Citation: Fassnidge, C. (2018). The Visually-Evoked Auditory Response. (Unpublished Doctoral thesis, City, Universtiy of London)

The Visually-Evoked Auditory Response

Christopher James Fassnidge

A THESIS SUBMITTED TO THE
DEPARTMENT OF PSYCHOLOGY
CITY, UNIVERSITY OF LONDON
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

March 2018

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank all of the people who supported me throughout my research. First my family, for all the love and support and especially to Harrison and Olivia for making me laugh like no one else. Thank you to my friends, and all the fellow PhD students here at City, especially Ali, Iro, and Maciej who gave me endless moral, practical and emotional support. I would also like to thank Kielan for his many sage words over the years. I must also acknowledge the hard work of the various students and volunteers who aided me in data collection, Adam, Andreana, Anthony, Claudia, Danny, Sylwia, Synøve, and Zainab. Thank you also to Kinga and to Cara for keeping me sane in these final months of write-up. Finally, my supervisor Dr Elliot Freeman for all his patience, support and expertise. I really couldn’t have done it without you.

Declaration

I grant powers of discretion to the University Librarian to copy this thesis in part or in whole without further reference to me. This covers only single copies for study purposes. The contents of this thesis are subject to normal conditions of acknowledgement. The content of this thesis is my own work.

Abstract

In synaesthesia a sensation in one modality triggers a consciously perceived sensation in another sensory modality or cognitive domain. In this thesis we investigate auditory sensation that are induced by dynamic visual stimuli, akin to hearing-motion synaesthesia (Saenz and Koch, 2008). We term this the Visually-Evoked Auditory Response (vEAR). We first establish the prevalence of vEAR in a random sample, with questionnaire responses indicating a higher prevalence (as many as 1 in 5) than canonical synaesthesias. We report that those who experience vEAR showed better performance compared to controls when discriminating between 'Morse-code' style rhythmic sequences in the visual domain, as did Saenz and Koch (2008). We also demonstrate that vEAR is perceptually real enough to interfere with hearing real world sounds. We then demonstrate that in control subjects Transcranial Alternating Current Stimulation (TACS), when applied over the temporal versus the occipital lobes, impairs auditory versus visual sequence discrimination respectively. However, temporal TACS improved visual and occipital TACS improved auditory sequence discrimination performance. This suggests the presence of normally-occurring mutual alpha-mediated competitive inhibition of the two cortices. This TACS effect was not seen in individuals with vEAR, indicating that their auditory and visual cortices are able to cooperate to perform the task despite disruption from TACS. Finally, we investigate the types of visual stimuli that best evoke vEAR, and the types of people who tend to experience it. We conducted a large online survey in which respondents rated the amount of vEAR evoked by a series of silent videos depicting types of motion. The predictiveness of a real-world sound was identified as a major contributor to ratings in all respondents, while motion energy (raw changes in light over space and time) specifically influenced ratings in those who experience vEAR. We also report demographic and trait questions relating to auditory perception that predict higher ratings, including the frequency one experiences music imagery in their head, or whether they have tinnitus or types of synaesthesia. We conclude that vEAR results from both high and low-level connectivity between the visual and auditory cortices and an atypical inhibition of these connections.

Abbreviations

A: Auditory

AV: Audiovisual

SD: Standard Deviation

SPL: Sound Pressure Level

TACS: Transcranial Alternation Current Stimulation

V: Visual

vEAR: Visually-evoked auditory response

Introduction

Chapter 1:

This chapter presents a brief outline of what synaesthesia is and outlines some of its variations. We present some contemporary debates in the literature including the role of learning in the specific sensory pairings experienced by synaesthetes, and the extent to which synaesthesia is expressed as a spectrum throughout the population as opposed to as a discrete continuous phenomenon. We highlight a relatively unknown form of synaesthesia in which flashes of light or moving visual objects are perceived with a concurrent sound and provide a plausible explanation for why this phenomenon, which we term the visually-evoked auditory response (vEAR), may be more prevalent in the population than other types of synaesthesia. Finally, we outline the aims of this thesis, which include i) establishing an estimate of the prevalence of vEAR, ii) exploring how perceptually real the visually-evoked sounds are to the individual, iii) investigating the neurophysiological mechanisms that may underlie the experience of vEAR and iv) examining what properties of a visual stimulus best evoke vEAR.

1.1 Introduction and Thesis Rationale

This thesis explores a little-known phenomenon that may potentially be type of synaesthesia, in which individuals perceive illusory auditory sensations when viewing dynamic visual stimuli such as moving objects or flashing lights. Prior to our research there was has been only one empirical report of this phenomenon in the literature (Saenz & Koch, 2008), and an incidental reference to it in another (Guttman, Gilroy & Blake, 2005). Despite this scarcity of research, anecdotal reports of this phenomenon can be found in the online and offline community of synaesthetes. Having spoken informally to some of these individuals both through online forums and at conferences, they tend to describe these visually-evoked auditory sensations as sounding like white noise or ‘whooshing’ sounds that accompany various types of visual motion, but some also report hearing spoken words accompanying lip movements when viewing muted televisions, or the footsteps of an individual walking in the distance out of earshot. Here we first use a combination of objective psychophysical tasks and subjective questioning in order to estimate how common this phenomenon is, as there has been no reported population prevalence of this particular variation of the condition to date. We then explore how perceptually real these visually-evoked auditory sensations are relative to externally originating sounds, and whether the two sets of auditory signals can interfere with one another. Next, we explore individual differences between those who experience these visually-evoked auditory sensations relative to those who do not, using transcranial electric stimulation to infer neurophysiological differences between the two groups, that could plausibly provide a cortical aetiology for the illusory auditory sensations. Finally, we explore the types of stimuli that best evoke these auditory sensations, with reference to both the physical and learned associative properties of the stimulus. We conclude by outlining what we believe is a plausible model to explain this unusual phenomenon, encompassing all our experimental findings with support from existing literature. The implications of our findings are discussed in the context of several contemporary discussions in the synaesthesia literature and questions for future research are addressed, as well as whether the visually-evoked auditory sensations described here strictly meet the criteria to be classified as a sub-type of synaesthesia, or whether they are likely to be a related yet distinct phenomenon.

In the following chapter we shall introduce the phenomenon of synaesthesia and briefly outline some of the sub-types that characterise the condition, with an emphasis on a little-known sub-type in which dynamic visual stimuli induce auditory sensations in the perceiver. We shall then introduce some contentious debates that exist in the field of synaesthesia research with reference to how our research can contribute to these on-going debates. Finally, we shall introduce some of the specific questions that this thesis will set out to address.

1.2 What is Synaesthesia?

Synaesthesia, from the Greek syn, (meaning together, a joining or union) and aesthesis (of or from the senses), is a perceptual anomaly whereby stimulation in one sensory modality, typically referred to as the inducer, leads to consistent and involuntary sensations in another sensory modality (or a separate property of the inducer modality), known as the concurrent (Grossenbacher, 1997). Although the defining criteria for synaesthesia vary somewhat in the literature (see Simner, 2012) there is a core set of components that are typically used to define the phenomenon. For example, the experience occurs automatically (Ward and Mattingley, 2007), requiring no conscious effort to perceive the concurrent and lacking an ability to 'turn it off'. The associations between inducer and concurrent are highly specific and consistent overtime (Baron-Cohen, Wyke, & Binnie, 1987). Those who experience synaesthesia generally report that the percepts they experience are perceived in addition to, rather than in place of, their veridical perception of the world (Simner, 2012; Ward and Simner, 2003), meaning for example a visual concurrent would not blind them to the true visual scene but would be perceived as an overlay. Despite this, synaesthetes do not tend to confuse their synaesthetic percepts with real world objects or sensations (Rich & Mattingley, 2002). Synaesthesia is generally developmental in nature, although there are reported cases of acquired synaesthesia following traumatic brain injury (e.g. Rao et al., 2007), training (Bor et al., 2014), or transiently following ingestion of certain psychoactive drugs such as psilocybin, LSD and mescaline (Brogard, 2013).

There are scientific reports of synaesthesia dating back more than two centuries (e.g. Galton, 1880, Calkins, 1895 who coined the term, and previously Sachs, 1812), although later advances in technology and experimental methodologies meant that the latter half of the twentieth century saw a surge of research activity in this area (see for example Baron-Cohen et al., 1987; Cytowic and Wood, 1982; Marks, 1975). Today synaesthesia continues to attract a considerable amount of research interest, not only because it provides a fascinating insight into a group of individuals who perceive the world in a manner apparently quite alien to the general population, but it also provides a valuable contribution to our understanding of how the brain integrates information from the different senses in typical perception. The field of contemporary synaesthesia research is somewhat diverse, in part due to the remarkable heterogeneity of the condition which gives rise to a diverse range of inducer-concurrent pairings and substantial variability in the perceptual vividness of the concurrent sensation (for a review see Ward, 2013). As a result synaesthesia research has seen a rapid and consistent increase in publication output (see Figure 1-0-1).

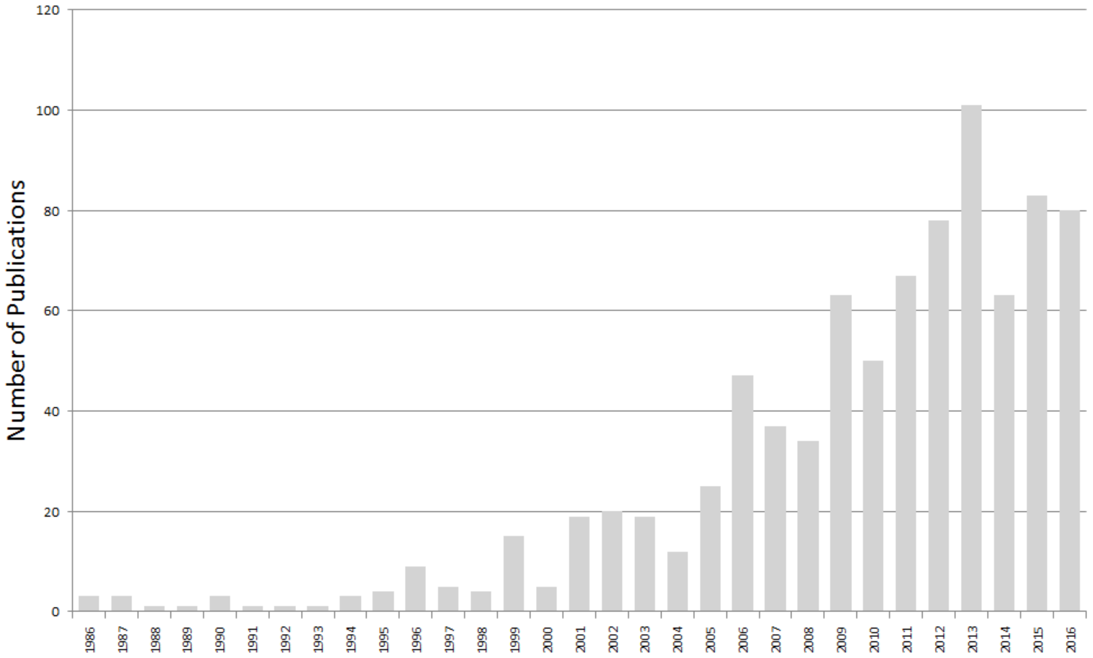


Figure 1-0-1: Published synaesthesia research per year 1996-2016.

(Articles containing the keyword “synaesthesia OR synaesthesia ”). Source: Web of Science

1.3 Types of Synaesthesia

There are many different sub-types of synaesthesia that have been reported, with at least 60 (Day, 2005, 2013) and possibly as many as 150 manifestations of the condition (Cytowic & Eagleman, 2009), and these tend to be named following the convention of combining the inducer and the concurrent separated by a hyphen (although this convention is not always followed, as in hearing-motion). For example, one of the more commonly reported variants of the condition is grapheme-colour synaesthesia (Rich, Bradshaw & Mattingley, 2005; Simner et al., 2005), in which individuals when reading text will always see particular letters or numbers as occurring in particular colours (for example the letter C might always be seen as red, or the number 9 as blue) no matter the true physical colour of the text. Other variants of the condition include tone-colour synaesthesia, (Ward, Huckstep & Tsakanikos, 2006) in which sounds, and in particular the frequency of musical tones, can trigger the sensation of colour, and lexical-gustatory synaesthesia (Ward & Simner, 2003), in which the synaesthete will experience a sensation of taste upon hearing a particular word. In the latter two examples the inducer-concurrent pairing cross from one sensory modality into another, while in grapheme-colour synaesthesia they cross different facets of visual perception, therefore it is not always the case that synaesthesia is strictly a multisensory phenomenon. In other forms of synaesthesia the inducer-concurrent pairing may blend more abstract concepts, such as days of the week, with sensory percepts such as colours (Shanon, 1982; Rich, Bradshaw, & Mattingley, 2005; Simner et al., 2006) or with numbers (Sagiv, Heer, & Robertson, 2006; Ward & Sagiv, 2007). Others still attribute non-sensory ‘personalities’ to the inducer, for example letters or numbers that are friendly, angry, or lazy (Day, 2005; Cytowic, 2002; Simner & Holenstein, 2007; Smilek et al., 2007). Whether these cases are considered as canonical synaesthesias depends on the definition of synaesthesia that one employs, and no universal standard has yet been agreed (although some contemporary debates in this area are outlined in section 1.4).

1.3.1 Audiovisual Synaesthesias

Of the many and varied types of synaesthesia reported to date several feature sensory crossover from auditory inducers to visual concurrents. The most prominent example of these is characterised by musical tones inducing visual percepts such as colours or coloured shapes (e.g. Chiou, Stelter, & Rich, 2013), a phenomenon known variously as sound-colour synaesthesia (e.g.

Ward, Huckstep, & Tsakanikos, 2006), tone-colour synaesthesia (e.g. Hänggi et al., 2008) or chromesthesia (e.g. Block, 1983).

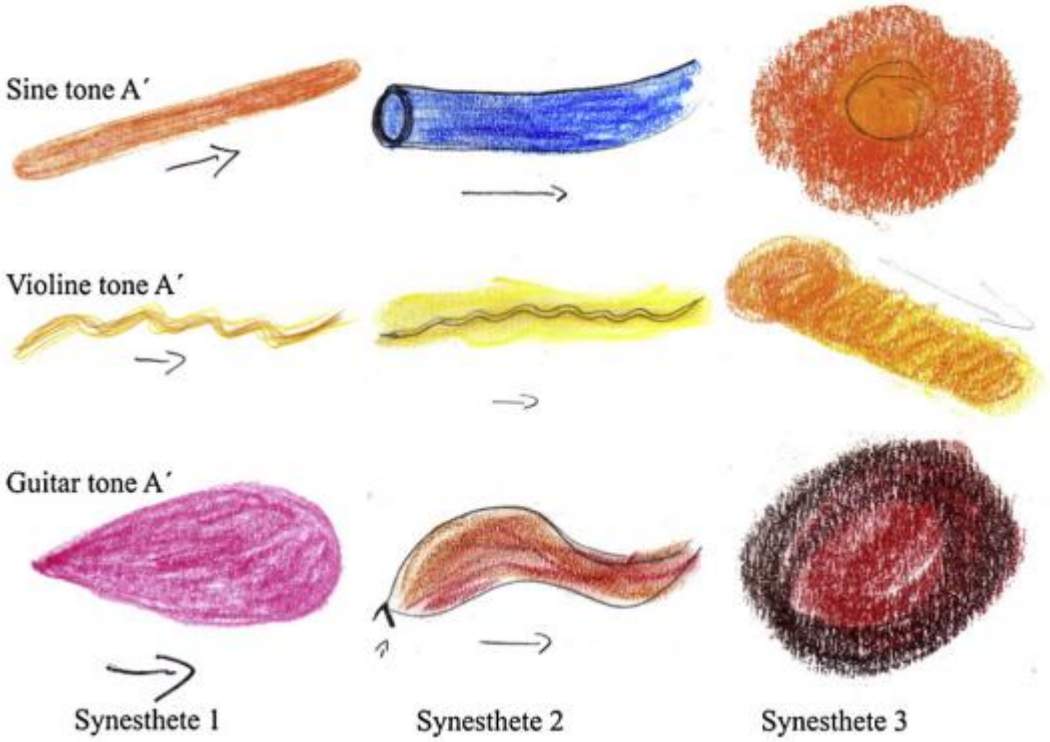


Figure 1-0-2: Acoustically induced synesthetic photisms of three individual synaesthetes.

(1 column per synaesthete). Inducers were single tones (sine, violin and guitar) in A. Photisms were perceived in three dimensions and the forms changed with the rise and fading of the tone, moving in the direction indicated by the arrows (Figure from Neufeld et al., 2012).

Those who experience this phenomenon tend to report perceiving a colour in some form, whether it is explicitly 'seen' in external space before their eyes (See Figure 1-0-2) or a more abstract 'feeling' of a colour, that is consistently triggered by a specific aspect of a sound, usually the frequency (pitch) but in other cases it can be influenced by factors such as the timbre, the type of instrument on which the sound is played, etc. In other examples the inducer is a non-musical sound, such as the human voice (Fernay, Reby, & Ward, 2012; Moos et al., 2013).

1.3.2 Hearing-Motion Synaesthesia

While such sound-to-vision synaesthesias have been widely reported there are very few reported examples of visual-to-sound synaesthesias to our knowledge, bar a few rare reports of individuals who hear distinct musical sounds associated with visual colours (Baron-Cohen, Burt, Smith-Laittan, Harrison, & Bolton, 1996; Goller, Otten, & Ward, 2009). Saenz & Koch (2008) described a previously unknown 'hearing-motion' synaesthesia in which visual events, such as flashing lights or dots moving on a screen, induce corollary auditory perceptions (following Guttman, Gilroy, & Blake, 2006 who had previously noted that individuals may be able to hear the rhythm of visual flashes). People with this condition report being able to consistently and involuntarily hear these visual stimuli, thus meeting one of the classical defining features of synaesthesia. In order to objectively test the veracity of these claimed hearing-motion abilities the authors devised a simple behavioural task on which performance would be enhanced if an individual possessed such abilities. This diagnostic test required the perceptual discrimination of paired 'Morse code' like sequences, presented either as sounds or as flashes. Participants were presented with two successive rhythmic sequence made up of eight stimuli of differing durations, either short (50 ms) or long (200 ms), and were then asked to report whether the two sequences were the same or different. On each trial sequences were either both auditory or both visual. Performance with auditory stimuli was generally high in both synaesthetes and in controls, benefitting from the greater temporal acuity of the auditory system (Glenberg et al., 1989; Guttman, Gilroy & Blake, 2005). By comparison performance in the visual condition was significantly poorer in the control group, however individuals who identified themselves as hearing-motion synaesthetes showed no significant difference in task performance between the two modalities, and significantly better visual sequence discrimination compared to controls. It was proposed that their visual sequencing performance benefits from the additional temporal information provided by their concurrent auditory sensation (Glenberg, et al., 1989; Guttman, Gilroy, & Blake, 2006).

Unlike in some forms of synaesthesia, the mapping of visual events to auditory concurrents could be considered to be somewhat generic, with participants' in Saenz and Koch's (2008) cohort describing their concurrents as whirring, tapping or beeping sensations, which are markedly less specific than, for example, instances in which particular graphemes or musical tones induce specific coloured percepts of a particular hue. Other anecdotal reports from our own lab describe the auditory sensations as being like a generic 'white noise' sound. It is then a matter of debate as to whether these people truly have synaesthesia in the canonical sense, or are instead experiencing a closely-related phenomenon. For these reasons we will instead be using the more agnostic description of the Visually-Evoked Auditory Response (vEAR) rather than the hearing-motion synaesthesia favoured by Saenz and Koch (2008). In this thesis we will explore whether this is in fact a normal (or at least a relatively common) sensory phenomenon, establishing the perceptual reality of the visually-evoked concurrent auditory sensation, and the neurophysiological mechanisms that might explain individual differences in the experience of the condition, as well as the types of visual stimuli that best evoke these auditory sensations.

1.4 Debates Surrounding Synaesthesia

The heterogeneity of synaesthesia raises a number of questions. Here we shall focus on a number of contemporary debates surrounding synaesthesia and the framework in which we understand the condition. These include the population prevalence of synaesthesia, what determines the specific pairings experienced by any given synaesthete, and the extent to which we all experience synaesthetic percepts in some sense, i.e. whether the condition is a dichotomous phenomenon that occurs only in a few individuals or whether it is continuous on some level throughout the population. Does synaesthesia globally share a common mechanistic aetiology, or does the architecture underlying each case differ somewhat? There is evidence to suggest that individuals with one form of synaesthesia have a higher probability of also experiencing a second form than would be the case in the general population (Day, 2005; Rogowska 2011), suggesting that there may be some common features across different synaesthetic sub-types. Other contemporary debates are more fundamental in nature, and might force us to re-examine the boundaries that define precisely what is, and what is not, a type of synaesthesia.

1.4.1 Rethinking Synaesthesia

In a discourse between several researchers, led by Simner (2012), there was an examination of whether some of the criteria by which synaesthesia has traditionally been defined need to be reevaluated in the light of recent findings, as they were now potentially excessively conservative. As a result, Simner suggests that characteristics of a subset of synaesthetes have been erroneously attributed to all synaesthetes and thus interpreted as defining features of synaesthesia. Simner's key arguments were:

1. That referring to synaesthesia as a 'merging of the senses' is not strictly accurate, as some involve cognitive or conceptual triggers and/or concurrents.
2. Second that insisting on the long-term consistency of inducer-concurrent pairings over time as a defining feature of synaesthesia is unnecessary, as it neglects the possibility that some individuals may experience a less consistent phenomenon that is nevertheless still perceptually real.
3. Third, Simner argues that the requirement that the induced concurrent qualia necessarily have a fixed position in external local space is also overly restrictive as many synaesthetes report, for example, a strong internal 'feeling' of a colour associated with a particular triggering inducer.
4. Finally she argues that a definition that is rooted solely in behavioural criteria is inadequate, and we should also be using some kind of physiological definition of synaesthesia, one that outlines one or more biological signatures of synaesthesia, such as hyper-connectivity between cortices or disinhibition of neural connections.

