraints. *Trends in Cognitive Sciences*, *5*(1), 36–41. https://doi.org/10.1016/S1364-6613(00)01571-0

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*(2), 181–197. https://doi.org/10.1017/S0952523800009640

Hubbard, E. M., & Ramachandran, V. S. (2005). Neurocognitive mechanisms of synesthesia. *Neuron*, *48*(3), 509–520. https://doi.org/10.1016/j.neuron.2005.10.012

Ipser, A., Agolli, V., Bajraktari, A., Al-Alawi, F., Djaafara, N., & Freeman, E. D. (2017). Sight and sound persistently out of synch: stable individual differences in audiovisual synchronisation revealed by implicit measures of lip-voice integration. *Scientific Reports*, *7*(October 2016), 46413. https://doi.org/10.1038/srep46413

Jacobs, L., Karpik, A., Bozian, D., & Gøthgen, S. (1981). Auditory-visual synesthesia sound-induced photisms. *Archives of Neurology*, *38*(4), 211–216.

Johnson, D., Allison, C., & Baron-Cohen, S. (2013). The prevalence of synesthesia. *Oxford Handbook of Synesthesia*, *1*.

Kaltenbach, J. A. (2011). Tinnitus: Models and mechanisms. *Hearing Research*, *276*(1–2), 52–60. https://doi.org/10.1016/j.heares.2010.12.003

Keverne, E. B. (1999). GABA-ergic neurons and the neurobiology of schizophrenia and other psychoses. *Brain Research Bulletin*, *48*(5), 467–473. https://doi.org/10.1016/S0361-9230(99)00025-8

Kumar, S., Sedley, W., Barnes, G. R., Teki, S., Friston, K. J., & Griffiths, T. D. (2014). A brain basis for musical hallucinations. *Cortex*, *52*, 86–97. https://doi.org/10.1016/j.cortex.2013.12.002

Lalwani, P., & Brang, D. (2019). Stochastic resonance model of synaesthesia. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *374*(1787), 20190029. https://doi.org/10.1098/rstb.2019.0029

Langer, N., Beeli, G., & Jäncke, L. (2010). When the sun prickles your nose: an EEG study identifying neural bases of photic sneezing. *PloS One*, *5*(2), e9208. https://doi.org/10.1371/journal.pone.0009208

Lessell, S., & Cohen, M. M. (1979). Phosphenes induced by sound. *Neurology*, *29*(11), 1524.

Mattingley, J. B., Rich, A. N., Yelland, G., & Bradshaw, J. L. (2001). Unconscious priming eliminates automatic binding of colour and alphanumeric form in synaesthesia. *Nature*, *410*, 580–582. https://doi.org/10.1038/35069062

Myles, K. M., Dixon, M. J., Smilek, D., & Merikle, P. M. (2003). Seeing double: The role of meaning in alphanumeric-colour synaesthesia. *Brain and Cognition*, *53*(2), 342–345. https://doi.org/10.1016/S0278-2626(03)00139-8

Nair, A., & Brang, D. (2019). Inducing synesthesia in non-synesthetes: Short-term visual deprivation facilitates auditory-evoked visual percepts. *Consciousness and Cognition*, *70*(February), 70–79. https://doi.org/10.1016/j.concog.2019.02.006

Neufeld, J., Sinke, C., Zedler, M., Dillo, W., Emrich, H. M., Bleich, S., & Szycik, G. R. (2012). Disinhibited feedback as a cause of synesthesia: evidence from a functional connectivity study on auditory-visual synesthetes. *Neuropsychologia*, *50*(7), 1471–1477. https://doi.org/10.1016/j.neuropsychologia.2012.02.032

Northoff, G., & Qin, P. (2011). How can the brain's resting state activity generate hallucinations? A "resting state hypothesis" of auditory verbal hallucinations. *Schizophrenia Research*, *127*(1–3), 202–214. https://doi.org/10.1016/j.schres.2010.11.009

Palmer, J. E., Chronicle, E. P., Rolan, P., & Mulleners, W. M. (2000). Cortical hyperexcitability is cortical under-inhibition: Evidence from a novel functional test of migraine patients. *Cephalalgia*, *20*(6), 525–532. https://doi.org/10.1046/j.1468-2982.2000.00075.x

Rothen, N., Bartl, G. J., Franklin, A., & Ward, J. (2017). Electrophysiological correlates and psychoacoustic characteristics of hearing-motion synaesthesia. *Neuropsychologia*, *106*, 280–288. https://doi.org/10.1016/j.neuropsychologia.2017.08.031