For each of these four arguments Simner provides supporting evidence from a range of empirical studies. Simner's thesis prompted a thorough response first from Eagleman (2012) and then

Cohen Kadosh and Terhune (2012), who were broadly supportive of her argument and each of whom also provided valuable additions to each component of Simner's argument. We shall now review the consensus reached in this commentary including the arguments in favour of reconceptualising synaesthesia as a spectrum rather than a dichotomy. We believe that our work with vEAR may potentially be pertinent to many of these arguments.

1.4.1.1 Neurophysiological Diagnostic Criteria

Simner suggests that some form of biological criteria should feature in our definition of synaesthesia, to supplement the behavioural measures currently used. These biological factors may encompass factors such as atypical levels of connectivity between neighbouring sensory cortices (Bargary & Mitchell, 2008), or alternatively the overall levels of connectivity may not underlie the condition but instead atypical levels of disinhibition of this connectivity maybe observed in synaesthetes (Grossenbacher, 2001; Neufeld et al., 2012). As Eagleman then notes, connectivity is necessarily spectral rather than all-or-none, meaning that diagnostic criteria along these lines may be ultimately impractical as it would involve the setting of arbitrary levels of connectivity beyond which one is considered a synaesthete. Eagleman suggests that there might be several diverse causes of different sub-types of synaesthesia, analogous to the multiple causes of deafness, such as inner ear damage, or cranial nerve lesions. Eagleman points to his recent family-linkage analysis of coloured-sequence synaesthesia (CSS), in which they found a candidate gene that may underlie CSS. However, this gene was not present in all families in which CSS appears to run, which suggests that there may be multiple biological pathways to developing CSS. This would imply that even within a single sub-type of synaesthesia there may be differing aetiologies between individuals. Cohen Kadosh and Terhune concur with this point, citing dyslexia as a condition once thought to be homogenous that we now recognise as comprising of different sub-types and origins. They do however strike a note of caution when it comes to a setting out a biological definition of synaesthesia, asking for example what then would be the dependent variable used to diagnose the condition, and how would it/they be operationalised? The fact that synaesthesia appears to manifest so rarely in the population might be indicative of this being a dichotomous, all-or-none phenomenon rather than one that can be seen along a spectrum throughout the population. If this is indeed the case then it is likely that the brains of synaesthetes are indeed special in some sense, whether in terms of the connectivity between regions or the excitability of regions pertinent to the qualia experienced by the synaesthete. For example, fMRI studies have shown that in grapheme-colour synaesthetes, brain areas such as the left lateral occipital cortex and in postcentral gyrus show a reduced BOLD signal when synaesthetic experiences are perceived (O'Hanlon, Newell, & Mitchell, 2013), whereas Hubbard et al., (2005) found an increase in V4 activation when synaesthetes viewed greyscale graphemes that evoke synaesthetic colour. It is therefore hypothetically possible that given a greater understanding of these neural idiosyncrasies we could one day identify the brains of a synaesthete based on physiological markers alone. However it is hard to conceive of a reason, if the dichotomous account of synaesthesia is valid, that this would be particularly advantageous compared to simply asking the individual to describe their experiences. As such, any such biological test would serve to do little more than corroborate the veracity of such claims. If on the other hand we reject this dichotomy, and accept that synaesthesia may manifest along a spectrum throughout the population, then it may well make more logical sense to include such physiological descriptions in defining in strictly mechanistic terms how these phenomena arise. Under this account too it is difficult to foresee how these physiological markers could be practically used to diagnose synaesthesia, however, as we are then presented with the uncertainties that fuzzy boundaries and perceptual thresholds of the condition would present. Therefore in this thesis we will refer to potential physiological mechanisms that may explain the experience of vEAR, such as the disinhibition and cross-activation hypotheses from the synaesthesia literature, in an effort to explain the phenomenon, but these will not be used in any sense to try and form a diagnostic tool. In particular, in chapters 3 and 4 we present physiological hypotheses to explain the aetiology of vEAR that go beyond the diagnostic, and may lead to new discoveries about the neurophysiological bases of a wide variety of unusual and anomalous forms of perception.

What are the physiological candidates for mechanisms that may underlie vEAR? Previous studies have identified groups of neurons in the auditory cortex that respond to various types of visual stimulation from higher areas or from subcortical regions via interconnections found in normal anatomy (Ghazanfar & Schroeder, 2006; Schroeder & Foxe, 2005). These connections may serve to aid spatial localisation of sound or alternatively to amplify the auditory response to visually identifiable sources (Schroeder & Foxe, 2005). Given this normal connectivity, the additional emergence of a conscious visually-evoked auditory concurrent might thus be readily explained by individual physiological variations in this connectivity. This variability may include whether these connections, or their interactions with higher areas, are inhibited or unmasked (Cohen Kadosh et al., 2009; Cohen Kadosh & Walsh, 2006; Grossenbacher & Lovelace, 2001; Neufeld et al., 2012), to greater or lesser degree.

1.4.1.2 Non-sensory components of Synaesthesia

As noted in section 1.3, some types of synaesthesia feature an inducer that is a cultural construct, such as a day of the week, or a cognitive component such as a numerical value, rather than a strictly sensory stimulus. Delineating between sensory and cognitive inducers becomes more complicated in sub-types such as grapheme-colour synaesthesia as both the physical form of the text and the semantic understanding of what the written shapes signify can both potentially be what triggers the synaesthetic concurrent. Simner (2012) notes that in most cases of grapheme-colour synaesthesia it is the cognitive element, rather than simply the physical shape of the grapheme, that triggers the association. For example, altering the typeface of the text does not typically alter the concurrent colour that is perceived by the synaesthete (Grossenbacher & Lovelace, 2001), and an ambiguous stimulus (e.g. “I”) may have a different associated colour whether it is presented in a string of numbers or of letters (e.g. I2345 vs. Imnop), although the characters are physically identical (Dixon et al., 2006, Myles et al., 2003). This suggests that it is not a purely sensory phenomenon, as the synaesthetic association is not being driven purely by what the eye sees, but also by a top-down interpretation of the stimulus. However, as Simner notes, although this is true of the majority of grapheme-colour synaesthetes it is not accurate in all cases, as in other cases the colour of the concurrent will vary with changes in the physical form of the grapheme (such as typeface, stylisation etc.), leading to the distinction between lower synaesthetes (triggered primarily by the physical form of the inducer) and higher synaesthetes (triggered by top-down interpretation of the inducer). This distinction may be pertinent to the experience of vEAR, as we might wish to examine whether the phenomenon is primarily driven by objective physical features of the visual stimulus (e.g. brightness, motion velocity, smooth vs. jerky motion etc.) and/or by a learned association between the stimulus and the expectation of a sound. This is one of the questions addressed in chapter 4.

1.4.1.3 Consistency of Inducer-Concurrent Pairings

Simner next points out that an overemphasis on the necessity for stable inducer-concurrent pairings may be neglecting the possibility that some synaesthetes experience transitory or variable synaesthetic experiences. She points out that some individuals claim to experience synaesthesia but routinely fail consistency tests. Either the claims of these people are untrue or our current conceptualisation of synaesthesia is inadequate for encompassing these individuals' experiences. On measures of consistency synaesthetes tend to score 80-100% consistent in their reported pairings, while non-synaesthetes score around 20%, but as Simner notes many people score in between this range. On this point Eagleman (2012) points to the extremely large dataset generated though his own online synaesthesia battery (synesthete.org; Eagleman et al., 2007), which includes objective tests of stimulus-colour pairing consistency, and notes that they do not find a bimodal distribution of scores, as one would expect if synaesthesia were an all-or-none phenomenon, but instead reports some highly consistent scores and a long tail representing less consistent scores (Eagleman, 2012), consistent with the conceptualisation of synaesthesia as a spectrum or a collection of related yet distinct phenomena. The issue of consistency too has implications for the study of vEAR, as informal reports from those who experience the phenomenon that we have spoken to often describe the concurrent sensation as a generic 'white noise' type of sound, making it harder to test participants for consistency, or some participants report that the concurrent is being influenced by the context of presentation. For example, after speaking to participants in our experiment outlined in chapter 2, some reported that the visually-evoked sounds took on the characteristics of the auditory beeps used in the auditory condition of the task, something that Saenz and Koch (2008) noted was also true of some of their synaesthete participants. This lack of specificity does not necessarily make these sensory phenomena any less perceptually real to the individual rather their inducer can evoke a range of different concurrents depending on context. One way to test this consistency is presented in chapter 4, where we present the results of a survey in which respondents were asked to rate a range of different visual stimuli (silent video clips) for the intensity of vEAR that each evokes. We then analyse the consistency of these specific subtypes of inducers in evoking sounds, based upon properties of the stimulus such as whether the content of the video is naturally associated with an accompanying sound, or the amount of motion energy contained within the scene. This latter property is a low-level physical characteristic of the visual stimulus and refers to the net amount of movement within a visual scene, which is obtained by quantifying changes in light in both space and time. Only those who report experiencing vEAR are sensitive to motion energy. In this sense we demonstrate consistency between particular types of visual stimulus and the vEAR it evokes in our respondents.

1.4.1.4 Spatially-Mapped Concurrents

Simner also criticizes the perceived diagnostic requirement for synaesthetic concurrents to have a perceived spatial location external to the synaesthete. Eagleman concurs with this point and highlights reports in the literature (e.g. Dixon, Smilek & Merikle, 2004) that some grapheme-colour synaesthetes reporting their associated concurrent colours as being located in external space, as if the text itself were printed in coloured ink, while others experienced a more abstract mental association with the colour ‘in the mind’s eye’. They term these ‘projectors’ and ‘associators’ respectively, with the latter forming a significant majority of grapheme-colour synaesthetes. Ward et al., (2007) further subdivide the projector synaesthetes into two sub-types;

surface-projectors, who will see the inducing stimulus itself, generally written text, as if it were printed in the associated colours, and space-projectors, who will perceive a coloured glow in the external space surrounding the inducer. This suggests that there are individual differences in how explicitly the association between inducer and concurrent is perceived by synaesthetes. The projector/associator distinction has since been applied to other forms of the condition, including sound-colour synaesthesia (Simner, 2012), and would seem to suggest that some form of graded synaesthetic spectrum does indeed exist, at least between individuals with synaesthesia.

Eagleman points out that Rouw and Scholte (2007) attempted to clearly behaviourally differentiate between projector and associator synaesthetes, but did not find bimodal distribution, instead results were a smooth continuum. This might suggest that these are related phenomena that represent either end of a spectrum, rather than two distinct categories. Eagleman then proposes a useful analogy, that of obesity and anorexia as opposite ends of a body weight continuum. There are clear medical reasons why in medicine it makes logical sense to study either end of this continuum in isolation, but we would not attempt to tenuously assign the average weight people in the centre into either category. Likewise with synaesthesia it may be that there is a spectrum of synaesthetic abilities that has simply been neglected by dogmatic sampling criteria that then feed a circular definition of synaesthesia. When we asked our participants ‘where’ the sounds characteristic of their vEAR were perceived to occur, our participants tended to report that they were inside their head, that they were mental sounds, or just the feeling of a sound. None reported perceiving them as occurring in external space, such as mistakenly attributing the spatial source of the flashes to the computer monitor, or from the computer speakers, for example. Nor did any report believing that the sounds were played by the experimenter in an attempt to deceive them. We can therefore assert that these visually-evoked sounds are not mapped to any point in external space to meet these diagnostic criteria for canonical synaesthesia. However, if we are able to demonstrate that these vEAR sounds are able to interfere with the detection of real-world sounds, as in chapter 2, then we may at least be able to support our assertion that these sounds are genuinely being 'heard' rather than imagined through some form of auditory imagery.

1.4.2 Is Synaesthesia Dichotomous or Continuous?

Taken as a whole the arguments by the authors involved in this debate appear sympathetic to the prospect of synaesthesia (or at the very least some of its sub-types) manifesting as a continuum. This is contrary to the traditional depiction of synaesthesia as a rare and unusual phenomenon that only affects a small percentage of the population, with estimates of the prevalence of the condition typically ranging between 2-4% of the population for the more common variants such as grapheme-colour synaesthesia (Simner at al., 2006; Ward, 2013). However, in estimating the prevalence of synaesthesia, the assumption being made is that there are synaesthetes and non-synaesthetes, and that we all fall into either the former or, predominantly, the latter camp. Proponents of the dichotomous view of synaesthesia might argue that it is likely to be the result of a genetic variants that are not frequently observed in the population (Asher, 2009; Brang & Ramachandran, 2011; Tomson et al., 2011). Under this theory the anatomy of synaesthetes is inherently different to the rest of the population and their perceptual world is unique, meaning there can be no meaningful middle ground between the presence and absence of synaesthesia. However, cases have been reported in which only one of two monozygotic twins experiences synaesthesia (Smilek et al., 2002; Smilek, Dixon, & Merikle, 2005) which suggests that other factors than genetics influence the development of the condition.

The alternative 'spectrum' account of synaesthesia is not necessarily incompatible with the evidence for a genetic component of synaesthesia, as any associated gene may be 'boosting' one's latent synaesthetic abilities, rather than creating them outright. The continuum account is supported by individual differences in the perceptual vividness of synesthetic abilities even between synaesthetes, as with the distinction between projector and associator synaesthetes. As Eagleman notes, the Diagnostic and Statistical Manual (5th Edition) of the American Psychiatric Association is moving to toward spectral definition of many conditions (e.g. autism) so it might be time to start considering synaesthesia along these criteria. Yet the lack of a bimodal distribution of scores on Eagleman's synaesthesia battery and in Rouw and Scholte (2007) work, in particular suggest that those who appear in the midrange of scores display some synaesthetic tendencies and are currently under-researched. The manner in which synaesthetes are typically identified for research purposes tends to implicitly presupposes that the dichotomous account of synaesthesia is correct. Synaesthetes are almost exclusively identified through self-selection, and likewise control subjects are also identified through similarly subjective methods by simply excluding them from the other category.

A thought experiment in which we assume the validity of the continuum hypothesis exposes why this sampling bias may occur. If synaesthesia is a common condition in which the intensity, the perceptual vividness, of sensory crossover varies throughout the population then it follows that for some the concurrent sensory percepts are considerably more subtle than in the standard account of synaesthesia, with many of these pairings at, around, or below their perceptual threshold. Consider too that this has been their sensory reality since birth, and is utterly unremarkable to the individual. It follows that for many of these people they would be completely unaware that they were in any way 'synaesthetic' and when questioned about sensory crossovers using subjective measures they would more than likely respond in the negative. This would then lead to a sampling bias in which only the extreme ends of the spectrum are captured and thought to be unrelated samples rather than opposing poles. This suggests that we might observe a very different pattern of results using subjective measures of synaesthesia experience to those gained through objective measures alone (such as consistency testing, or the "pop-out" test of grapheme-colour synaesthesia used by Rich & Karstoft, 2013). This highlights the importance of the mixed methodology approach employed in this thesis, such as combining our psychophysical data in chapter 2 with a subjective questionnaire about the individual experience of each participant while they performed our tasks, or the online vEAR survey that we report in chapter 4.

At this point we wish to note that we refer in this thesis at several points to people who do versus do not experience vEAR. In doing so we are referring to those who report that they are consciously aware of the experience, and this is not intended to suggest that we favour a dichotomous 'all-or-none' explanation of vEAR. We acknowledge that there may be individuals who are experiencing subtle sensory crosstalk of which they are not consciously aware but that may be experienced sub-threshold. In the following chapter we present evidence that some individuals may be experiencing vEAR without realising it until their attention is drawn to the phenomenon under laboratory conditions.

1.4.3 Relationship with Normal Multisensory Integration

Cohen in a 2017 essay argues that synaesthesia is continuous with ordinary multisensory integration or, as the title somewhat provocatively suggests, that “we’re all synaesthetes now”. His argument is based upon the following premises; 1) that the defining features of synaesthesia are present to the point of ubiquity in the typical population, 2) that there are both similarities and differences between synaesthetic and non-synaesthetic cross-modal associations, and 3) that synaesthetes demonstrate enhanced performance on a number of integrative perceptual tasks relative to controls. These three arguments seemingly support the idea of the synaesthete as a ‘super-integrator’. Synaesthetes are not unique in integrating information from multiple senses; it is a normal human ability that we all do all the time. We are constantly surrounded by sights, sounds, smells, flavours and textures, and every object we come across may harbour a multitude of sensory properties. On encountering a tree, for example, we may see the colour and shape of fruit, smell the sweet scent of blossom, feel the coarse texture of bark or hear the sound of leaves rustling in the wind. Yet we perceive the structure as a single unified object in space, rather than as several disconnected streams of incoming stimuli. We must therefore have evolved a mechanism in the brain for binding those sensory qualia that arise from a common origin in the outside world. Our understanding of how we integrate information from the different senses has undergone considerable refinement in recent decades. In earlier orthodoxies, information from each sensory modality was thought initially to be processed independently in early unisensory regions, such as primary visual cortex, before undergoing upstream binding in higher-level regions

(Felleman & Van Essen, 1991; Murray et al., 2016; Treisman & Gelade, 1980). In this framework it made sense to compartmentalise early cortices as auditory, visual, or somatosensory. However, as our understanding of sensory processing in the brain has advanced, focus has shifted away from this strictly modular and unisensory view to an understanding that places a greater emphasis on multisensory interactions. Indeed, we now know that multisensory integration does in fact occur in lower-level cortical regions, previously thought to be strictly unisensory, and even as early as the primary cortical level (for reviews see Kayser & Logothetis, 2007; Schroeder & Foxe, 2005).

Ghazanafar & Schroeder (2006) examined the accumulating evidence for early multisensory integration and suggesting that the neocortex is “essentially multisensory” in nature. This idea has since gained widespread support (van Atteveldt et al., 2014; de Meo et al., 2015; ten Oever et al., 2015), with Murray et al. (2016) recently proposing that *“the convergence and integration of information from different senses within low-level cortices is a rule rather than an exception”* (p.161). Similarly Fulkerson (2014) advocated a model of ‘sensory pluralism’ with an emphasis on the multiple interactions that occur in sensory cortices. In addition, anatomical connections have been found between early visual areas such as V1 and early auditory cortices in a number of species, including several species of rodent (Henshke et al., 2015; Laramée et al. 2013; Vaudano et al., 1991), cats (Clemo et al., 2008) and in nonhuman primates (Cappe & Barone, 2005; Falchier et al., 2009), and in individuals who are born blind the primary visual cortex has been shown to respond to nonvisual inputs such as tactile stimuli in the form of Braille (Cohen et al., 1997, 1999; Likova, 2012; Sadato et al., 2002) and sounds (Bedny, Richardson, & Saxe, 2015), demonstrating that under the right circumstances, cortical rewiring can occur outside the realm of synaesthesia giving rise to other forms of cross-modal phenomena.

We know then that multisensory interactions occur all the time in the normal population, and that cortical projections exist between primary cortices of each sensory modality, suggesting that the potential cortical architecture to support synaesthetic-like traits in the normal population are in place. Furthermore, even in the general population there is a propensity to experience what have been called cross-modal correspondences, defined as *"a tendency for a sensory feature, or attribute, in one modality, either physically present or merely imagined, to be matched (or associated) with a sensory feature in another sensory modality "* (Spence & Parise, 2012, p. 410). These are distinct from synaesthesia (according to commonly used diagnostic criteria) as the associations are not explicitly perceived as sensations (as in synaesthesia), rather they are intuitively associated when asked to make seemingly arbitrary pairings between stimuli. The key distinction between cross-modal correspondences and synaesthesia therefore is that in the latter an input in one sensory or cognitive pathway triggers sensory or cognitive experiences in another domain that are explicitly perceived in a manner in which the concurrent is perceptually real to the synaesthete. In cross-modal correspondences, however, there is no concurrent percept, rather there is a general sense of two different sensory pairings feeling appropriate on a conceptual or aesthetic level. Might cross-modal correspondences and synaesthesia share a common origin in the brain? Certainly, there appear to be a correlation between the reported sensory pairings of some synaesthetes and the cross-modal correspondences experienced by the wider population (Cohen Kadosh et al., 2007; Cohen Kadosh & Henik, 2007; Eagleman, 2009; Simner et al., 2005; Ward, Huckstep, & Tsakanikos, 2006). For example, in music–colour synaesthesia synaesthetes and non-synaesthetes alike have a tendency to associate high frequency sounds with lighter colours and low frequency tones with darker colours, and although only the synaesthetes are aware of consciously perceiving these pairings, non-synaesthetes nevertheless will intuitively match them when prompted (Ward, Huckstep, & Tsakanikos, 2006). Similar associations between grapheme-colour mappings have been demonstrated between synaesthetes and non-synaesthetes (Simner et al., 2005).

Cross-modal effects like these can influence performance on a number of tasks. One example of these would be a selective attention task, in which the perceiver is asked to attend to a stimulus in one sensory modality and ignore information in another. Previous research has demonstrated a congruence effect (e.g. Melara & O'Brien, 1987), whereby task performance in enhanced when the stimuli in both the attended and unattended modality match on some criteria, such as auditory pitch and lightness of colour, where participants are quicker to classify high pitched sounds when paired with (unattended) white visual stimuli and vice versa for low sounds with black visual stimuli (See Martino and Marks, 2000, for a review and other examples). Martino & Marks (2000) demonstrated that this congruence effect is also present when using low vs. high frequency vibrotactile stimuli paired with either a black or white visual stimulus (where congruent pairings were Black with low frequency and white with high frequency tactile stimulation). Analysis of response times demonstrated that this was due to a facilitation effect of matching pairs, where responses were quicker than with unisensory reaction times at baseline whereas reaction times to incongruent pairs were the same as when presented alone at baseline. These results demonstrate that cross-modal interactions may depend on synaesthesia-like associations across the senses that are not exclusive to those with canonical synaesthesia.

Such examples of cross-modal associations presumably follow normal rules of multisensory perception, in which information from more than one modality are perceptually 'bound' if they occur in sufficient temporal and/or spatial coincidence (Meredith & Stein, 1986; Meredith, Nemitz & Stein, 1987). Bien and colleagues (2012) examined what occurs when the bottom-up characteristics such as these are ambiguous, and whether the cross-modal associations described above play a role in addressing this ambiguity. To test their pitch-size hypothesis, which suggests that small objects will be grouped with high pitch sounds and larger objects with lower pitched sounds, they used an auditory spatial localisation task, known as the ventriloquist paradigm (Driver, 1996), in which spatial correspondence is manipulated to explore the boundaries of the binding process. Bien et al. (2012) results showed that despite the violations of spatial coincidence, congruent pairings are more often integrated than incongruent pairings. This ventriloquism effect was then disrupted using Transcranial Magnetic Stimulation (TMS) that was applied to the right intraparietal sulcus, a region known to play a role in multisensory integration. This provides further evidence that low level synaesthetic-like mappings are a normal feature of multisensory integration in the brain, and that this in turn supports the reconceptualisation of synaesthesia as operating along a spectrum.

However, others have argued in favour of the separatist view that synaesthesia is indeed a rare phenomenon, distinct from other cross-modal phenomena, that manifest only in a select group of individuals, and counter any claims of a middle ground between synaesthesia and typical perception. For example, Deroy and Spence (2013b) critique claims that experiences such as cross-modal correspondences represent a form of ‘weak’ synaesthesia and point to the defining criteria of synaesthesia as involving conscious percepts which are absent in cross-modal correspondences. This argument could be seen as somewhat circular, as it limits discussion of synaesthesia to definitions that have been set on the basis of previous research, and ignores the possibility that borderline cases of synaesthesia involve percepts that perhaps the individual has habituated to and thus learned to ignore, but could nevertheless be explored using bespoke psychophysical and neuroimaging methodologies.

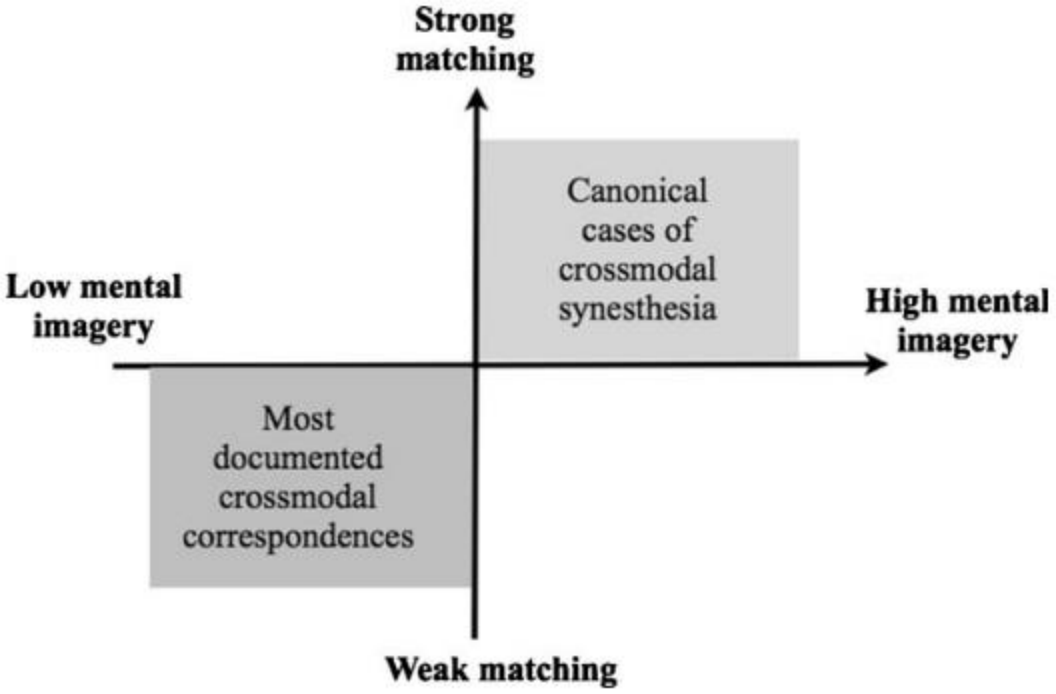


Figure 1-0-3: Model of relationship between synaesthesia and cross-modal correspondences.

Note that the model could represent either a spectrum of cases or discrete categories. (Based on Rader & Tellegen, reproduced from Deroy and Spence 2013b, p. 652).

Deroy and Spence (2013b) also note that while there are certainly similarities between certain synaesthesia sub-types and cross-modal correspondences, in many cases there are also key differences. One such discrepancy they suggest is the ordered nature of the pairings in cross-modal correspondences. While the sensory pairings in synaesthesia are often, but not always, arbitrary and tend to be diverse in their variety between synaesthetes, there is often a plausible explanation for the mappings of cross-modal correspondences, such as frequency of exposure in the natural world (reviewed in Spence, 2011). To take the object size/sound frequency correspondence discussed above (Bien, et al., 2012), one could easily imagine that large objects are associated with lower frequencies and smaller objects with higher frequencies as this is commonly the case with the human voice, or a dog's bark, for example, in both cases we likely have learnt that larger body size often correlates with a deeper, booming voice. However as noted above, in other types of synaesthesia such as music-colour, the sensory correspondences reported in the general population appear to correlate with those of the synaesthete population (Ward, Huckstep, & Tsakanikos, 2006). In cases such as these the argument for crossmodal correspondences as 'weak' form of synaesthesia is strongest, although given the heterogeneous nature of the condition it is likely that this is not a universal property of either synaesthesia or of crossmodal correspondences.

Despite these learned correspondences being common throughout the population, it is possible that synaesthetes are the 'super-integrators' at the top of the spectrum (see Figure 1-0-1), and while most of us only integrate at the level of a vague sense of two sensory components 'feeling right' together, those at the far end of the scale display a propensity to integrate even more unusual cross-modal phenomena that are explicitly perceived on a sensory level. Claims that synaesthesia is an enhancement of normal multisensory processing gains some support from a number of studies which claim to have induced synaesthesia-like percepts in non-synaesthetes through training (e.g. Bor et al., 2014; Rothen, Wantz & Meier, 2011; reviewed in Rothen and Meier, 2014). For example, daily training sessions for seven days, in which letters are paired consistently with an associated colour, has been shown to induce a ‘synaesthetic’ Stroop effect whereby incongruent letter-colour pairings impair response times on a colour-letter matching task, although this did not extend to experiencing any overt percepts in response to the trained stimuli (Meier & Rothen, 2007). This synaesthetic Stroop effect can also be induced through the reading of books in which letters are consistently presented in an associated colour (Colizoli, Murre, & Rouw, 2012).

There are also reports of synaesthesia-like experiences being induced by the ingestion of certain psychoactive substances such as lysergic acid diethylamide (LSD), mescaline, and psilocybin (Luke & Terhune, 2013) and methamphetamine (Ahmadi, Keshtkar, & Pridmore, 2011). However, the sensory pairings experienced under these conditions are not usually fixed (i.e. colour pairings may vary between different periods of drug ingestion), unlike the canonical criteria for synaesthesia (Sinke et al., 2012). Other studies have attempted to induce synaesthesia via post-hypnotic suggestions, with mixed success. Cohen Kadosh et al. (2009) report associating particular letters with particular colours under hypnosis. Participants were then given a simple Yes/No detection task where letters were presented over a series of coloured backgrounds. Task performance was impaired when the letter was presented on a background matching the colour made under hypnotic suggestion, but this effect was not seen in controls, suggesting that the letter was to some degree being perceived or associated with the hypnotically-suggested colour. However other studies have reported that while hypnotic suggestions are able to induce reports of perceived letter-colour associations, these are not perceived in a comparable way to true synaesthesia as it did not aid performance on an embedded figures test (Anderson et al., 2014).

If synaesthesia can be transiently experienced in non-synaesthetes through hypnosis, training, or through the use of psychedelics, it would suggest that the cortical architecture by which synaesthesia manifests is not unique to synaesthetes but that the connectivity is widely present in the population, as it is considerably more plausible that these substances alter chemical expression at existing synapses rather than generate new, temporary ones. It may be that training-induced synaesthesias are achieved by reinforcing multisensory neural connections that are common to all of us, through Hebbian learning, the process by which synaptic connections which often fire in synchrony become stronger, while drugs may temporarily alter patterns of inhibition or transmission of neurotransmitters such as serotonin (Luke & Terhune, 2013). This would allow the individual to transiently access the same abilities to which synaesthetes have permanent access. Moreover, there is evidence to suggest that in individuals with synaesthesia learning and environmental exposure plays some part in determining the specific cross-modal pairings experienced by synaesthetes.

1.4.4 Are Synaesthetic Pairings Environmentally Influenced?

It has been suggested that early environmental experience can influence the specific sensory pairings experienced by adult synaesthetes, and there is some evidence to date that suggests it may (for a review see Watson et al., 2014). Evidence for the role of environmental exposure in the development of synaesthesia can be seen in the high occurrence of inducers that form part of a formal sequence (e.g. Monday, Tuesday, Wednesday...; A, B, C...; 1, 2, 3...) (Rich, Bradshaw, & Mattingley, 2005), suggesting that these pairings are formed as part of a formally prescribed learning process. Witthoft, Winawer, and Eagleman (2015) reported that the specific colour-letter pairings experienced by grapheme-colour synaesthetes correlate with exposure to the coloured letter fridge magnet toys that have been popular with children for several decades. This may partially explain why there is a great deal of consistency in the specific pairings experienced between synaesthetes, for example, Rich, Bradshaw & Mattingley (2005) also reported that the letter 'Y' elicited the colour yellow in 45% of their sample while the letter 'D' elicited brown for 47%. These pairings are learnt in childhood, with child grapheme-colour synaesthetes having concurrent colours for approximately 35% or letters by the age of 7, rising to 70% of letters by the age of 11 (Simner et al., 2009; Simner and Bain, 2013). Furthermore, studies of coloured-sequence synaesthesia (CSS) in twins have shown that the concordance of inducer-colour pairings are 73.9% in monozygotic twins with CSS, compared to 36.4% in dizygotic twins with CSS (Bosley & Eagleman, 2015). This suggests that although biology clearly plays a role in determining the specific sensory pairings perceived in CSS, environmental exposure must also play a significant role. These findings do not explain why some individuals do develop synaesthesia and others do not, but it does suggest that in those that do, childhood exposure may partially dictate the specific inducer-concurrent pairings that persist into adulthood.

An interesting question was posed by Cohen Kadosh, Henik & Walsh (2009); is synaesthesia learned or lost? Some have suggested that in the first few months of life synaesthesia is the norm, exemplified by the neonatal synaesthesia hypothesis. This was first suggested by Maurer &

Maurer (1988), later by Baron-Cohen, 1996), after observations that cross-modal learning effects in the audiovisual and visuotactile domains are both common and most pronounced in the first few months of life, as are preferences for certain cross-modal correspondences over others (Meltzoff & Borton, 1979), before these preferences decrease until the age of around 8 months (Wagner & Dobkins, 2011) when the senses are considerably more modular. However, DeRoy and Spence (2013a) present a sceptical history of the neonatal synaesthesia hypothesis, arguing for example that these neonatal multisensory percepts lack the specificity and consistency to qualify as true synaesthesia, and that they are closer to the general cross-modal correspondences experience by the majority of the adult population. However, see section 1.4.1.3 for a discussion of why an emphasis on diagnostic criteria such as consistency of pairings may not necessarily be a suitable yardstick for defining synaesthesia.

Having established that early environmental exposure may play a role in the development of synaesthetic pairings, and that these associations may be formed at a preverbal developmental stage, we might ask whether synaesthesia is more common for sensory pairings that co-occur more frequently in the environment compared to those with arbitrary or bizarre pairings, such as lexical-gustatory synaesthesia? If environmental exposure in infancy influences the specific inducer-concurrent pairings experienced by a synaesthete (e.g. Bosley & Eagleman, 2015; Mankin & Simner, 2017; Witthoft , Winawer, and Eagleman, 2015) then one might expect that stimuli which are highly correlated in the natural world would be more commonly experienced as inducer-concurrent pairings than the more unusual and arbitrary pairings, such as lexical-gustatory synaesthesia, that are not constantly reinforced through learning, repetition and experience. One example candidate for such a common real-world pairing could well be sound and visual motion, which are highly correlated in our environment, with abundant examples of moving objects that emit temporally correlated sounds, including lip-movements and speech, a bouncing ball, motion-to-impact sounds (a door slamming, clapping hands, or footsteps). Furthermore these pairings can be experienced from birth and do not rely on the child to be sufficiently cognitively developed to possess concepts such as numbers and letters, meaning they are reinforced from a considerably earlier age, and crucially prior to the critical period of development after which synaptic pruning occurs. This is particularly crucial, because once these associations form via Hebbian learning they may be less likely to be pruned back, and therefore persist into adulthood, consistent with our prediction that vEAR should be more prevalent in the adult population than canonical forms of synaesthesia.

1.5 Summary and Thesis outline

Although there is a large body of literature exploring synaesthesia, prior to our own research the Saenz and Koch (2008) paper was the only published exploration of hearing-motion synaesthesia (or vEAR as we will refer to this phenomenon). There are then a great number of unanswered questions pertaining to vEAR. Equally there are a number of ongoing debates pertaining more broadly to the field of synaesthesia research, many of which could potentially be informed by the findings of this thesis. For example, are the visually-evoked sounds perceptually real enough and suitably vivid to interfere with the ability to hear genuine real-world sounds? Canonical reports of synaesthesia often report that there are specific and consistent pairings between the inducer stimulus and the triggered concurrent percepts that remain persistent over time. For example, middle C on the piano might evoke the colour blue, D red, and E green. Based on the reports in Saenz and Koch (2008) and more informally from conversations with our participants, the relationship between the visual inducer and the evoked auditory sensations are less specific, often described using terms such as whirring, buzzing, whooshing or humming. For this reason we emphasise that we are not at this stage referring to the vEAR phenomenon as a type of synaesthesia. However, if we are able to demonstrate that the evoked sounds that characterise vEAR are able to interfere with the detection of real-world auditory signals then we are at least able to confidently say that they are ‘sounds’ that are ‘heard’.

We do not know the population prevalence of vEAR, although as discussed above we believe there are reasons to predict that it may be more common than reports of canonical synaesthesia might suggest. It is also not clear if this is a unique ability possessed by a small number of individuals who perceive a vivid and conscious visually-evoked sound that is absent in the larger population, or whether we all experience the vEAR phenomenon to some extent, many of us without realising it. It is possible that vEAR is experienced as a graded, continuous phenomenon throughout the population with the intensity (loudness) of the visually-evoked sound displaying considerable individual differences. There is an ongoing debate in the synaesthesia research about the extent to which the condition is spectral or is dichotomous, and if we are able to demonstrate that the experience of vEAR is indeed graded along a continuum then can say that there are occasions when an inducing stimulus in one modality can trigger percepts in another modality that vary in intensity throughout the population. Furthermore, if the prevalence of vEAR is considerably greater than canonical reports of synaesthesia then it might indicate that learned associations between the senses that are formed early in life have some influence of the development of these cross-modal sensations, as visual motion is highly predictive of sound in the natural environment, particularly compared to some of the more esoteric pairings experienced by some synaesthetes.

Finally, we do not know the mechanisms by which the visually-evoked auditory qualia characteristic of vEAR originate. For example, it might be that there are extraneous connections between the visual and auditory cortices in people with vEAR that are absent in those who do not. Alternatively, the physical connections themselves may be no different to those in the wider population, rather this audiovisual connectivity is usually inhibited. If so, vEAR might reflect greater levels of baseline activity or of disinhibited feedback between the regions in people with vEAR. This too is a debate that exists in the wider synaesthesia literature to which our work with vEAR may contribute.

To conclude, the aims of the research presented in this thesis are to establish a plausible estimate of the prevalence of vEAR, to establish whether these are truly auditory signals that can interfere with real auditory signal detection, and to examine whether this is a true sensory phenomenon rather than a cognitive or attentional bias. We aim to explore how the brains of people who do versus do not experience vEAR may differ from one another, specifically in how their auditory and visual cortices interact to produce (or not) these visually-evoked auditory sensation. Finally, we will examine what types of visual stimulus best evoke these sensation, and the properties of these stimuli that best evoke the sensation. By examining whether it is, for example, low-level physical characteristics of the stimulus or higher learned associations between a stimulus and the prediction of a sound, we can then speculate further on the neural mechanisms that may be at work in those individuals who experience sounds when viewing silent visual motion. Each of these aims will then be used to inform some of the above debates about synaesthesia, and whether vEAR meets enough of the accepted criteria to be considered a synaesthetic sub-type.

Experiment 1: The prevalence, advantages, and perceptual reality of visually-evoked sounds[1].

Chapter 2:

This chapter attempts to estimate the prevalence of vEAR in a random sample using both objective measures following Saenz and Koch (2008) and additional and subjective methods. In subjects who experience vEAR we measure how perceptually real the evoked sounds are using an objective measure of the effect of visual flashes on auditor detection following Lovelace, Stein and Wallace (2003). Potential non-sensory explanations for our findings are discussed, and evidence to rule out alternative explanations such as an attentional bias toward vision is presented. Results are discussed with reference to the contemporary debates surrounding synaesthesia outlined in the previous chapter.

2.1 Introduction

In the present chapter we primarily address two questions: how prevalent is the ability to 'hear' visual events as described by Saenz and Koch (2008), and how perceptually 'real' are these sounds? Estimating the prevalence of synaesthesia is complicated by the relative lack of consensus surrounding a working definition of synaesthesia, and by the remarkable heterogeneity

---

[1] This chapter is based on data that has previously been published in Fassnidge, C., Marcotti, C.C., & Freeman, E.D. (2017). A deafening flash! Visual interference of auditory signal detection. Consciousness and Cognition, **49**, 15-24.

of the condition, but figures of about 2 to 4% of the population have been proposed for the most common forms of the condition, such as grapheme-colour synaesthesia (Simner et al., 2006; Ward, 2013). This low prevalence has led some to suggest that the aetiology of synaesthesia may lie in a rare genotype (Brang, Williams, & Ramachandran, 2012; Tomson et al., 2011), which in turn may give rise to unusual patterns of neural cross-wiring or cross-activation between adjacent cortical regions (Ramachandran & Hubbard, 2001; Rouw & Scholte, 2007). However, an alternative school of thought argues that some forms of synaesthesia might be grounded on the normal Hebbian learning mechanisms involved in forming and reinforcing associations between different modalities and sensory dimensions (Brang, Williams, & Ramachandran, 2012; Cohen, 2013; Cohen Kadosh, Henik, Catena, Walsh, & Fuentes, 2009; Cytowic, 2003; Grossenbacher & Lovelace, 2001; Ramachandran & Hubbard, 2001; Ward, Huckstep, & Tsakanikos, 2006). This theory need not necessarily run counter to the argument for a genetic cause of synaesthesia. Rather, one may be genetically disposed to synaesthesia in general, with the specific inducer-concurrent pairings perceived by the individual synaesthete learnt and reinforced through environmental exposure. This means that the rarity of synaesthesia might be explained by how infrequently the kinds of exotic associations that typify many forms of synaesthesia are found together in nature. For example, repeated exposure to consistent letter-colour pairings (found in fridge-magnets, educational materials, or experimental stimuli) might shape and reinforce grapheme-colour synaesthesia (Bor, Rothen, Schwartzman, Clayton, & Seth, 2014; Witthoft, Winawer, & Eagleman, 2015), but grapheme-colour synaesthesia might still be rare because such consistent correspondences between letters and colours are themselves rare and thus do not typically reinforce strong associations. How then might the extremely common co-occurrence of the domains of sound and visual motion affect the prevalence of vEAR? We propose that vEAR may be more prevalent than other synaesthesias due to the high levels of audiovisual co-occurrence in the natural world.

2.1.1 Prevalence of vEAR

Our first goal was to estimate the prevalence of visually-evoked auditory sensations in a random sample. Saenz and Koch's (2008) study was the first and, until our research, only published report of hearing-motion synaesthesia, but the prevalence of the condition cannot be assessed from this study, as they used a small sample (n=14) and the few participants who were identified as synaesthetes were not randomly sampled but self-selected. By grouping participants in this way, with non-synaesthetes presumably chosen specifically for *not* showing any synaesthetic tendencies, they were arguably imposing a dichotomy, with the tacit presumption that synaesthesia and typical perception are binary in nature. This assumption risks neglecting the possibility that vEAR may present as a continuum running through the population, or that some form of proto-synaesthesia, where cross-modal sensations are present but below threshold, unknown to the participant but detectable through standard psychophysical practices. Indeed, there are valid reasons to consider that vEAR may show a unique prevalence profile compare to canonical synaesthesia because sound and motion are frequently co-occurring phenomena in our environment, compared to some of the more bizarre and arbitrary sensory groupings that have been reported in the literature. If learning or environmental exposure plays a part in the development of synaesthesia it would follow that frequently co-occurring stimuli during the critical period in infancy prior to the extensive synaptic pruning after the age of around 3 (Craika & Bialystok, 2006; Drachman, 2005) would lead to higher rates of the corresponding synaesthesia in adulthood.

Saenz and Koch devised a task to test their participants' claims to hear visual events. This required the participants to discriminate between rhythmic sequences of stimuli, presented either as beeps or as flashes (see chapter 1, section 1.3.2). Saenz and Koch presented their data sorted by visual sequence discrimination accuracy (Saenz & Koch, 2008, Supplemental data Figure S1, see Figure 2-1 here), with each subject's equivalent auditory accuracy presented adjacently (see Figure 2-1). The subjects who reported that they could hear visual events are highlighted in grey, and these same subjects also did not exhibit significant differences between the two modalities on the Sequence Discrimination Task, while the controls did (outlined in chapter 1). Saenz and Koch present this as objective confirmation that these individuals were indeed hearing the visual flashes, putatively because auditory recoding of the visual sequences benefits sequence discrimination due to better temporal acuity in auditory modality relative to the visual modality (Glenberg et al., 1989; Guttman, Gilroy & Blake, 2005). However, it is unclear what the variance of this discrepancy between visual and auditory sequence discrimination would look like in a randomly selected sample of the population. We might observe in several participants behavioural results that resemble those of the synaesthetes, in that visual and auditory sequence discrimination do not significantly differ. However, if these individuals did not report the ability to hear visual events then it would be clear that the Sequence Discrimination Task alone is not sufficient for diagnosing or corroborating vEAR, and a more objective measure of the effect of visual stimuli on audition may be required.