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., … Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception*, *35*(8), 1024–1033. Retrieved from http://people.brunel.ac.uk/~hsstnns/reprints/Simner_at_al_2006_Prevalence.pdf

Smilek, D., Dixon, M. J., Cudahy, C., & Merikle, P. M. (2001). Synaesthetic photisms influence visual perception. *Journal of Cognitive Neuroscience*, *13*(7), 930–936. https://doi.org/10.1162/089892901753165845

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Terhune, D. B., Murray, E., Near, J., Stagg, C. J., Cowey, A., & Kadosh, R. C. (2015). Phosphene perception relates to visual cortex glutamate levels and covaries with atypical visuospatial awareness. *Cerebral Cortex*, *25*(11), 4341–4350. https://doi.org/10.1093/cercor/bhv015

Terhune, D. B., Song, S. M., Duta, M. D., & Cohen Kadosh, R. (2014). Probing the neurochemical basis of synaesthesia using psychophysics. *Frontiers in Human Neuroscience*, *8*(February), 89. https://doi.org/10.3389/fnhum.2014.00089

Tibber, M. S., Anderson, E. J., Bobin, T., Antonova, E., Seabright, A., Wright, B., … Dakin, S. C. (2013). Visual Surround Suppression in Schizophrenia. *Frontiers in Psychology*, *4*(February), 1–13. https://doi.org/10.3389/fpsyg.2013.00088

Tibber, M. S., Kelly, M. G., Jansari, A., Dakin, S. C., & Shepherd, A. J. (2014). An inability to exclude visual noise in migraine. *Investigative Ophthalmology and Visual Science*, *55*(4), 2539–2546. https://doi.org/10.1167/iovs.14-13877

Tomson, S. N., Avidan, N., Lee, K., Sarma, A. K., Tushe, R., Milewicz, D. M., … Eagleman, D. M. (2011). The genetics of colored sequence synesthesia: suggestive evidence of linkage to 16q and genetic heterogeneity for the condition. *Behav Brain Res*, *223*(1), 48–52. https://doi.org/10.1016/j.bbr.2011.03.071

Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), 14580–14585. https://doi.org/10.1073/pnas.1221454110

Van Der Werf, Y. D., Altena, E., Van Dijk, K. D., Strijers, R. L. M., De Rijke, W., Stam, C. J., & Van Someren, E. J. W. (2010). Is disturbed intracortical excitability a stable trait of chronic insomnia? A study using transcranial magnetic stimulation before and after multimodal sleep therapy. *Biological Psychiatry*, *68*(10), 950–955. https://doi.org/10.1016/j.biopsych.2010.06.028

van Leeuwen, T. M., den Ouden, H. E. M., & Hagoort, P. (2011). Effective Connectivity Determines the Nature of Subjective Experience in Grapheme-Color Synesthesia. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.0569-11.2011

Wilkins, A., Nimmo-smith, I., Tait, A., Mcmanus, C., Sala, S. Della, Tilley, A., … Scott, S. (1984). A neurological basis for visual discomfort. *Brain*, *107*(4), 989–1017. https://doi.org/10.1093/brain/107.4.989

Witthoft, N., Winawer, J., & Eagleman, D. M. (2015). Prevalence of Learned Grapheme-Color Pairings in a Large Online Sample of Synesthetes. *PLoS One*, *10*(3), e0118996. https://doi.org/10.1371/journal.pone.0118996

Xing, J., & Heeger, D. J. (2001). Measurement and modeling of center-surround suppression and enhancement. *Vision Research*, *41*(5), 571–583. https://doi.org/10.1016/S0042-6989(00)00270-4

Yoon, J. H., Rokem, A. S., Silver, M. A., Minzenberg, M. J., Ursu, S., Ragland, J. D., & Carter, C. S. (2009). Diminished Orientation-Specific Surround Suppression of Visual Processing in Schizophrenia. *Schizophrenia Bulletin*, *35*(6), 1078–1084. https://doi.org/10.1093/schbul/sbp064