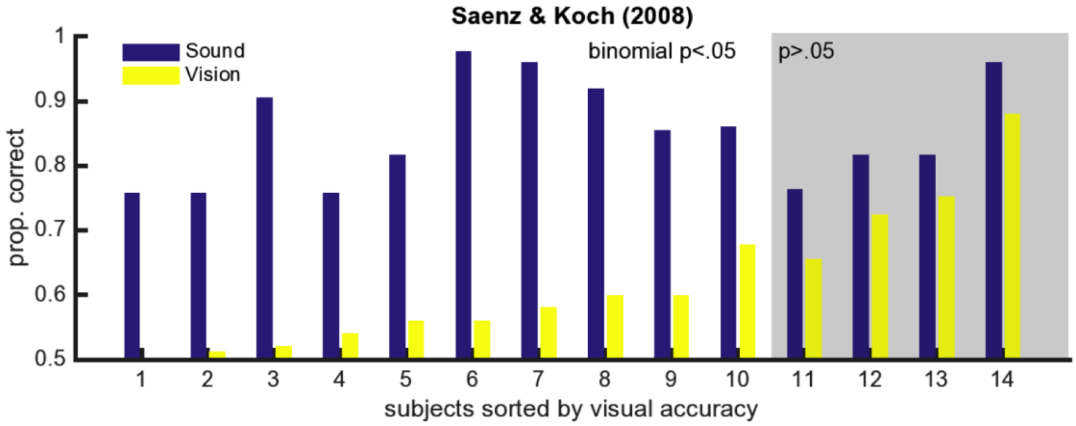


Figure 2-1: Visual vs. Auditory Sequence discrimination Accuracy replotted from Saenz & Koch (2008).

Data are presented ranked by visual sequence discrimination accuracy from low to high. Grey area denotes no significant difference between visual and auditory sequence discrimination (N.B, only synaesthete participants meet these criteria).

Saenz and Koch (2008) report their findings as if their synaesthetes and controls are two distinct groups, however we can see that there is no clear dichotomy in terms of the visual sequence discrimination ability between those who do versus do not report hearing visual motion (see Figure 2-1). Rather, we can see that when participants are ranked according to these scores, there is a smooth progression from poor visual sequencing to high accuracy at discriminating visual sequences. This would be consistent with our proposed conceptualisation of vEAR as a continuum, with a threshold beyond which vEAR comes into conscious awareness. In chapter 4 we will explore the potential mechanisms by which this threshold might operate by exploring what properties of a visual stimulus predict a more vivid experience of vEAR, with two candidates being the learned predictiveness between a visual stimulus and an expected sound, as well as low-level physical properties of the stimulus such as the raw amount of motion energy contained within the stimulus, regardless of any meaningful content of the stimulus. It may be that an individual requires both of these conditions to be met in order to experience vEAR.

2.1.2 Objective Measures of Visually-Evoked Sounds

Our second goal was to assess the extent of the perceptual reality of the sounds that Saenz & Koch participants reported experiencing in the study. The original Sequence Discrimination Task employed by Saenz & Koch is putatively a measure of the participant's ability to recode visual sequences into auditory sequences, but was unsuitable for corroborating the synaesthetes' claims of hearing the visual events as it is ostensibly a visual task which only infers the presence of auditory sensations, rather than measuring their effect on hearing directly. Our second goal was therefore to probe the effects of visual stimulation on actual auditory signal detection, and measure the correlation of such effects with performance on Saenz & Koch's sequence discrimination paradigm. To achieve this aim we employed a second objective psychophysical task, directly measuring the effect of any visually-evoked auditory sensations on an auditory signal detection task.

In a demonstration of the ability of visual stimuli to directly modulate auditory perception, Lovelace, Stein, & Wallace (2003) reported that detection of a faint auditory stimulus was improved in the presence of a non-predictive irrelevant light, versus when the sound was presented alone. It is unclear why this effect occurs, but it may be due to the normally-occurring connections between visual and auditory cortices, meaning that excitation from the visual stimulus also increases excitability of the auditory cortex, with the effect of increasing auditory detection acuity. However, we propose this increased auditory sensitivity may be cancelled out by any vEAR sensations that accompany the presence of the visual stimulus, and thus actually impairing auditory detection in the presence of a visual stimulus in participants who experience vEAR.

Following Lovelace and colleagues we measured our participants' sensitivity for detecting a faint burst of white noise both with and without the presence of a task-irrelevant and temporally non-predictive visual event, specifically a high-contrast drifting radial grating. This would allow us to examine whether any visually-induced auditory sensations experienced either consciously or unconsciously by the participant would in any way modulate their detection of real-world auditory signals.

2.1.3 Controlling for bias/sensory dominance

Humans appear to have a natural attentional bias toward visual events compared to our other senses (Posner et al., 1976; Sinnett, Spence & Soto-Faraco, 2007). We wished to rule out the possibility that individual differences in the extent to which our participants are biased to attend toward vision rather than audition may explain our findings. Individual differences in this inherent visual dominance may bias visual sequence discrimination performance relative to auditory sequence discrimination performance, as it would lead to commensurate differences in how a visual stimulus would capture their attention relative to an auditory stimulus. Likewise on the Auditory Detection Task, high levels of visual dominance may distract the participant on audiovisual trials, on which a flash accompanies an auditory stimulus, leading to lower detection rates that are due to a failure to attend to the auditory target and thus are not attributable to individual differences in sensory processing per se.

We therefore included a measure of individual differences in the dominance of visual versus auditory modalities when a participant is required to attend to the two modalities simultaneously, following Colavita (1974). The Colavita effect is a robust, widely studied phenomenon that measures the extent to which an individual's vision dominates over another sensory modality, usually audition, when the two modalities must compete for attentional resources (see Spence, 2009). To measure this effect, participants are asked to make speeded responses to a stream of stimuli, which are primarily unimodal (typically 80% of trials are unimodal divided equally per modality e.g. Koppen & Spence, 2007a; Sinnett, Spence, & Soto-Faraco, 2007), and the remaining trials are bimodal, on which the two sets of stimuli will appear simultaneously. Response methods vary between studies, with either 2-button response, with a button per modality and requiring the two to be pressed together on bimodal trials, or a 3-button response with a dedicated third button for bimodal trials. The standard Colavita finding is that on bimodal trials participants will disproportionately neglect to respond to the non-visual component of the trial, either by pressing only the visual button (2-button paradigm) or by pressing the visual button rather than the bimodal button (3-button paradigm), therefore providing a metric of visual bias. In the present study we employed the 2 button response paradigm, in-keeping with the majority of research in this area.

2.2 Hypotheses

Stronger visual relative to auditory sequence discrimination abilities would be consistent with a higher likelihood of some experience of vEAR, given the parity of performance across modalities in Saenz & Koch's sample of hearing-motion synaesthetes. Having hypothesised that these visually-evoked auditory sensations may manifest along a continuum throughout the population, we predicted that on the Saenz & Koch (2008) Sequence Discrimination Task we would observe a broad range of performance of the visual task relative to auditory task. If, however, vEAR manifests as a dichotomy between those who do versus those who do not hear visual events, we would predict a bimodal distribution of performance on the visual sequence discrimination, rather than the broad range of performance we predicted. This hypothesis is based on the assumption that vEAR is an automatic, pre-attentive and perceptually real auditory sensation. An alternative to this hypothesis is that participants are instead using a learned cognitive strategy to exploit auditory imagery, whereby the individual deliberately and consciously recodes visual flashes into imagined mental sounds by replaying the sequence internally as an auditory record of the sequence.

These alternative hypotheses are tested operationally by measuring the effects of irrelevant flashes on the detection of faint auditory stimuli. We hypothesised that if vEAR is perceptually real, rather than a deliberate cognitive strategy as outlined above, then subjects who performed better on the visual sequence discrimination would exhibit lower detection rates for an auditory target presented at threshold when the target was presented in the presence of an irrelevant visual stimulus, compared to no visual stimulus, because the visually-evoked sound would add extra noise to the signal. If visual events impair auditory detection then this would suggest that this is a true cross-modal sensory phenomenon, rather than a learned strategy which should not impair performance on the Auditory Detection Task. Our measure of the Colavita effect was included to eliminate a possible alternative explanation, that visual events are distracting attention away from the auditory target stimulus. If this were the case we would expect the impairment of auditory detection to be greater in individuals showing stronger visual bias in the Colavita task.

We also predicted that we would find a higher prevalence rate for vEAR compared to canonical types of synaesthesia. Previous researchers have suggested that in some cases the association between inducer and concurrent is to some extent learned, via the standard Hebbian mechanisms by which an association between stimuli is reinforced (Brang, Williams, & Ramachandran, 2012; Cohen Kadosh et al., 2009; Cytowic, 2003; Grossenbacher & Lovelace, 2001; Ramachandran & Hubbard, 2001; Ward, Huckstep & Tsakanikos, 2006). A frequently cited example of evidence for this possibility is the association between exposure to letter-colour pairings in children's toys (fridge magnets, educational materials) and later development of grapheme-colour synaesthesia with these same specific letter-colour pairings in adulthood (Bor, et al., 2014; Witthoft, Winawer & Eagleman, 2015). We could argue that vEAR is distinct from other types of synaesthesia in that sound and motion are, in the real world, often highly correlated; the motion of walking is accompanied by the sound of footsteps, for example, speech with lip movements, or the impact of a moving object may result in the sound of a collision. If a degree of learned association were involved in early development of synaesthesia then we might expect that more frequent exposure to these real-world sensory pairings would lead to vEAR being more common than canonical synaesthesia, particularly those with bizarre and arbitrary sensory pairings.

2.3 Methods

2.3.1 Participants

A total of 40 naïve participants took part in the research (24 female and 16 male, aged 19-36 (mean 24.5, standard deviation 3.54) and were paid for their participation. All participants had normal or corrected vision and were screened for any neurological or hearing impairments by self-report. These participants all took part in both the Sequence Discrimination Task and the Auditory Detection Task. A subset of these original participants then completed an additional test of sensory dominance following Colavita (1974). This subset consisted of 24 participants (6 Male, 18 Female), aged 19-36 with a mean age of 24.83, standard deviation 4.92. Participants were recruited through opportunity sampling, and we neither actively included nor excluded individuals with synaesthesia from participating. All procedures were carried out after informed consent was obtained and were approved by the local Psychology ethics committee.

2.3.2 Apparatus and stimuli

The experimental procedure was conducted using an Apple Mac Mini connected to a 17” Sony HMD-A420 cathode ray tube (CRT) display. Auditory stimuli were presented through two Labtec PC speakers both positioned next to each other directly in front of and below the centre of the monitor. Video mode was 800x600 pixels with a 120 Hz refresh rate and a viewing distance was approximately 57cm (controlled using a chin rest). A small white fixation point marked the centre of the display. Responses were made using the arrow keys on a standard computer keyboard. Experimental procedures and stimuli were programmed using Psychtoolbox for Matlab.

In the Sequence Discrimination Task, the visual stimuli used consisted of white circular discs of 81cdm^{-2} luminance, presented centrally on a black background. Disk diameter was 3 degrees of visual angle. Auditory stimuli were sine wave tones with a maximum loudness of 91dBA sound pressure level (SPL) and a frequency of 360 Hz. ‘Short’ and ‘Long’ events were presented for periods of either 75 ms or 300 ms respectively, during which stimulation amplitude immediately decayed linearly from maximum to zero amplitude over a period equal to the duration of the stimulus.

In the Auditory Detection Task auditory stimuli were 300 ms in duration bursts of white noise bursts (see Figure 2-6) and amplitude was modulated by a Tukey window with 150 ms rise and fall time. This was presented over a continuous white noise background of 45dBA SPL. The visual stimulus was a grey-level radial grating, of maximum luminance 72cdm^{-2}, diameter 4deg, and frequency of four cycles per revolution, with the grating phase incremented at two cycles per second giving the appearance of a rotating windmill. This animated stimulus was chosen with the aim of amplifying any auditory sensations induced by visual motion. Response prompts and interval marker digits were displayed in white 18pt Helvetica.

Finally, in the Colavita visual dominance task (Figure 2-4), visual stimuli were circular white discs of luminance 48cdm^{-2} and diameter 4.5deg, presented centrally on a black background. Auditory stimuli were white noise bursts of 60dBA SPL. All stimuli were presented for 50 ms.

2.3.3 Design

The independent variable used in the Sequence Discrimination Task was the modality of the stimulus sequence (either visual or auditory) and the dependent variable was the subjects' discrimination accuracy at making the same/different judgement. In the Auditory Detection Task the independent variable once again was the modality of the stimulus, in this case either auditory or audiovisual, with auditory target detection accuracy as the dependent variable. In both tasks results were then coded as the proportion of correct responses versus false alarms. This allowed for analysis based on standard signal detection models (Green & Swets, 1966). We also included a measure of visual sensory dominance over audition, based on Colavita (1974) and Koppen et al., (2009) as individual differences in visual dominance could bias our results. Here the independent variable was trial modality (visual, auditory, or bimodal/audiovisual) and the dependent variable was the number of missed auditory responses on bimodal trials.

We also included a series of debrief questions, asking the subject whether they used any particular strategy to assist them on the Sequence Discrimination Task. We followed this with a series of questions to establish the extent to which they did or did not experience sensory pairings in everyday life, graded in such a wat as to build from a distanced probing down to the specific detail of the experience, in order to avoid leading the participant in one direction or another.

2.3.4 Procedure

In the Sequence Discrimination Task the procedure closely followed Saenz & Koch (2008). On each trial two successive rhythmic patterns of stimuli were presented. In half of the trials the events were all visual, and in the other half all auditory (See Figure 2-2). The modality of each trial was randomized between trials. Within each sequence, constituent stimuli (events) could be either short (75 ms) or long (300 ms) with a total of eight events per sequence. Sequences were randomly generated and consisted of a minimum of four and a maximum of five transitions (i.e. a short event followed by a long or a long followed by a short). There was an inter-event interval of 100 ms between events, and an interval of 500 ms between the first and second sequence. On half of the trials, the two sequences were identical, and on the other half they differed. In 'different' trials, the first two events and the last event were always identical between pairs, while the order of the remaining events was randomly permuted. Immediately following the second sequence, participants were required to indicate whether they thought the two sequences were same or different by pressing either the left or right arrow key on a PC keyboard, respectively. No error feedback was given as we were interested primarily in our participants' natural abilities, rather than training them to acquire a skill. The response initiated the next trial. After the final trial in each block of 20 trials, participants could take a break and the next block would begin when they pressed the spacebar. There were a total of 100 trials per session.

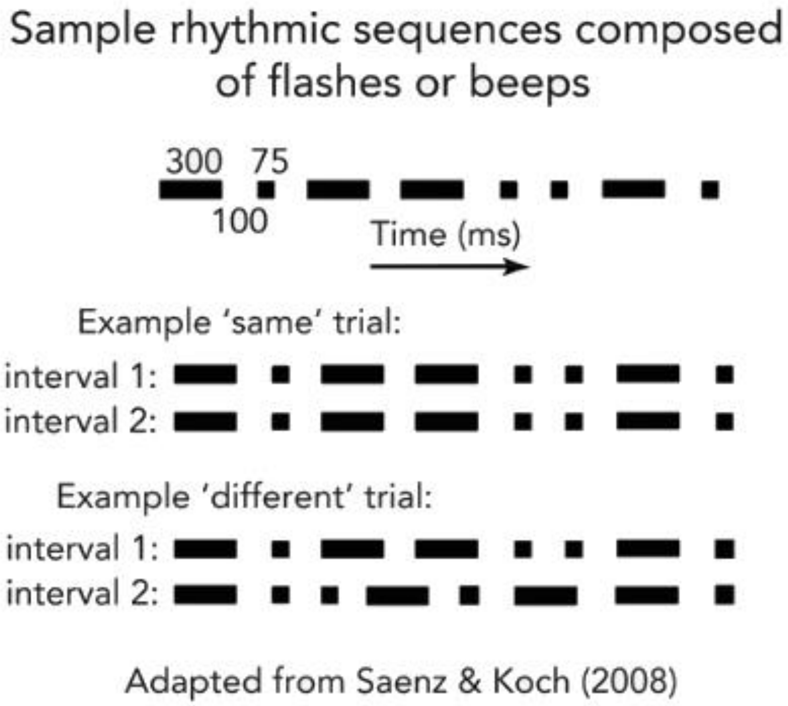


Figure 2-2: An example of a same vs. a different trial on the sequence discrimination.

Before beginning the Auditory Detection Task, we used a two-alternative forced choice staircase procedure to find auditory detection thresholds in the absence of any visual stimuli. Participants were required to detect a white noise burst in the presence of on-going background white noise. Each block began with a central fixation dot, which was followed after a key press by a central digit "1" for 500 ms, followed by a blank stimulation interval of 500 ms. This was immediately followed by a "2" for another 500 ms, then another blank stimulation interval. This was terminated by a visual prompt to make a response. The target stimulus was a 300 ms burst of white noise, which could be presented, after a delay of 150 ms, in either one of the two stimulation intervals. This procedure ensured that there was no visual stimulation present during the auditory stimulation, while providing clear visual temporal markers bracketing each stimulation interval (similar to Lovelace et al., 2003). The participant had to indicate whether the target was in either the first or second interval, using left or right arrow keys respectively. Participants heard a single click if their response was correct, and two clicks if incorrect. On each trial, target intensity was chosen depending on responses to the previous trials, using the Quest algorithm (Watson & Pelli, 1983). Each block contained 40 trials, and there were 6 blocks in total. After each block the algorithm calculated the auditory thresholds, which were then averaged to achieve a final threshold estimate. This was used to set the amplitude of the target for the main Auditory Detection Task, which then remained fixed.

The main Auditory Detection Task consisted of single-interval trials in which a target was either present or absent. Participants had to indicate whether the target was present or absent, using left or right arrow keys respectively, and received error feedback. The target timing and stimulus characteristics were the same as described before. Prior to each trial a central fixation dot was presented for 500 ms, which then disappeared at the onset of the 500 ms stimulation interval. The end of the stimulus interval was marked by the appearance of a response prompt, instructing participants to press the left or right arrow to indicate whether they thought the target was present or absent, respectively. On half of the trials the interval contained no visual stimuli (auditory-only trial), while the other half contained a white rotating radial grating that was presented throughout the 500 ms interval (audiovisual trials). Because it filled the entire stimulation interval, the visual stimulus provided no more temporally predictive information than the fixation offsets and onsets that already flanked the stimulation interval. Participants were told that the visual stimulus could be ignored because provided no information. Auditory and audiovisual trials were blocked and the order of blocks were randomly permuted for each participant. Each block contained 40 trials and there were 6 blocks in total (three for each condition).

The entire experiment took approximately 45-60 minutes for each participant to complete, with some inherent variability based upon factors such as how long each individual waited in the breaks between blocks on either task, and how long it took the experimenter to find the individual auditory threshold for each participant in the Auditory Detection Task.

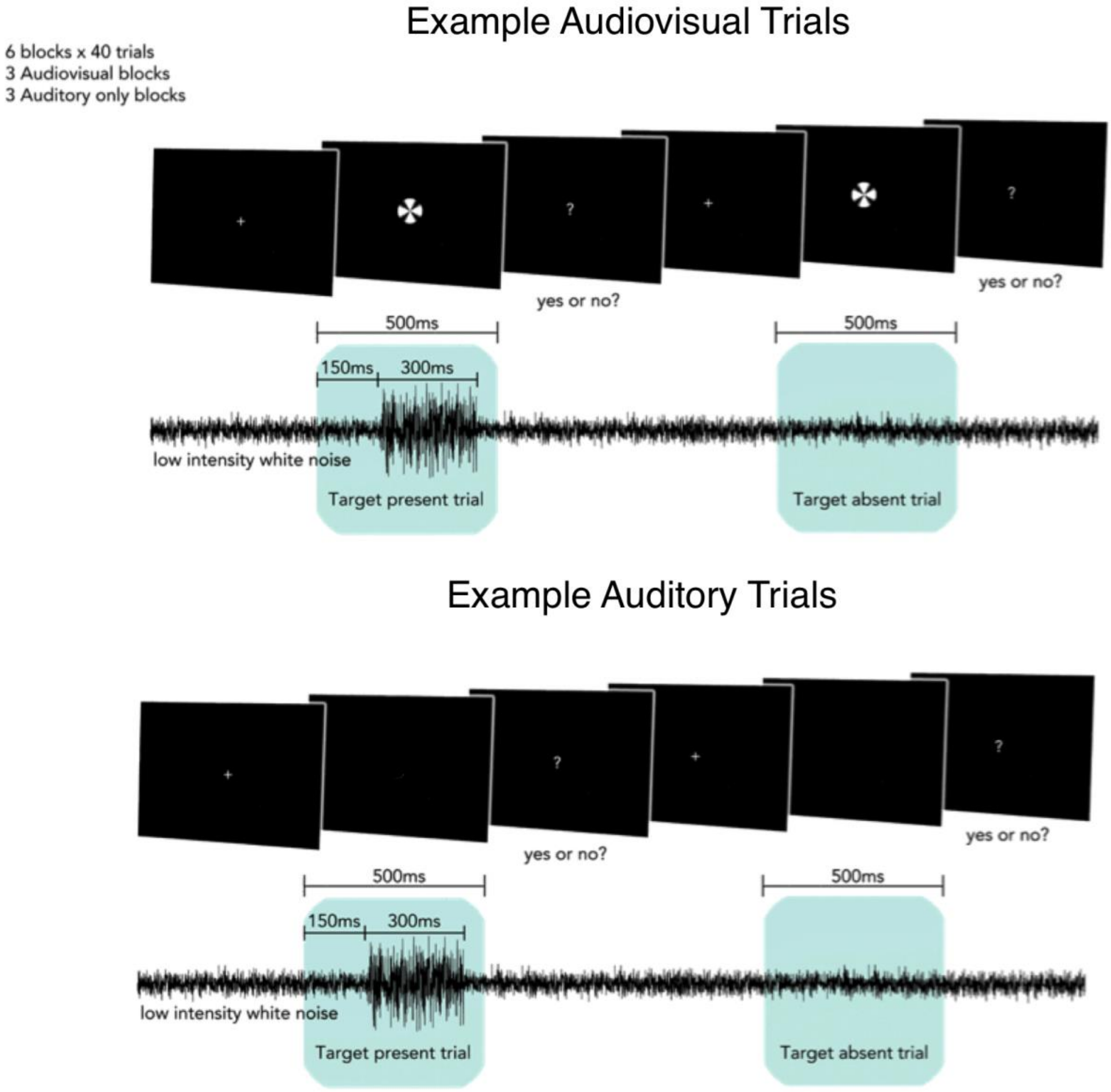


Figure 2-3: An example of an audiovisual (above) vs. an auditory (below) trial on the Auditory Detection Task.

After participants had completed both tasks we then administered a debrief questionnaire, asking them a series of short questions to gauge whether they experienced any vEAR type experiences during the task. The questions were:

What strategy were you using when you saw flashes in the visual sequencing experiment?

Were you aware of using the flashes as if they were sounds, e.g. “flash, flash-flash” = “beep, beep-beep”

Did you actually hear faint sounds when you saw flashes?

In everyday life, are you ever aware of hearing sounds when you see flashing lights or movement, e.g. shop displays, car indicators, or people walking?

Do you ever experience colours associated with letters, or with music, or tastes or smells associated with sounds?

Have you ever been diagnosed a synaesthete, or do you suspect you might be one?

In a follow-up session with 24 of our participants we conducted a test of Sensory Dominance following Colavita (1974), measuring the extent of an individual’s level of visual dominance over audition when the participant is required to attend to both modalities simultaneously. Each participant was presented with 10 blocks of 100 randomised trials which comprised of 40% visual trials, 40% auditory trials, and 20% bimodal (AV) trials. Each block consisted of a stream of stimuli, each of 50 ms duration, and each separated by a random interval of between 1300 ms and 1700 ms.

On each trial the participant was instructed to press either the left arrow key when detecting an auditory stimulus, or the right arrow key on detecting a visual stimulus. (This key combination was counterbalanced between participants). On bimodal trials participants were instructed to press both keys simultaneously. Participants were instructed to respond as rapidly as possible.

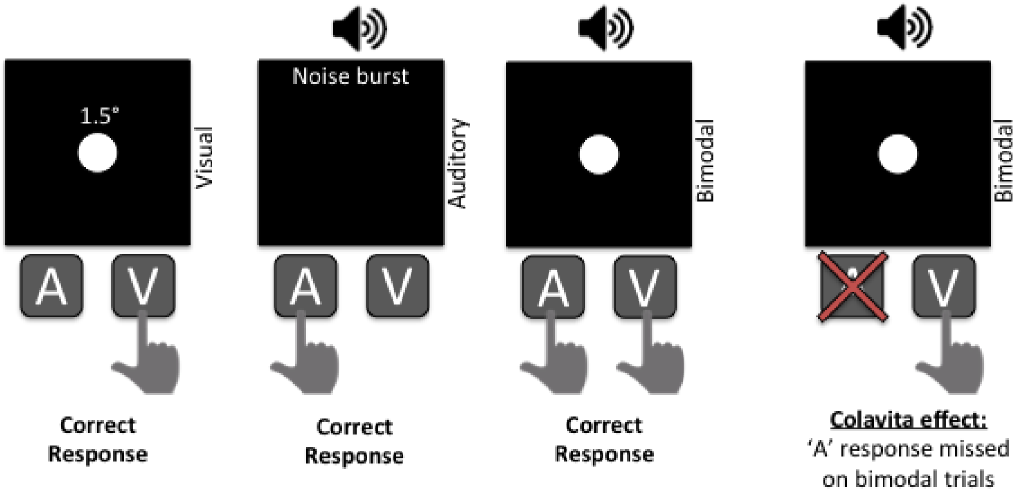


Figure 2-4: Trial types on the Colavita Visual Dominance Task. An example of a Visual, Auditory and Bimodal trial, plus the classic Colavita effect of missed Auditory responses on Bimodal trials.

2.4 Results

We removed data from one participant as their results on the Sequence Discrimination Task were at chance in both visual and auditory modality trial types. This left a total of 39 participants in the dataset.

2.4.1 Questionnaire results

We administered a debrief questionnaire after the two main tasks had been performed in order to gauge any particular strategies the participant had used to discriminate between sequences, and whether they had been or are ever aware of hearing the visual events. Two participants declined to complete the debrief questionnaire due to late running of their experimental session, leaving a total of 37 participants in the dataset.

	Used Flashes as if Beeps	Heard Faint Sound with Flash	Ever Aware of Hearing Flashes?	Do you Experience Sensory Pairings?
Sum	27	8	4	3
% (N=37)	73%	22%	11%	8%

Table 2-1: Results of the debrief questionnaire.

Results of the debrief questionnaire are depicted in Table 2-1. The first column shows percentage of participants who reported deliberately recoding the flashes and repeating them in their head as sounds. The second column shows the percentage who reported actually hearing a spontaneous sound accompanying the flashes. The third column shows the percentage who had previously been aware of hearing visual events in their daily life. The final column depicts the percentage of participants who report experiencing some kind of synaesthesia-like pairing between the senses.

In response to the first question, *'What strategy were you using when you saw flashes in the visual sequencing experiment?'*, 22 out of 37 participants (73%) reported without prompting that they consciously and deliberately attempted to convert or recode the visual sequences into internal mental sounds, imagined them as sounds or *'played them back as sounds'* in their head. 17 out of

37 participants (46%) reported trying to count and remember the positions of long versus short flashes. Some reported attempting both strategies.

In response to the next question, ‘Were you aware of using the flashes as if they were sounds?’, 27 participants (73%) reported that they were. When asked ‘In the Sequence Discrimination Task did you actually hear faint sounds when you saw flashes?’, eight participants answered said that they did. These responses were interpreted in a conservative manner, so if an individual was hesitant or unsure about whether they heard any auditory sensation on seeing the flashes, this was recorded as a negative response. We then asked ‘In everyday life, are you ever aware of hearing sounds when you see flashing lights or movement?’ and 4 participants reported that they did, despite only one of these individuals answering yes to the previous question. No participants reported that they had been diagnosed or considered themselves to be synaesthetes, although when asked ‘Do you ever experience colours associated with letters, or with music, or tastes or smells associated with sounds?’, 3 reported that they did sometimes experience phenomena like these. Of these two reported sometimes seeing colours when listening to music (in one case only particular types of heavy rock music) and the other reported that they associated numbers with particular personality traits, citing the example of the number 8 being ‘lazy’.

2.4.2 Psychophysical Data

2.4.2.1 Sequence Discrimination Task

The d’ results for the Sequence Discrimination Task were analysed in an ANOVA, in which we grouped participants by whether they reported when asked an awareness of hearing the flashes while carrying out the visual sequence discrimination. Performance was significantly better in the auditory modality [Visual Mean d’ (SE): 1.74 (3.14); Auditory: 3.14 (1.22); F(1,35)=41.8, p<.00001], in-keeping with the findings of Saenz & Koch (2008). Furthermore, overall performance was better for those participants who reported that they were aware of hearing the visual flashes [‘Yes-Responders’ Mean d' (SE): 3.02 (0.29); ‘No-Responders’: 2.27 (0.15); F(1,35)= 5.17, p=.029]. The interaction was not significant [F(1,35)=.90, ns].

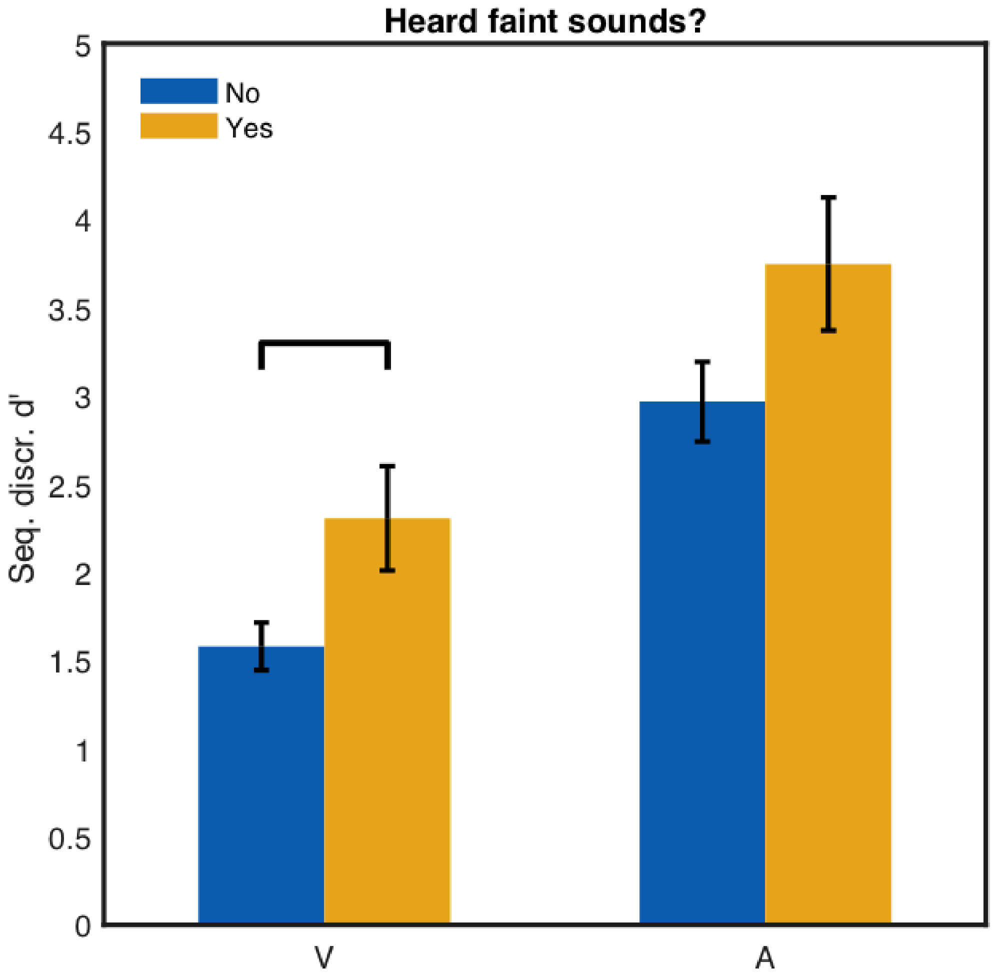


Figure 2-5: Mean d’ sensitivity for Visual vs. Auditory sequence discrimination. Grouped by responses to the question ‘Did you hear faint sounds when you saw flashes?’

We then generated Bonferroni-corrected comparisons which demonstrated that only the visual sequence discrimination benefited significantly in the participants who reported being aware of experiencing accompanying auditory sensations [Visual: t(35) = 2.43, p<0.02; Auditory t(35) = 1.64, p<0.11]. The corresponding analysis for the question *'were you aware of using the flashes as if they were sounds'* showed no significant main effect of group [F(1,35)=.48, ns]. There was however a significant interaction with task modality [F(1,35)= 5.96, p=.02], where those who answered 'Yes' had higher auditory d's than the others [Yes-Use Mean d' (SE): 3.32 (0.23); No-Use: (SE): 2.63 (0.38)], although post-hoc comparisons did not show any significant differences. There was sizeable individual variation in the extent to which a participant's auditory sequence discrimination ability exceeded their visual discrimination ability (Figure 2-6). This pattern of distribution is consistent with our prediction of vEAR manifesting as continuum throughout the sample, rather than the bimodal distribution we would have expected to see if vEAR was a dichotomous 'all-or-none' phenomenon.

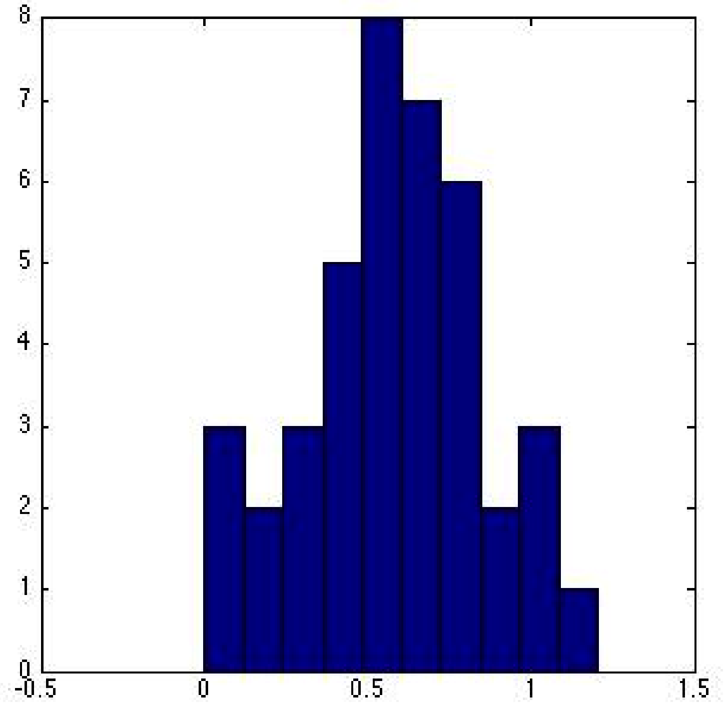


Figure 2-6: Distribution of auditory advantage (relative to visual performance) on the Sequence Discrimination Task. Scores are computed by subtracting visual sequence discrimination d’ score from auditory equivalent d’ score.

Analysis of criterion scores indicated that participants were also significantly less cautious about making ‘different’ responses to visual sequences (Mean criterion 1.67, SE 0.078) compared to auditory (Mean 2.25, SE 0.12) [t(35) = 5.6, p<0.00001]. However an ANOVA confirmed that this finding did not interact with whether or not the participant reported hearing the flashes [F(1,35)= 5.17, p=>.05, ‘Yes-Responders’ Visual Mean criterion (SE): 1.784 (0.208), Auditory Mean criterion

(SE): 2.422 (0.284); ‘No-Responders’ Visual Mean criterion (SE): 1.608 (0.083), Auditory Mean criterion (SE): 2.184 (0.132)]. Also, there was no significant main effect of hearing flashes [F(1,35) = .74, ns]

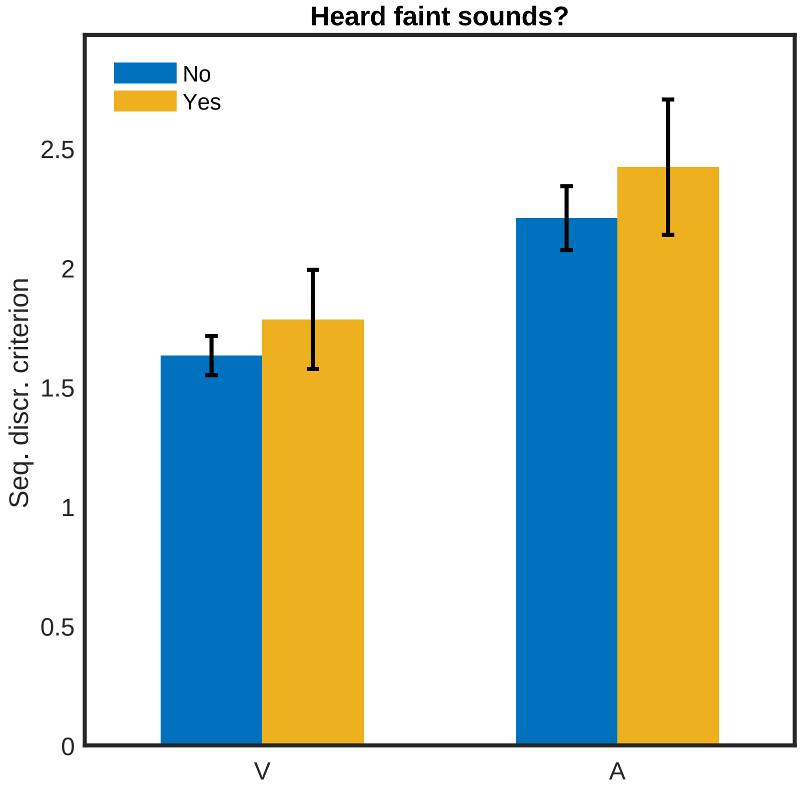


Figure 2-7: Mean criterion scores for visual and auditory sequence discrimination, by whether participant heard flashes.

2.4.2.2 Auditory Detection Task

For the Auditory Detection Task, we also found considerable individual differences in auditory detection d', as well as in the effects of the presence of the visual stimulus (Figure 2-8 B), but no significant main effect of stimulus modality [Auditory only: Mean 1.26 (.21); Audiovisual: 1.33 (0.21); [F(1,35) = 1.21, ns]. There was no significant main effect of participants' response to the question *'Did you hear faint sounds when you saw flashes'* [F(1,35) = 0.22, ns] nor interaction [F(1,35) = 0.42, ns]. There was also no significant difference in criterion scores between the audiovisual (Mean 0.076, SE 0.046) and the auditory condition (Mean 0.11, SE 0.041) [t(38) = 1.07, p=0.29].

2.4.2.3 Correlation between tasks

The 'hearing-motion' synaesthetes tested by Saenz and Koch (2008), differed less in their sequence discrimination accuracy across modalities compared to control subjects, with a putative explanation being that they were using similar resources, i.e. greater auditory temporal acuity relative to vision, to perform the task in different modalities. We hypothesised that participants showing this pattern of results would have lower auditory detection sensitivity in the presence of visual stimulation, compared to without, because any auditory sensation evoked by the visual stimulus would decrease the signal to noise ratio when detecting the auditory stimulus. In order to assess this prediction we first subtracted the auditory d' from the visual d' for sequence discrimination ($SEQd'_{V-A}$). Here higher values (less negative) indicate greater equality of sequence discrimination abilities across modalities, and behavioural results more akin to the HM synaesthetes in Saenz & Koch (2008). Our direct measure of the visual influence on auditory perception was the Auditory Detection Task. Here the results were computed by subtracting d’ in the auditory-only condition from d’ in the audiovisual condition (DETd’$_{AV-A}$). Here negative values would be consistent with an effect of the visual stimulus interfering with auditory detection.

There was a significant negative correlation between scores on these two measures [Pearson’s r(37) = -0.42, p<0.0075], indicating that the closer visual sequence discrimination ability was to auditory discrimination ability on the first task, the more visual stimulation impaired auditory detection, or the less it benefited on the second task (Figure 2-8 C). This result is consistent with the notion that internal (i.e. synaesthetic in nature) auditory noise evoked by the visual stimulus can actually interfere with an externally-originating auditory signal. We also found a significant positive correlation between DETd’$_{AV-A}$ and auditory sequence discrimination [r(37)=0.40, p=0.01], but not visual sequence discrimination [r(37)=0.01, p=0.96]. These results suggest that the visual stimulus interfered with auditory detection more in individuals whose auditory sequence discrimination ability is poorer than those with better auditory sequence discrimination abilities.

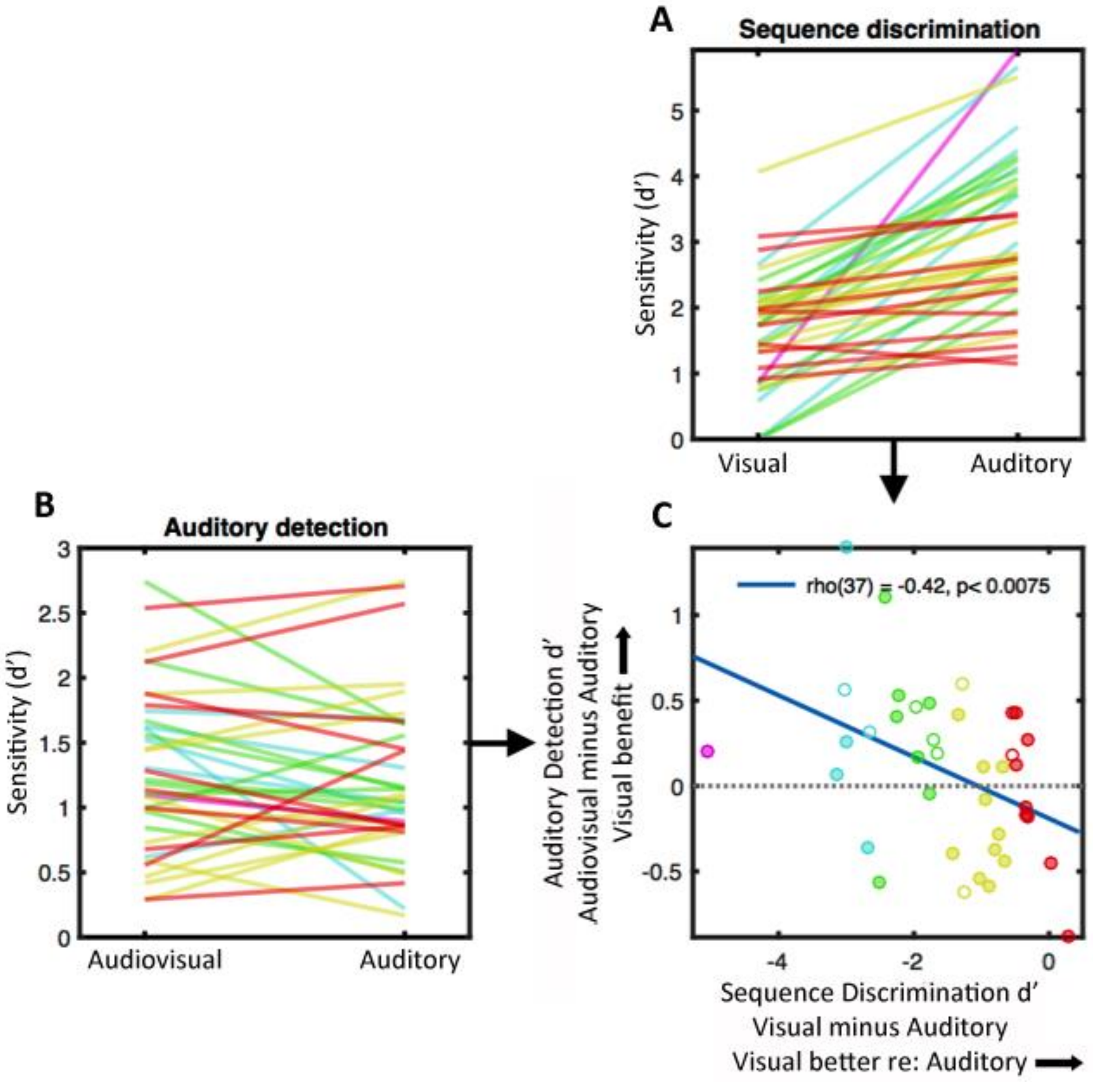


Figure 2-8: Results of sequence discrimination, Auditory Detection Tasks, with correlation.