Table 1 Alternative wordings of trait questions in vEAR survey

TRAIT	**POSITIVE STATEMENT**	**NEGATIVE STATEMENT**
MUSICALITY	*I have a good ear for music*	*I do not have a good ear for music* [1]
TINNITUS	*I suffer from tinnitus*	*I do not suffer from tinnitus* [1]
PHOSPHENES	*I have experienced vivid flashes evoked by sounds*	*I have never experienced flashes evoked by sounds* [1]
EARWORMS	*I very frequently experience earworms*	*I never experience earworms* [1]
VEAR	*I can watch visual movement or flashing without experiencing sounds in my head* [1]	*I cannot watch visual motion or flashing without experiencing vivid sounds in my head*
MIGRAINE	*I get migraines very frequently*	*I never experience migraines*
AURA[2]	*When I get migraines, I usually experience visual disturbances*	*When I get migraines, I never experience visual disturbances*
PATTERN GLARE	*I can look at certain high contrast repetitive patterns without experiencing discomfort or visual distortions* [1]	*I cannot look at certain high contrast repetitive patterns without experiencing discomfort or visual distortions.*
PHOTIC SNEEZING[1]	*I can walk in bright sunshine without experiencing the urge to sneeze* [1]	*I cannot walk in bright sunshine without experiencing the urge to sneeze*
SPEECH IN NOISE	*I am bad at following conversations when there is background noise*	*I am good at following conversations when there is background noise* [1]
INSOMNIA	*I find it hard to sleep at night*	*I find it easy to sleep at night* [1]

[1] *reverse-coded*

[2] *response options included 'not applicable'*

Table 2 Pearson's correlations of video ratings with trait ratings

	Positively-worded	**Negatively-worded**
Musicality	r(4566)= 0.20, p<0.001	r(4662)= 0.09, p<0.001
Tinnitus	r(4656)= 0.14, p<0.001	r(4572)= 0.10, p<0.001
Phosphenes	r(4597)= 0.29, p<0.001	r(4631)= 0.21, p<0.001
Earworm	r(4594)= 0.23, p<0.001	r(4634)= 0.04, p<0.003[1]
vEAR	r(4623)= 0.46, p<0.001	r(4605)= 0.32, p<0.001
Migraine	r(4632)= 0.13, p<0.001	r(4596)= 0.12, p<0.001
aura	r(2721)= 0.10, p<0.001	r(2574)= 0.08, p<0.001
PatternGlare	r(4562)= 0.18, p<0.001	r(4666)= 0.19, p<0.001
PhoticSneezing	r(4633)= 0.10, p<0.001	r(4595)= 0.11, p<0.001
Conversations	r(4710)= 0.11, p<0.001	r(4518)= 0.05, p<0.001
Insomnia	r(4720)= 0.13, p<0.001	r(4508)= 0.15, p<0.001

[1] *Only non-significant result after Bonferonni correction (p=.07).*

Table 3 Regressions of each trait against Motion Energy sensitivity (ME) and Surround Suppression (SS)

Trait	ME_slope[1]	ME_t	ME_p	ME_p_{corr}	SS_slope[1]	SS_t	SS_p	SS_p_{corr}	R^2 %
Musicality	0.054	4.184	<0.001	<0.001	-0.008	-3.730	<0.001	0.002	0.3
Tinnitus	0.059	3.823	<0.001	0.001	-0.004	-1.526	0.127	0.743	0.2
Phosphenes	0.082	5.322	<0.001	<0.001	-0.008	-3.161	0.002	0.016	0.4
Earworm	0.077	5.674	<0.001	<0.001	0.001	0.565	0.572	1.000	0.4
vEAR	0.139	10.177	<0.001	<0.001	-0.008	-3.625	<0.001	0.003	0.3
Migraine	0.066	4.160	<0.001	<0.001	0.001	0.226	0.821	1.000	0.2
PatternGlare	0.060	4.117	<0.001	<0.001	-0.002	-0.947	0.344	0.985	0.2
PhoticSneezing	0.040	2.460	0.014	0.131	-0.005	-2.065	0.039	0.328	0.1
Conversations	0.032	2.202	0.028	0.245	0.006	2.398	0.016	0.153	0.1
Insomnia	0.051	3.361	0.001	0.008	-0.006	-2.433	0.015	0.140	0.2

[1] *Coefficients from the regression equation:* $Rating_{trait} = ME + SS + \varepsilon$

Table 4 Results of factor analysis of traits

Trait	*Factor1*	*Factor2*
Musicality	0.06	0.07
Tinnitus	0.28	0.07
Phosphenes	0.06	0.42
Earworm	0.28	-0.04
vEAR	-0.06	0.51
Migraine	0.33	0.04
PatternGlare	0.40	0.06
PhoticSneezing	0.14	0.10
Conversations	0.40	-0.11
Insomnia	0.38	-0.02