A: Sequence discrimination sensitivity (d’) for individual participants; colour spectrum distinguishes participants with small (red) versus larger (through to blue) advantage for auditory sequence discrimination relative to visual (i.e. smaller auditory advantage is closer to performance of HMS participants in Saenz & Koch (2008)).

B: Sensitivity (d’) for auditory signal detection, in the presence of an irrelevant visual stimulus (Audiovisual) vs. alone (Auditory); colour scheme is the same as in panel A.

C: Scatterplot of the benefit of an irrelevant visual flash on d’ for auditory signal detection (AV-A on y-axis), against the advantage of auditory relative to visual sequence detection (A-V, on x-axis), with the same colour scheme as above. Open symbols represent participants who reported ‘hearing’ the visual stimuli.

The correlation between d’ difference scores could potentially arise as a result of each task’s absolute d’ measurements being positively correlated, however each subject to ceiling versus floor effects respectively. For example, if subject exhibits performance on the Sequence Discrimination Task that is near the floor, and near ceiling on the Auditory Detection Task, then better performance in one subject will widen the difference between Auditory versus Visual sequence discrimination scores, but compress the difference between Audiovisual detection and Auditory detection scores, resulting in an apparent negative correlation across subjects. However there was no significant positive correlation between absolute measures (averaged across visual and auditory conditions for each task) [r(37)=-0.12, p< 0.47], so we discount this as an explanation of our findings.

We then repeated the above analysis but using criterion scores instead of d’ for the detection task, in order to establish whether these results reflect true sensory experiences or differences in response characteristics. Visual sequence discrimination d’ positively correlated with $\mathrm{DETc}_{\mathrm{AV\text{-}A}}$ [r(37)=0.35, p=0.03]. This indicates that participants with higher visual sequence discrimination abilities were more cautious when responding in the Auditory Detection Task in audiovisual trials compared to purely auditory trials. There are two potential explanations for this finding. First, there may be a tendency in these participants to discount auditory signals that co-occur with internal visually-evoked sounds, meaning that they might mistake the real sound for an internal sound and thus report that they heard nothing. Alternatively, it may be due in part to some form of bias towards vision when the two senses are forced to compete for attentional resources.

2.4.2.4 Colavita Sensory Dominance Task

One potential explanation for the negative correlation described above could be individual differences in the extent to which an individual is biased toward attending to visual stimuli over auditory stimuli when both are stimuli are competing for attentional resources. If an individual is more biased toward visual events we would expect them to have an advantage on the visual sequence discrimination relative to the auditory sequence discrimination compared to individuals who do not display such a large visual bias. This would be beneficial when the visual stimulus is relevant, as in the visual sequence discrimination, but conversely would be detrimental on the audiovisual trials of the Auditory Detection Task, when the irrelevant visual stimulus is merely a distraction to which their attention is drawn more than individuals with less of a visual bias. To address this possibility we included a measure of visual dominance over audition following Colavita (1974). The visual bias hypothesis outlined above would predict that scores on this measure should correlate with those on our sequence discrimination and Auditory Detection Tasks. The classic ‘Colavita effect’ is that, when required to make speeded responses to either visual, auditory, or bimodal stimuli, participants will disproportionately fail to detect (or respond to) the auditory component of the bimodal trials, responding only to the visual component on said trials. The extent to which these specific errors are made by any given subject is a measure of their ‘bias’ toward vision over audition. In our subset as a whole we observed this effect, with errors on bimodal trials neglecting the auditory component twice as frequently as the visual component was neglected [Visual error rate: M=10.8%, SE=1.9; Auditory: Mean 5.1%, SE .11; t(23)=3.46, p =.002]. Crucially however, there was no significant correlation between visual bias and our difference measure of visual interference in auditory detection ($\text{DETd'}_{AV\text{-}A}$) [r(22)=-0.34, p=0.10], nor was there a correlation with visual bias and visual sequence d' [r(22)=0.05, p=0.81]. There was no significant correlation between $DETc_{AV\text{-}A}$ and auditory sequence discrimination [r(37)=0.14, ns], or with $SEQd'_{V\text{-}A}$ [r(37)=0.12, ns]. We then performed a multiple regression analysis, which confirmed that the relationship between sequence discrimination and auditory signal detection ($SEQd'_{V\text{-}A}$ vs. $DETd'_{AV\text{-}A}$) was still reliable after controlling for sensory dominance [t(21) = 2.46, p= .022, R^2= 3.1], which itself was not found to be significantly predictive [t(21) = 1.78, p= .091]. The lack of a correlation between the Colavita effect and our other measures should be interpreted with some caution, given the small subsample who completed this measure it is possible that we lacked sufficient power to detect any effect, and also because our use of Colavita to measure this type of visual bias is somewhat speculative. It may be that other measures may be more sensitive to any association between attentional bias toward visual stimuli and visually-evoked auditory sensations.

2.5 Discussion

These results provide the first estimate of the prevalence of the visually-induced auditory sensations described by Saenz and Koch (2008), thus building considerably on their findings. We also offer the first evidence that these sensations are perceptually 'real' enough to impair detection of real world auditory signals. 22% of our participants reported hearing mental sounds when viewing the visual flashes on the Sequence Discrimination Task. These participants performed significantly better in a visual element of the task compared to those who did not report experiencing vEAR, in line with previous suggestions that the ability to recode temporally dynamic visual events as sounds may influence visual discrimination of rhythmic sequences (Guttman, Gilroy & Blake, 2006; Saenz & Koch, 2008). Our findings are the first to suggest that this auditory-recoding of visual sequences not only indirectly affects visual performance but that it also directly affects auditory signal detection ability, suggesting that vEAR can interfere with normal hearing.

Our finding that some people spontaneously recode temporally complex visual stimuli into internal sounds is not without precedent. Of course Saenz and Koch report this ability in their small sample of synaesthetes, but prior to that Guttman, Gilroy & Blake (2005) reported a very similar phenomenon, asking participants to make a same/different judgement on rhythmic visual sequence, noting "*a natural tendency to hear the temporal sequencing of these [visual] changes*" (p.2). They then performed several experiments to see how visual vs. auditory distractor stimuli affect visual sequence discrimination. Results showed that simultaneously presented sounds that were incongruent with visual rhythms disrupted task performance, while congruent sounds enhanced performance, indicating that task performance was reliant on auditory rather than visual information. Task irrelevant changes in the visual sequences (e.g. varying the contrast levels across the visual stimulus train) only slightly impaired task-performance indicating that performance primarily relied on the temporal structure that was recoded into auditory information. These results suggest that participants did mentally recode the visual sequences in auditory format, as incongruent auditory information significantly impaired visual task performance despite being irrelevant to the visual task. Furthermore they then compared the presentation of task-irrelevant auditory information with the first visual sequence (i.e. at encoding) versus during the second visual sequence (i.e. during the comparison/retrieval process). Results showed that the greater disruption occurred when the distractors were presented at encoding, again suggesting that this process is characterised by a recoding of visual into auditory information.

However, both Saenz and Koch (2008) and Guttman, Gilroy & Blake (2005) only measured the effects of audition on visual sequence discrimination performance. The former used an indirect test of putatively synaesthetic sounds on visual performance, while the latter explored how real-world sounds impaired auditory recoding of these visual sequences. Neither explored how these visually-evoked sounds affected hearing. We hypothesised that if there is such a natural tendency in some individuals for visual flashes to evoke an auditory response, then these should impact on the detection of real sounds. Our results confirmed this: participants who showed similar visual and auditory sequence discriminability that resembled Saenz & Koch's synaesthetes tended to have poorer auditory detection sensitivity when accompanied by irrelevant visual flashes. To our knowledge this is the first study to corroborate the existence of these visually-induced auditory percepts against real auditory stimuli, rather than simply implying their existence indirectly with indirect visual measures.

The results of the present study support the hypothesis that it is this natural and spontaneous ability to hear visual rhythms, rather than deliberately attempting to consciously adopt recoding as a strategy, on which enhanced visual sequence discrimination ability depends. In Saenz & Koch's (2008) study, 'hearing-motion' synaesthetes who reported routinely hearing visually-evoked sounds were almost as good at discriminating visual flash sequences as they were discriminating auditory tone sequences, while in non-synaesthetes sequence discrimination was significantly poorer for flashes than for tones. It was inferred from this that the synaesthetes were uniquely able to recode the flashes as sounds and thus benefit from the better temporal acuity of the auditory modality (Glenberg et al., 1989; Guttman, Gilroy & Blake et al., 2006). In the present study some participants did report a deliberate attempt to consciously recode visual sequences into sound, however, those who did so did not show the same improved visual sequencing as those for whom the auditory sensations were spontaneously perceived. Our results suggest that vEAR is an involuntary experience as it is seemingly present whether it presents an advantage, as in the Sequence Discrimination Task, or a disadvantage, as on the Auditory Detection Task. This spontaneous automaticity is also one of the defining criteria of synaesthetic percepts.

Our subjective measures in the form of a debrief questionnaire support our assumption that a proportion of our sample were indeed spontaneously recoding visual events into sound. Despite not actively recruiting synaesthetes, 22% of our participants reported when asked that they had heard faint sounds accompanying the visual flashes, and as with Saenz and Koch (2008) these same participants displayed significantly better visual sequence discrimination compared to those who said they did not hear the visual flashes. There is an interesting discrepancy between the 22% prevalence of participants who reported being able to hear the visual sequences in the lab versus the lower 11% prevalence of individuals who reported that they had been aware of the ability to hear visual events in the past. There are several possible explanations for this discrepancy. First, it may be that the prevalence figure was slightly inflated by a response bias, with participants giving the experimenter the answer that they wanted to hear. However we were particularly careful not to lead the participants to a particular answer, and more importantly, if this were the case then we would not expect this 20% of participants to significantly outperform the rest on the measure of visual sequence discrimination ability. A second potential explanation for the discrepancy is that some of the participants may only have become aware of the visually-evoked auditory sensations in the lab, as it is not often that one sits in a dark silent room watching silent visual flashes for forty minutes. Furthermore, if one is naturally inclined to hear these faint auditory sensations and presumably has been so inclined since birth, then they may be so unremarkable as to remain unnoticed until the experimenter inquired about them. It may also be the case that the auditory sensations vary in intensity between individuals, and what distinguishes the 11% who routinely hear flashes from the larger 22% who were aware of them in the lab is that these are the ones for whom vEAR is particularly vividly perceived. This explanation would be consistent with our hypothesis that vEAR may manifest as spectrum from those who hear nothing at all, to those who may become aware of the phenomenon under apposite circumstances, to those who are constantly aware of the sensations at all times. In any case, even this lower prevalence rate of 11% is still considerably higher than the estimated prevalence rates of other types of synaesthesia, which vary between 2-4% of the population.

Although our subjective data does predict some aspects of participants' objective performance, these objective markers of visually-evoked auditory sensations (i.e. increased visual sequence discrimination relative to auditory sequence discrimination, and impaired auditory detection in the presence of a visual event) could also occur without the individual reporting any conscious awareness of hearing flashes. Other dissociation of this kind between subjective and objective measures of synaesthesia have been reported in efforts to induce synaesthesia through training, which have shown robust objective effects on information processing but little evidence of that the subject is perceiving any subjective concurrent sensation (reviewed in Deroy & Spence, 2013b). We can highlight two methodological differences between our Sequence Discrimination Task and that of the original Saenz and Koch (2008) study that could potentially explain this partial dissociation of subjective and objective measures outlined above. Firstly, we employed random sampling while the original contrast between presumably self-reporting synaesthetes and a cohort who did not report experiencing any form of audiovisual synaesthesia. It is therefore conceivable that this control group may have contained individuals who experience low-level vEAR without conscious awareness, and had not been made aware of it through introspection as described above (N.B. the original Saenz and Koch paper did not include the subjective debrief questionnaire that we employed). Secondly, other differences between our stimuli and those used in the original paradigm, such as slightly longer events and shorter retention delays between intervals, may have been more sensitive to latent vEAR tendencies found in neurotypical participants.

Given that the results of our objective and subjective measures are at least partially dissociated, one might ask whether the two are both being jointly influenced by some other factor other than cross-modal visual-to-auditory sensations. One such candidate might be a bias in certain participants to attend to visual stimuli over auditory stimuli. Here visual sequences would be more memorable or attention-capturing than auditory sequences, while on the Auditory Detection Task a visual flash may distract the participant away from the auditory target. We specifically included a measure of the Colavita effect (Colavita, 1974) that is putatively capable of measuring such a bias. While we did replicate the Colavita effect in our sample, this was not predictive of scores on either of our other measures, and we therefore do not believe that our findings can be attributed to individual differences in visual attentional bias.

The high prevalence of vEAR that we report here is consistent with the hypothesis set out in chapter 1 section 1.4.4 that certain synaesthetic sub-types may occur more frequently if the inducer stimulus and the concurrent sensation are highly correlated and frequently occurring in the natural world. This has support from previous evidence to suggest that synaesthetic pairings can be reinforced by repeated early years exposure (Bor et al., 2014; Witthoft et al., 2015). Unlike some of the more unusual sensory pairings experienced by some synaesthetes, visual and auditory events are frequently co-occurring and therefore highly predictive of one another in the world around us, with examples such as speech sounds and lip movements, footsteps and stride, objects colliding just a few examples of when a motion would predict an accompanying sound. As a consequence if an individual has some natural disposition toward synaesthesia, these common audiovisual associations are likely to manifest more often than rare pairings such as graphemes and colours. However, one might well ask why, if vEAR is so common, are there so few reports of these or other visually-evoked auditory concurrents (such as Baron-Cohen et al., 1996; Goller et al., 2009; Saenz & Koch, 2008). This may counter intuitively be precisely because of the same highly predictive nature of visual motion and accompanying sounds outlined above. For example, when a faint auditory sensation occurs with an accompanying visual event it would not be out of place, as these sensations co-occur very frequently, whereas a visually-evoked taste or smell would be more striking. If as we suggest, vEAR may exist along a continuum it is possible that those who experience ‘weak’ vEAR may not be aware of their condition as that their synaesthetic concurrents may not be of sufficient intensity to be consciously perceived. Alternatively, some of those with vEAR may have learnt not to attend to these sensations, either erroneously assuming them to be real phenomena, such as incorrectly believing they can hear the footsteps of a person walking ahead when in fact the sound exists only in their mind, or with the ‘synaesthetic’ sound of a footstep ‘drowned out’ by the real sound of the foot hitting the floor. It should not be surprising then that some of our participants who experienced vEAR had not been aware of these sensations before their attention was drawn to them under controlled laboratory conditions, as others have synaesthetes have done before when engaged in introspective attention (Tyler, 2005). Over the course of my PhD I have even noticed myself experiencing auditory sensations triggered by visual events, such as flashing lights and quite vividly once while driving. The road was lined with trees and the evening sun coming in from the side, meaning the shadows of the trees came and went in quick succession, and I was aware of a distinct fluttering noise as this happened. Whether this was the result of increased awareness and introspection, or a learned response to three years of exposure to our sequence discrimination paradigm I do not know.

2.6 Summary

In summary, in this chapter we have presented for the first time evidence that a remarkably high proportion of randomly-sampled neurotypical individuals reported experiencing visually-evoked auditory sensations. We show that these auditory sensations can be sufficiently perceptually real to impair detection of real-world sounds and can benefit visual sequence discrimination abilities. This higher prevalence rate as compared to other comparable phenomena may be due in part to the increased exposure to audiovisual co-occurrence in the natural world, which leads to an increased statistical association and thus predictability of a sound from dynamic visual events. The continuous distribution of performance across individuals that we see on our objective measures seems to suggest that this is a common ability and therefore is unlikely to be the result of particularly abnormal or atypical neurophysiology. In the next chapter we turn our attention to the brain mechanisms that may underlie the experience of vEAR.

Experiment 2: Using Transcranial Alternating Current Stimulation to Explore the Neurophysiological basis of vEAR

Chapter 3:

This chapter examines two plausible neurophysiological mechanisms that may underlie vEAR. The first of these is that in individuals who experience vEAR have greater than average levels of intercortical connectivity between the auditory and visual cortices. The second is that the normal connections between auditory and visual cortices are largely inhibited in the general population but disinhibited in those who experience vEAR. We explore these two hypotheses by using Transcranial Alternating Current Stimulation (TACS) to modulate the primary visual cortex versus the primary auditory cortex while participants perform the Sequence Discrimination Task outlined in chapter 2. By comparing performance under each stimulation condition and task modality in those who do versus do not experience vEAR we discuss the implications for the two potential mechanisms outlined above and suggest that disinhibition of audiovisual connections is likely the more plausible explanation.

3.1 Introduction

In the previous chapter we saw that an irrelevant visual flash can impair auditory signal detection in subjects who display patterns of visual versus auditory sequence discrimination ability behaviourally similar to the hearing-motion traits described by Saenz and Koch (2008). We also saw that this effect cannot be explained by individual differences in a subject's level of visual attentional bias relative to audition. We now begin to explore the underlying neurological differences between those who report experiencing vEAR compared to those who do not. Here we use Transcranial Alternating Current Stimulation (TACS), a method of noninvasively modulating the electrical activity of the brain by focally applying a weak current over a particular point on the scalp. Our key questions were: if vEAR depends on crosstalk between visual and auditory cortex, does TACS modulate these interactions, and if so, does this modulation depend on the individual propensity to experience vEAR? This second question is of particular importance as it begins to establish the possibility of cortical and/or connective differences in high versus low scorers on our measure of hearing-motion abilities, and thus a potential mechanism to explain vEAR.

At present we know little in the way of the characteristics of the types of people who experience vEAR, nor the types of trait that correlate with it. However, based on our assertion that vEAR may be more common than other types of synaesthesia due to the high degree of correlation between movement and sound in the natural world we propose that individuals who are particularly attuned to these two modalities might be particularly susceptible to developing vEAR. One such candidate demographic would be highly trained musicians, who from an early age will have spent considerable amounts of time rehearsing, in which they must read the physical cues of their fellow musicians and their conductor as they play. In this environment each movement will be highly predictive of an accompanying sound, for example as the violinist extends and retracts their bow arm the instrument will sound, and as a timpani player brings their arm down we will expect to hear a thud. Having hypothesized that these individuals might be prone to vEAR, we included in our sample a number of musicians recruited from the Royal College of Music, in order to maximise our chances of finding a sufficient number of participants who experience the phenomenon of vEAR. We will later return in chapter 4 to the topic of other personal traits that may correlate with experience of vEAR.

3.1.1 Cross-activation versus Disinhibition as a Mechanistic Underpinning of vEAR

There are several models that attempt to explain why synaesthetes experience the distinctive sensory and cognitive crossovers that are hallmarks of the condition. Two potentially pertinent models in the case of vEAR would be the cross-activation model (Ramachandran & Hubbard, 2001) and the disinhibited feedback model (Grossenbacher & Lovelace, 2001).

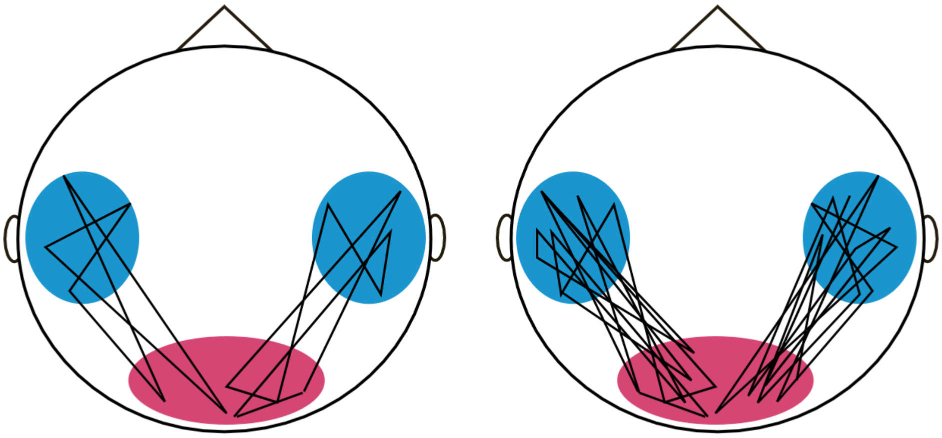


Figure 3-1: The Cross-activation Model. This suggests that vEAR is the result of an excess of cortical connections between visual and auditory areas.

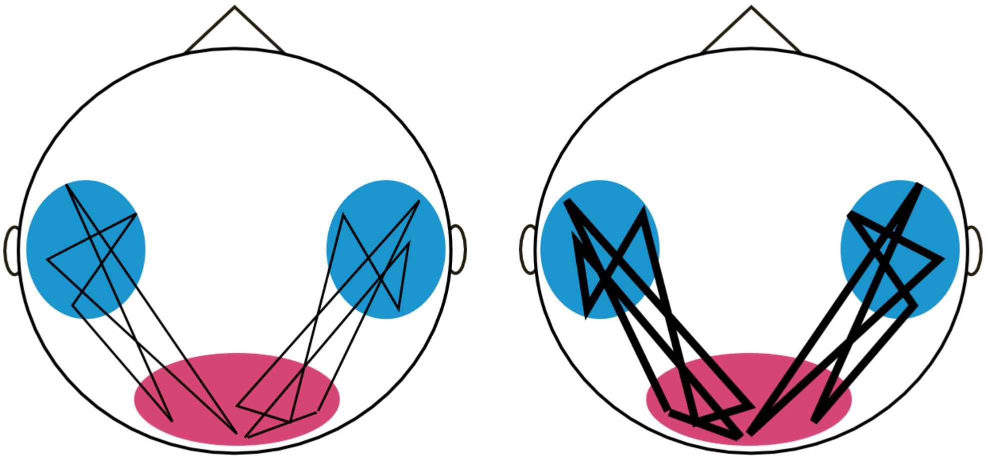


Figure 3-2: The Disinhibition Model. This suggests that vEAR is the result of greater levels of disinhibition (thickness of line) in otherwise normal connections between visual and auditory areas.

The former suggests that synaesthesia is the result of excessive cortical connections between the brain areas that represent the inducer and the concurrent associated by the synaesthete (See Figure 3-1), while the latter proposes that it is not the overall levels of connectivity but an atypical activation of the concurrent-representing brain regions as a result disinhibited feedback from multisensory areas such as the parietal cortex (Neufeld et al., 2012) (See Figure 3-2). Disinhibition may occur through abnormal feedback within the sensory processing network, e.g. from the anterior fusiform gyrus back to posterior inferior temporal regions and V4 (Grossenbacher and Lovelace 2001). Smilek et al. (2001) suggested that higher brain areas such as the superior temporal sulcus or the temporo-parietal occipital junction, might mediate such disinhibited feedback, allowing activation to return to lower sensory areas.

In some synaesthesia sub-types such as grapheme-colour synaesthesia, the inducer and concurrent representations are directly cortically adjacent (Brang, Hubbard, Coulson, Huang, & Ramachandran, 2010; Hubbard, Arman, Ramachandran, & Boynton, 2005; Wade et al., 2002). This direct proximity has led some to conclude that the cross-activation hypothesis explains the mechanisms for grapheme-colour synaesthesia (Ramachandran & Hubbard, 2001). Neuroimaging studies appear to support this account with neighbouring cortical areas that represent colour (e.g. V4) and graphemes (e.g. Brang, Hubbard, Coulson, Huang, & Ramachandran, 2010). Some have suggested that decreased synaptic pruning between these adjacent areas after the critical developmental period may explain the later development of synaesthesia (Hubbard, 2007). However, grapheme-colour synaesthesia also appears to be characterised by feedback from the parietal cortex (Rouw & Scholte, 2007, 2010; van Leeuwen, Petersson, & Hagoort, 2010). Taken together these studies indicate that both cross-activation and disinhibition of feedback may both play a role in the aetiology of grapheme-colour synaesthesia, which has led to the development of a hybrid of the cross-activation and disinhibition models known as the Re-entrant model (Hubbard, 2007), which attempts to link the local and distal cortical connections that may underlie synaesthesia.

There have been numerous attempts to identify physiological differences in the brains of synaesthetes relative to control subjects. These include differences in grey matter volume, white matter volume, and the ratio between the two as well as connections between brain regions. However this line of research is substantially constrained by the enormous heterogeneity of synaesthesia, and efforts to date have largely focused on grapheme-colour synaesthesia.

Diffusion tensor imaging (DTI) is a method of magnetic resonance imaging that allows the mapping of water molecule diffusion, and has been used to map white matter connectivity in the brain. To date only a handful of studies have used DTI to explore the brain structure of synaesthetes (e.g. (Hänggi et al. 2008; Jäncke et al. 2009; Rouw and Scholte 2007). Rouw and Scholte (2007) studied structural connectivity in grapheme-colour synaesthetes using DTI, specifically to measure fractional anisotropy (FA) which quantifies the directionality of cortical connections One key finding was increased FA in the white matter proximal to the fusiform gyrus, a brain area involved in colour perception (McKeefry & Zeki 1997), in the synaesthetes relative to controls. The fusiform gyrus is adjacent to neighbouring areas that are specialised in grapheme recognition (Cohen et al. 2000). This would appear to provide evidence for the cross-activation theory of synaesthesia discussed above. Rouw and Scholte (2007) also reported that this increased white matter volume in right inferior temporal cortex was more pronounced in projector synaesthetes relative to associator synaesthetes. In addition, there was greater connectivity in the superior parietal and frontal cortex in synaesthetes.

Jäncke et al. (2009) also explored neuroanatomical differences between grapheme-colour synaesthetes and controls using surface-based morphometry to measure cortical thickness, volume, and cortical surface area, as well as FA measures of white matter coherence. Synesthetes had on average increased thickness, volume and surface area in the grey matter of both the left and right fusiform gyrus and adjacent regions, and increased colume specifically in V4, relative to controls. Structural connectivity as measured with FA was greater surrounding the fusiform gyrus in grapheme-colour synesthetes relative to controls.

These two studies appear to demonstrate hyper-connectivity in specific neighbouring brain regions in grapheme-colour synaesthetes compared with non-synesthetes. This does not negate the possibility that structural differences are also complemented differences in function connectivity. Such functional differences are likely to follow given the structural differences between synaesthetes and controls reported above, and it is possible that these structural differences may be the result of long-term differences in functional connectivity (e.g. Bezzola et al. 2011; Loui et al. 2011; Mackey, Whitaker, and Bunge 2012).

In the present study we aim to examine how the mechanisms of cross-activation and disinhibition may potentially explain the phenomenon of vEAR. We propose two potential explanations, the first of which is that inter-participant variability in vEAR is explained primarily by individual differences in the local cross-activation of the visual and auditory cortices. Under this account greater white matter tracts between the two regions could bring about auditory sensation when viewing visual events. The second hypothesis is that feedback from V1 back to A1 is typically inhibited (e.g. Iurilli, et al., 2012; Mattingley, et al., 1997), but disinhibition occurs in individuals who experience vEAR, leading to visual flashes evoking a response in both visual and auditory cortical areas.

There are very few reports of visual motion inducing auditory qualia, beyond Guttman, Gilroy & Blake (2006), Saenz and Koch (2008), and our results outlined in chapter 2 (Fassnidge, Cecconi-Marcotti & Freeman, 2017). There are, however, examples of the inverse of this phenomenon. Scheef et al. (2009) recently provided evidence that V5, a region known for processing visual motion, can be modulated by auditory input, suggesting a multisensory role for this area. The authors artificially paired biological motion stimuli with either an auditory stimulus derived from the visual motion through a process they call ‘sonification’ (concordant condition) or a discordant auditory stimulus. Results as assessed via fMRI showed that V5, a primarily visual region, was activated more in the concordant auditory stimulus condition than in the discordant condition. This may suggest that V5 receives input from auditory areas when the auditory signal carries information pertinent to the visual event. In addition, Bueti & Macaluso (2010) found that activity in visual areas, known to respond to bodily motion, is modulated by auditory stimuli that are predictive of visual motion, such as the sound of a hand clapping. If this type of crosstalk between visual and auditory cortices is bidirectional then it could potentially account for the experience of vEAR.

3.1.2 Transcranial Electric Stimulation

There is a rich history of the scientific application of electric currents to modulate brain activity (see Priori, 2003 for a review). One modern branch of this research encompasses a variety of noninvasive techniques collectively known as Transcranial Electric Stimulation (TES), in which two or more electrodes are placed on the surface of the scalp (or elsewhere as a reference point) and a current is then run through these electrodes. An example set-up can be seen in Figure 3-3.

Figure 3-3: An example stimulator as used in the present experiment. Image reproduced from http://www.neurocaregroup.com

This may be an alternating (AC) or direct (DC) current (See Figure 3-4), or a randomly fluctuating current within a specified spectrum, with these techniques known as Transcranial Alternating Current Stimulation (TACS), Transcranial Direct Current Stimulation (TDCS), and Transcranial Random Noise Stimulation (TRNS) respectively (Paulus, Nitsche, & Antal, 2016). Here we will focus primarily on TACS, the method employed in the present study. Unlike in TDCS, the electrodes during TACS each alternate between acting as the anode and the cathode once per half cycle of each oscillation (assuming a sinusoidal waveform). This reversal means that the net current delivered per cycle is zero, and thus will not affect neuronal membrane potential in the stimulated cortical region (Paulus, Nitsche, & Antal, 2016). This pattern of depolarization and hyperpolarization, assumed to be induced in neurons affected by these currents, is sufficient to entrain the stimulated cortical region into the oscillatory frequency set by the stimulation (e.g. Helfrich et al., 2014; Vosskuhl, Strüber, & Herrmann 2015).

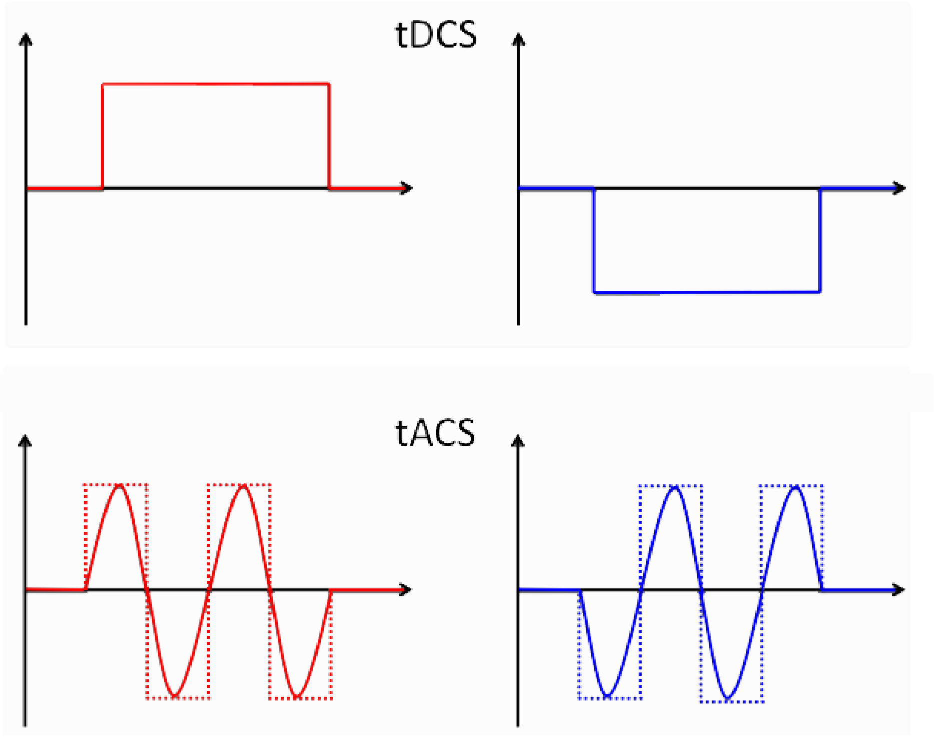


Figure 3-4: Different stimulation paradigms.

During TDCS (top) a direct current is applied through one positive (anode) and one negative (cathode) channel, while with TACS (bottom) an alternative current is applied. This may be either sinusoidal (solid line) or rectangular (dotted line), with the two channels in counter-phase (from Herrman et al., 2013).

In recent years the physiological effects of TACS have been the subject of some controversy. Efforts have been made to study them in vivo through the use of intracranial recordings in animals. For example, Fröhlich and McCormick (2010) applied intracranial stimulation to live ferrets while simultaneously recording local field potentials (LFPs) and multiunit activity (MUA). Prior to stimulation onset neuronal firing in MUAs was synchronized to the oscillatory frequency of the LFPs (See Figure 3-5, left). Cortical slices were then taken and applied with in vitro stimulation, with simultaneous MUA once again being recorded. This revealed that even a weak sinusoidal current of less than 0.5 V/m was sufficient to entrain firing patterns in the MUAs (Figure 3-5, right).

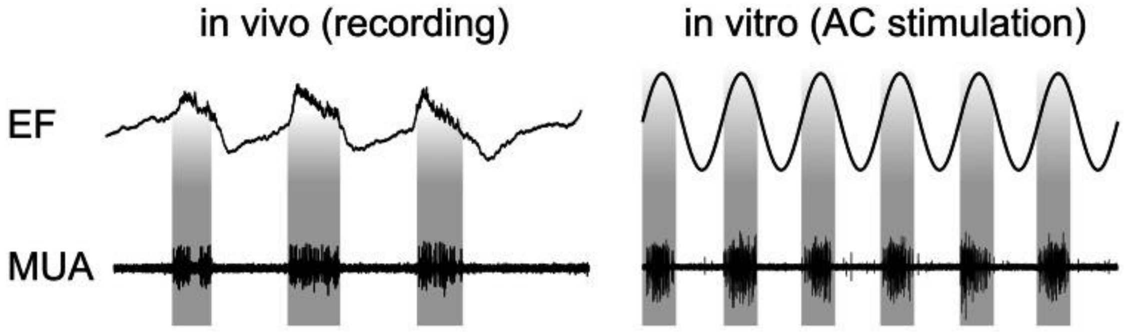


Figure 3-5: Physiological mechanisms of TACS.

Left: Recordings taken *In vivo* in ferrets displaying spontaneous neuronal activity in Multiunit Activity (MUA) can synchronize with phase of local electric field (EF).

Right: Slices of cortex stimulated electrically with a sinusoidal current also led to synchronization.

(Cited by Herrmann et al., 2013, adapted from Fröhlich and McCormick, 2010).

Building on this in vitro work, Ozen et al. (2010) examined whether such weak currents as those used in TACS studies were actually sufficient to penetrate the skull, and if so was this in turn adequate to affect cortical activity. Ozen and colleagues replicated the effect of TACS observed in humans using rats by stimulating the skull's surface with electrodes while simultaneously recording intracranial neural activity. Their results confirmed that even a current of ~1 V/m was able to synchronize neuronal firing with the peak of the extracranial sinusoidal current.

3.1.2.1 Modulation of Perception using TACS

If then TACS can modulate the electrical activity of the cortex, is this effect sufficient to modulate perception in any detectable way? Work by Cecere, Rees and Romei (2015) suggests that it can. They examined the effect of TACS on the double flash illusion, in which a single flash of light may be perceived as two successive flashes if accompanied by two rapid auditory stimuli (Shams, Kamitani & Shimojo, 2000). This exploits one of the fundamental rules of multisensory integration, that temporally congruent stimuli will be perceived as sharing a common origin (King & Palmer, 1985; Meredith, Nemitz & Stein, 1987; Stevenson et al., 2012) and are thus perceptually bound. The temporal window between the two auditory stimuli required for the double flash illusion to occur is small, around <100 ms, although larger in some clinical groups such as those with autism spectrum disorder (Foss-Feig et al., 2010) or schizophrenia (Haß et al., 2017). Outside of this binding window, successive auditory stimuli do not induce the illusion of concomitant flashes. There is evidence to suggest that this window in which the sounds must occur in order for the illusion to be perceived is driven by individual differences in alpha frequency in the occipital lobe (Cecere, Rees, & Romei, 2015). Cecere and colleagues explored whether changes in TACS stimulation frequency are able to modulate the temporal window in which the double flash illusion is maximally perceived. Theorising that the alpha oscillatory cycle may be a candidate for the mechanism for defining this binding window they applied TACS to participants at ±2Hz their individual alpha frequency (which had been established beforehand with EEG). The effect was to reduce or increase the window in which the illusion was most often perceived in line with this TACS-driven change in alpha oscillation. This clearly demonstrates that TACS can entrain oscillatory patterns in the human brain that result in perceptual changes at the individual level (see Figure 3-6).

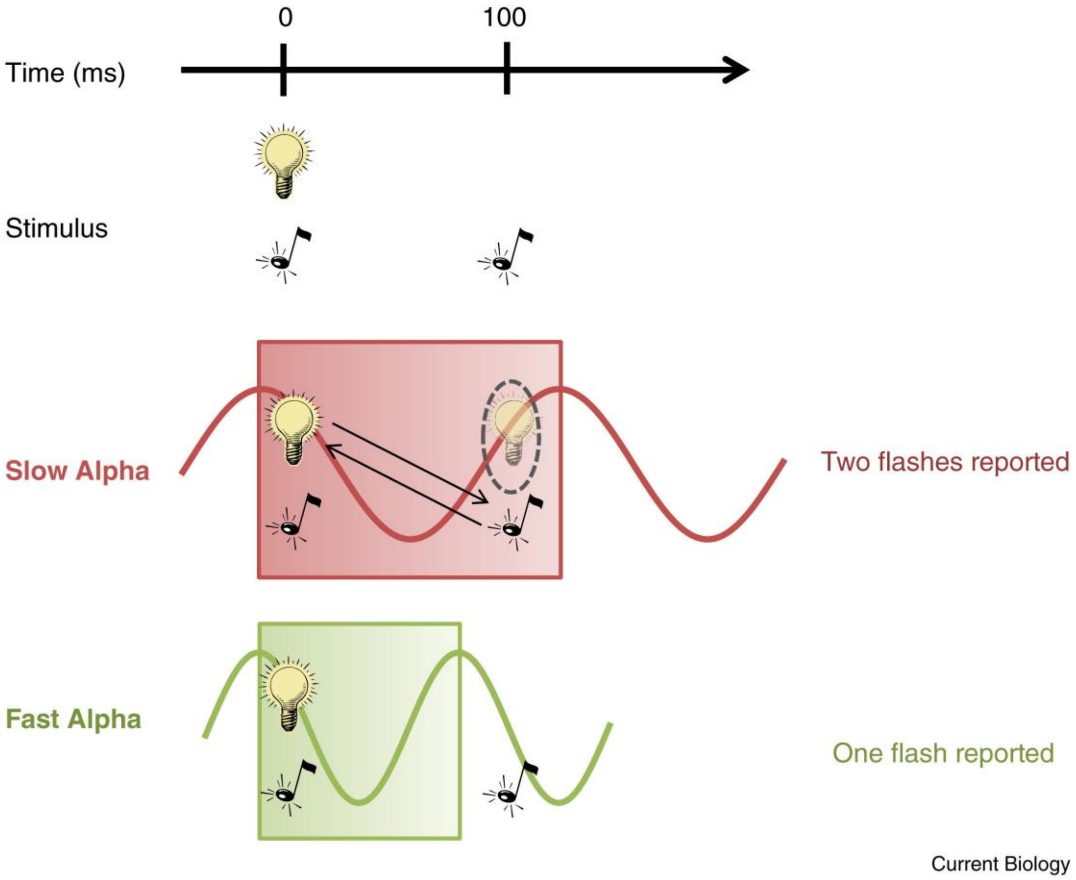


Figure 3-6: Modulation of the Double Flash illusion with TACS reported by Cecere, Rees, & Romei (2015), reproduced from Kerlin & Shapiro (2015).

In the present study we employed alpha frequency TACS as a method of disrupting normal sensory processing on the Sequence Discrimination Task outlined in chapter 2. This will be achieved by comparing task performance under TACS to that under sham stimulation, when the cortices perform as normal. We make no specific predictions about the direction of the effect of stimulation (i.e. that it will improve versus disrupt sequence discrimination ability) only that it will lead to a noticeable and quantifiable modulation in task performance.

Specifically our aim was to use TACS to establish the contribution toward performance of the temporal versus occipital cortices on both auditory and visual sequence discrimination abilities. On the premise that TACS will focally modulate the cortical activity at the stimulation site, this will allow us to examine the relative contribution to task performance of the two cortices and how they differ between those who do versus do not experience vEAR. TACS can then allow us to distinguish between two common hypotheses about the neural basis of intersensory crosstalk. The first of these, the cross-activation hypothesis (e.g. Hubbard, Brang, & Ramachandran, 2011), postulates that synaesthetic percepts are the result of atypical interconnectivity between two or more brain regions, leading to the intersensory crosstalk experienced by synaesthetes. If vEAR is explained by such increased cross-connectivity, and TACS interferes with this crosstalk in some way, then we would expect to see a greater effect of TACS in participants who experience vEAR than in those who do not. For example, in those who experience vEAR then TACS to the temporal cortex might modulate both visual and auditory sequence discrimination because crosstalk is interrupted, while in those without vEAR who do not experience this crosstalk, we would only expect performance modulation in the modality relevant to the TACS stimulation site. An alternative hypothesis is that this crosstalk is actually a common trait, rather than being unique to synaesthetes, but that it tends to be inhibited in those who do not experience synaesthesia (Grossenbacher & Lovelace 2001). If the visual and auditory cortices are mutually inhibiting one another (Mattingley et al., 1997) and alpha TACS interferes with this inhibition (Klimesch, Sauseng, & Hanslmayr, 2007), this would predict greater effect of stimulation in non-vEAR participants than in vEAR participants, e.g. visual sequence discrimination performance might be affected more by temporal stimulation in non-vEAR participants than those who experience vEAR.

3.2 Methods

3.2.1 Participants

A total of 36 naïve participants took part in the experiment. These included 20 participants recruited from the student population and the local community (age range 18-31 years, mean 23.1, SD 3.74, 7 male. There were also 16 classical musicians from the London Royal College of Music (RCM, age range 18-55, mean 24.44, SD 9.92, 9 male), having undergone between 5 and 46 years of regular musical training (M 15.3 years, SD 9.9). Five participants identified themselves as synaesthetes prior to participation, predominantly experiencing grapheme-colour and/or music-colour associations. All participants had normal or corrected vision and were screened for contraindications such as neurological and/or health conditions (i.e. a family history of epilepsy) that would preclude them from participating on the grounds of safety. Participants were recruited through opportunity sampling at City, University of London and the Royal College of Music, and all procedures were carried out after informed consent was obtained and were approved by the local Psychology ethics committee.

3.2.2 Apparatus and stimuli

The experimental procedure was conducted using an Apple Mac Mini connected to a 17” Sony HMD-A420 cathode ray tube (CRT) display. Auditory stimuli were presented through two Labtec PC speakers both positioned next to each other directly in front of and below the centre of the monitor. Video mode was 800x600 pixels with a 120 Hz refresh rate and a viewing distance was approximately 57cm (controlled using a chin rest). A small white fixation point marked the centre of the display. Subject responses were collected using the arrow keys on a standard computer keyboard. Experimental procedures and stimuli were programmed using Psychtoolbox for Matlab.

Visual stimuli consisted of circular white discs of 81cdm^{-2} luminance, presented centrally on a black background. Disk diameter was 3 degrees of visual angle. Auditory stimuli were sine wave tones with a maximum loudness of 91dBA sound pressure level (SPL) and a frequency of 360 Hz. ‘Short’ and ‘Long’ events were presented for periods of either 75 ms or 300 ms respectively, during which stimulation amplitude immediately decayed linearly from maximum to zero amplitude. Stimulation was delivered using a battery-driven alternating-current stimulator

(Magstim, UK) through two 5 cm x 5 cm conductive-rubber electrodes housed in sponges that had been saturated with saline solution.

3.2.2.1 Stimulation

Electrodes were secured using a rubber head strap and placed over either the occipital pole (O1, O2) or the temporal lobe (T3, T4), depending on the condition, following the international 10-20 system. The stimulation site was randomly counterbalanced between sessions. The temporal stimulation site was located by measuring the distance between preauricular joints, with the vertex taken as the midpoint. The point 10% above the preauricular joint bilaterally was identified as the site of the primary auditory cortex, corresponding to positions T3 and T4 on the international 10-20 system (see Figure 3-7). The occipital stimulation site was located by measuring the distance between the inion and the nasion, with the vertex taken as the midpoint. The point 10% above the inion was identified as the site of the primary visual cortex, and the electrodes placed either side of this point, corresponding to positions O1 and O2 on the international 10-20 system (see Figure 3-7).

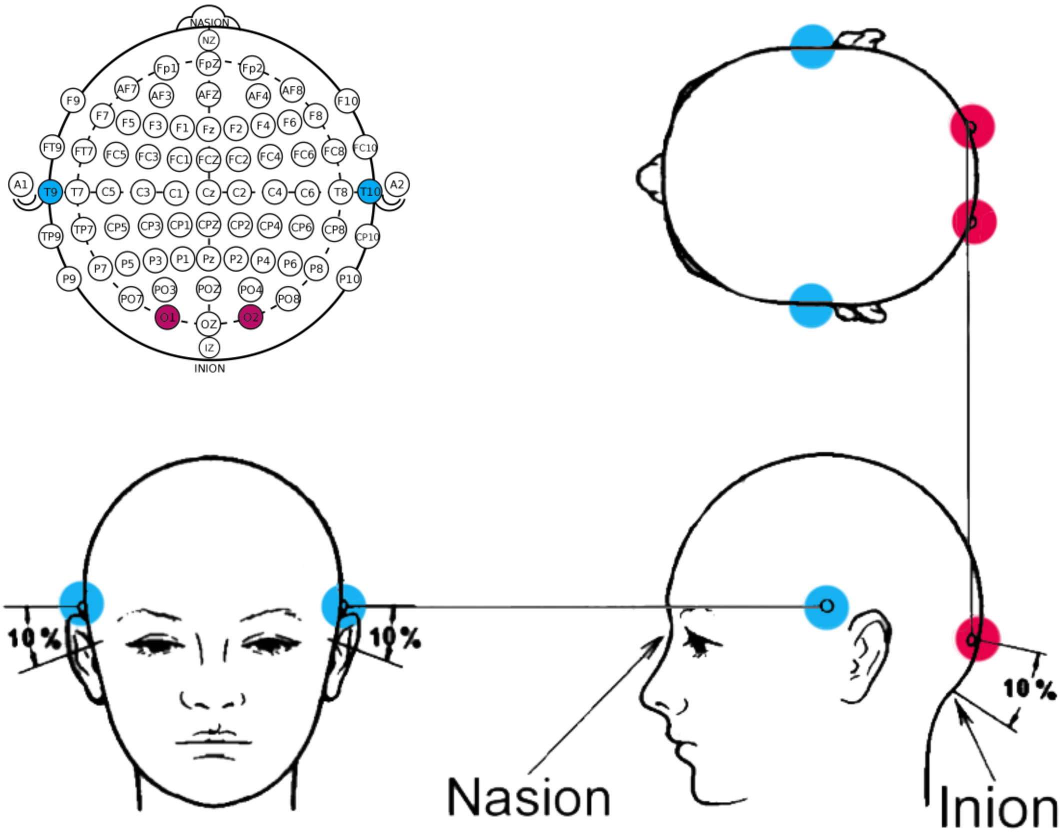


Figure 3-7: TACS Stimulation Sites.

Temporal T3 & T4 (blue highlight) and occipital site O1 and O2 (red highlight) according to the international 10-20 system.

Adapted from http://tronda-electronics.blogspot.co.uk/2011/11/eeg-background.html

In each experimental session participants performed the Sequence Discrimination Task under both sham and stimulation conditions, which were counterbalanced within session. Both the experimenter and the participant were blind as to whether sham or stimulation was being delivered. A minimum five-minute break was enforced between each condition in order to minimise carry-over effects, and a minimum of 24 hours separated each experimental session (occipital stimulation/sham versus temporal stimulation/sham).

In the stimulation condition, the current was 1000µA, 10 Hz, delivered bilaterally for 15 minutes during the Sequence Discrimination Task. The stimulator was pre-programmed to fade in over the course of 25 cycles (2.5 seconds), with each cycle equalling one sinusoidal waveform, and stimulation was delivered for a total of fifteen minutes, which was equal to 9000 cycles, ending with stimulation fading out over another 25 cycles. Impedance was kept below 10 kΩ for all sessions, with the saline concentration increased by dripping more solution onto the sponge if the impedance went above this level, until satisfactory.

In the sham condition the current remained at 1000µA and a frequency of 10Hz with a fade in and out over 25 cycles, although the total number of cycles was 300, the equivalent of thirty seconds of stimulation. This was included to replicate any initial ‘tingling’ sensation experienced in the stimulation session in order that the participant was less likely to identify which condition contained true stimulation.

3.2.3 Design

Here we employed three within-subject independent variables, each consisting of two levels. The first IV was the modality of the rhythmic sequence (visual flashes versus auditory beeps), the second was the site to which stimulation was applied (occipital pole versus temporal lobe), and the third was the type of stimulation being applied (10 Hz TACS versus sham). The behavioural task was the rhythmic Sequence Discrimination Task described carried out as described in chapter 2, again following Saenz & Koch (2008).

Participants underwent two separate experimental sessions, performing the Sequence Discrimination Task twice on each occasion. In one session they received neurostimulation delivered to the temporal cortex and in the other to the occipital cortex. The order of stimulation site was counterbalanced between participants. In each experimental session participants performed the experimental task once under sham and once under true stimulation conditions, the order of which was also counterbalanced between participants. Both the experimenter and the participant were blind as to whether sham or true stimulation was being delivered in any given iteration of the task. Within each experimental session a minimum five-minute break was enforced between iterations in order to minimise carry-over effects, and a minimum of 24 hours separated each experimental session (occipital session versus temporal session).

3.2.4 Procedure

Potential participants who expressed an interest in taking part were first provided with an information sheet to read explaining the nature of TACS. If they decided to participate their first session was booked no less than 24 hours later, to allow time for the participant to change their mind and to avoid pressurising anybody to participate. In the first experimental session participants were again given an information sheet and completed a safety screening questionnaire before providing informed consent. This was in order to exclude any individual with a history of epilepsy either personally or in their family, although this was a purely precautionary measure as there are no reported instances of TACS triggering a seizure. A first-aid trained researcher was present at all times as stipulated by the local ethics committee. The experimental task was explained and demonstrated, with the opportunity for a practice block if the participant wished. The relevant stimulation location for that session was then located using a measuring tape following the procedure outlined in the stimulation section above. The stimulation site depended on the randomised counterbalancing that had been established beforehand. Electrodes were then applied to the relevant area and secured by the use of a rubber strap. To ensure double-blinding a second experimenter was on hand to set up the stimulator prior to each task iteration, delivering either sham or alpha stimulation according to the counterbalancing. Once again participants were presented with 100 trials split into 5 blocks each consisting of 20 trials. In each block the stimulus modality was randomised.

On each trial two successive rhythmic patterns of stimuli were presented. In half of the trials the events were all visual, and in the other half all auditory. The modality of each sequence was randomized between trials. Within each sequence, constituent stimuli (events) could be either short (75 ms) or long (300 ms) with a total of eight events per sequence. Sequences were randomly generated and consisted of a minimum of four and a maximum of five transitions (i.e. a short event followed by a long or a long followed by a short). There was an inter-event interval of 100 ms between events, and an interval of 500 ms between the first and second sequence. On half of the trials, the two sequences were identical, and on the other half they differed. In ‘different’ trials, the first two events and the last event were always identical between pairs, while the order of the remaining events was randomly permuted. Immediately following the second sequence, participants were required to indicate whether they thought the two sequences were the same or different by pressing either the left or right arrow key on a PC keyboard, respectively. No error feedback was given. The response initiated the next trial. After the final trial in each block of 20 trials, participants could take a short break and the next block would begin when they pressed the spacebar. There were a total of 100 trials per session.

After the first task iteration the participant was given a mandatory minimum five minute break. Following this the second experimenter returned to set up the stimulator, and participants performed a second iteration of the sequence discrimination either with sham or TACS, whichever condition had not been run in the first iteration. On finishing this iteration the electrodes and straps were removed. A second experimental session followed no less than 24 hours later, in which the exact procedures outlined above were repeated for the other stimulation location. After the final iteration of the task in the second session, participants were asked the debrief questionnaire set out in chapter 2, including asking whether they had been aware of hearing the visual flashes, before being thanked for their time and being provided either with cash payment or course credits in the case of psychology undergraduates.

3.3 Results

Fisher’s exact tests revealed that frequency of reporting hearing flashes was higher in musicians [odds ratio=22.00, p=0.0001], and those reporting synaesthetic experiences [odds ratio=24.00, p=0.0001]. RCM musicians also tended to report synaesthesia more frequently [odds ratio=9.43, p=0.0043 odds ratio=9.43, p=0.0043]. Figure 3-8 shows the proportions as stacked bar charts.

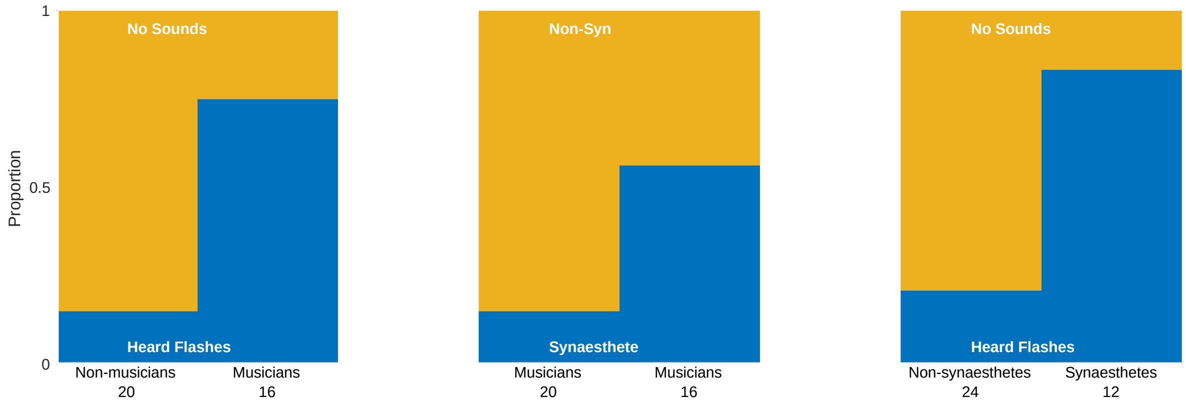


Figure 3-8: Stacked bar charts illustrating relative proportions of participants in different groups.

Sensitivity measures (d’) for participants’ same/different discrimination judgements were first calculated following standard psychophysical methods (Green & Swets, 1966). The mean d’ scores were then computed grouped by stimulation type (alpha TACS versus sham), stimulation site (occipital versus temporal), and task modality (visual sequences versus auditory sequences). The mean for each of these is presented in **Error! Reference source not found.**Table 3-1.

	Stimulation	
	Alpha	Sham

		Mean d’	Standard Error of the Mean	Mean d’	Standard Error of the Mean
Visual Task	Occipital Site	1.834	0.170	1.981	0.152
	Temporal Site	1.746	0.166	1.591	0.174
Auditory Task	Occipital Site	3.384	0.222	3.229	0.229
	Temporal Site	3.035	0.212	3.347	0.221

Table 3-1: Mean sequence discrimination sensitivity (d’) for visual and auditory task broken down by stimulation site and type (active vs. sham), with standard error.

To check whether there were any consistent carry-over effects of tACS on performance in the sham condition when following stimulation, we analysed d’ for just the sham conditions in an ANOVA for each of the two electrode montages, grouping data by the order of sham condition. Sham order had no significant main effect, and did not interact significantly with Modality (Visual vs. Auditory).

d’ scores were analysed in an ANOVA including Site (Occipital versus Temporal), Modality (Visual vs Auditory) and Stimulation (Sham vs. tACS) as repeated measures, and Yes/No responses to the ‘Hearing Flashes’ debriefing question as a grouping variable. Performance was significantly better overall in participants who reported ‘Hearing Flashes’ [F(1,34)=12.34, p=.001, η^2_p=.27], and auditory (A) d’ was significantly higher than visual (V) performance on average [F(1,35)=134.90, p<.00001, η^2_p=.79]. Means (and SE) for the different groups and conditions are as follows: ‘Yes’: V 2.39 (0.17); A 4.12 (0.33), N=15; ‘No: V 1.60, (0.19); A 3.05 (0.81), N = 21. There was no significant main effect of tACS versus Sham, or for Site. However, there was a significant interaction between

Site, Stimulation and Modality [F(1,34)=6.43, p=.016, η^2_p=.16]. Furthermore, these variables interacted significantly with ‘Hearing Flashes’ [F(1,34)=7.00, p=.012, η^2_p=.17].

To quantify the effects of tACS, we subtracted Sham from tACS d’ scores for each stimulation site separately, so that negative scores represent a decrement in performance. The results are shown in Figure 3-9, with asterisk and bracket annotations indicating significance of Tukey multiple comparisons at p<.05. The pattern of results appear to reveal reciprocal effect on task performance of tACS applied at different sites, particularly in non-vEAR participants. For example, A performance was significantly impaired by temporal tACS relative to sham and to occipital stimulation, but there was a non-significant trend for A performance to slightly improve with occipital stimulation. V performance was also significantly poorer with occipital stimulation compared to A, which again showed a non-significant trend to improved discriminability. In contrast, there were no significant deviations from sham performance in vEAR participants.

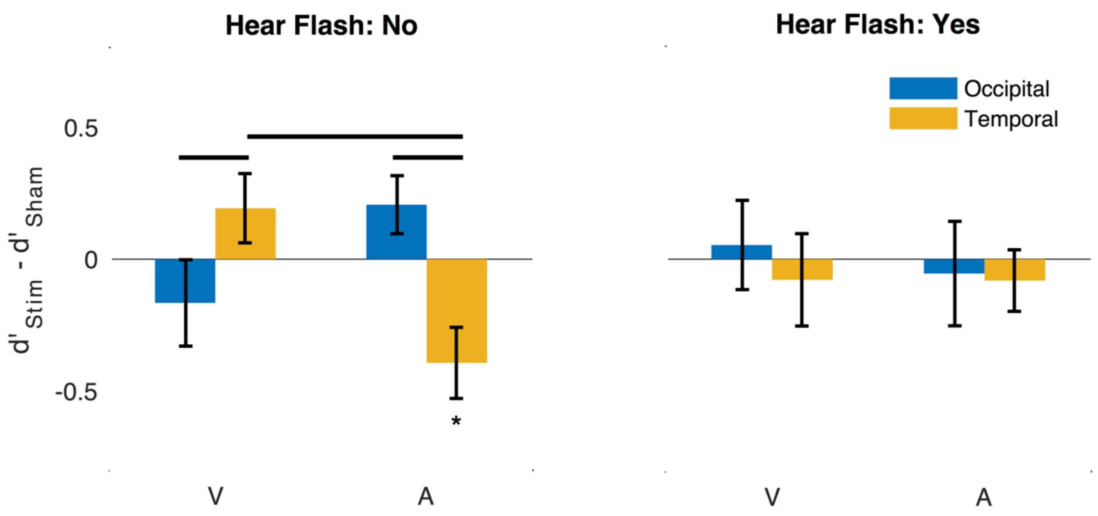


Figure 3-9: Effects of tACS on d' sensitivity (relative to sham stimulation) to differences in visual (V) or auditory (A) sequences, relative to sham, or two occipital versus temporal electrode montages (dark and light grey respectively). Positive values indicate an improvement due to tACS. Error bars show standard error; asterisks and horizontal brackets show significant differences (p<.05, Tukey comparisons). Left graph represents participants who did not report visually-evoked auditory sensations (N=21) evoked by the flashes in the visual sequence discrimination condition. Right graph is for participants (mostly musicians) who did report visually-evoked auditory sensations (N=15).

In case the trends seen in Figure 3 had been weakened by participants who overall experienced less effects of tACS than others, we correlated individual scores for each task, under occipital stimulation against temporal stimulation (see Figure 3-10). This analysis found significant negative correlations specifically in non-vEAR participants (left graph), confirming that greater decrements in performance related to one stimulation site coincided reciprocally with greater improvement at the other site. In particular, impairment of A performance under temporal stimulation significantly correlated with improved performance of the same task under occipital stimulation [r(19) = -.69, p<.001]. A similar significant negative correlation was observed for V performance [r(19) = -.79, p<.001], where greater impairments under occipital stimulation coincided with greater improvements under temporal stimulation. No such trends were observed for vEAR participants (right graph).

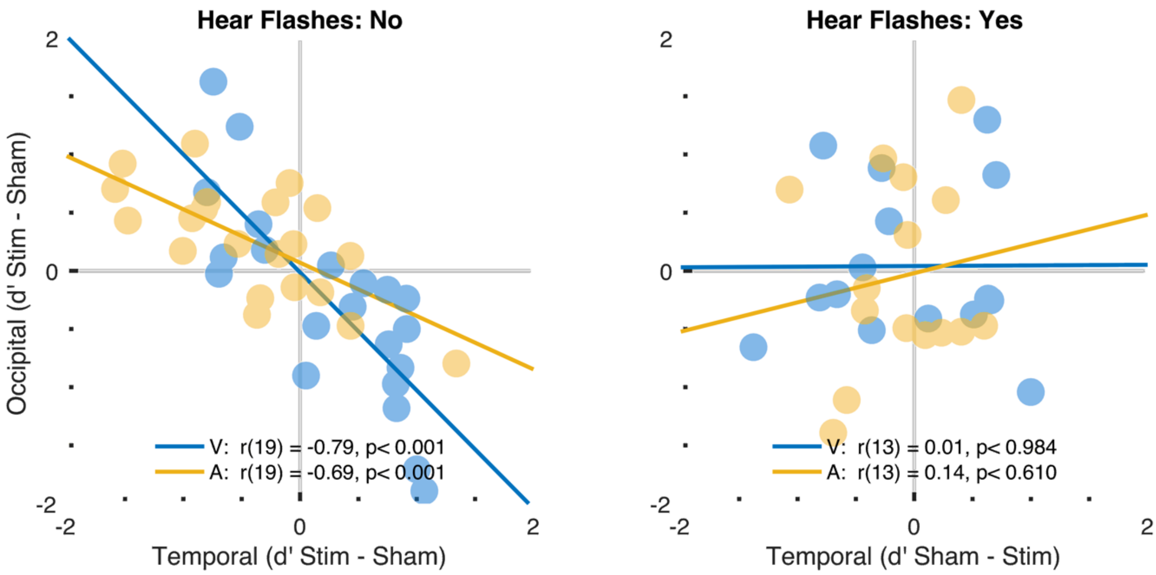


Figure 3-10: Scatterplot of effects of tACS electrode montage on d’ sensitivity (relative to sham stimulation). Separate colours for Visual (dark blue) and Auditory (light yellow) sequence discrimination tasks. Separate graphs for participants who did not report hearing flashes (left) and who did (right). Pearson correlation statistics shown in legend.

3.4 Discussion

Here we aimed to discover whether TACS differentially modulated visual compared to auditory sequence discrimination ability in those who do versus do not experience vEAR. We achieved this by comparing the effect of TACS when delivered to the temporal versus the occipital lobes. We were also interested in how musicians differ from non-musicians in this regard, due to their many years of exposure to highly correlated motion (e.g. in the form of other musicians' body movements) with sound (the output of their instruments) and whether this translated into a higher prevalence of vEAR. Our results demonstrate that the effect of TACS was dependent on the site to which stimulation was applied (occipital versus temporal cortex), the modality of the task (visual versus auditory sequence discrimination) and whether or not participants reported being aware of hearing sounds accompanying the flashes. In those who did not experience vEAR, TACS degraded task performance in the modality associated with stimulation site (i.e. visual sequence discrimination with occipital stimulation and auditory with temporal stimulation), but actually enhanced performance in the opposing modality (i.e. visual sequence discrimination with temporal stimulation etc.). In those who did report experiencing vEAR there was no effect of TACS on sequence discrimination ability.

We had suggested two potential mechanisms that may underlie the experience of vEAR, following existing models proposed to explain other forms of synaesthesia. The first of these is the cross-activation hypothesis, in which vEAR would be the result of an abundance of connections between auditory and visual cortex that are not found in the typical population. This is contrasted with the disinhibition hypothesis, which suggests that it is not the overall level of connectivity between the cortices that differ between those who do versus do not experience vEAR but the differential levels of inhibition between these two populations. If vEAR is primarily explained through cross-activation then we suggested that TACS would have a greater effect in those who experience vEAR, as stimulation of either the temporal or occipital cortex would modulate performance on both the visual and auditory task. Conversely if disinhibition of intercortical crosstalk explains vEAR then we would expect to see a lesser effect of TACS in those who do experience vEAR compared to those who do not, as was the case in our results.

What we had not anticipated was the crossover effect observed in the participants who did not experience vEAR, for whom TACS degraded task performance in the modality relevant to stimulation site but actually enhanced performance in the opposing modality (See Figure 3-11).

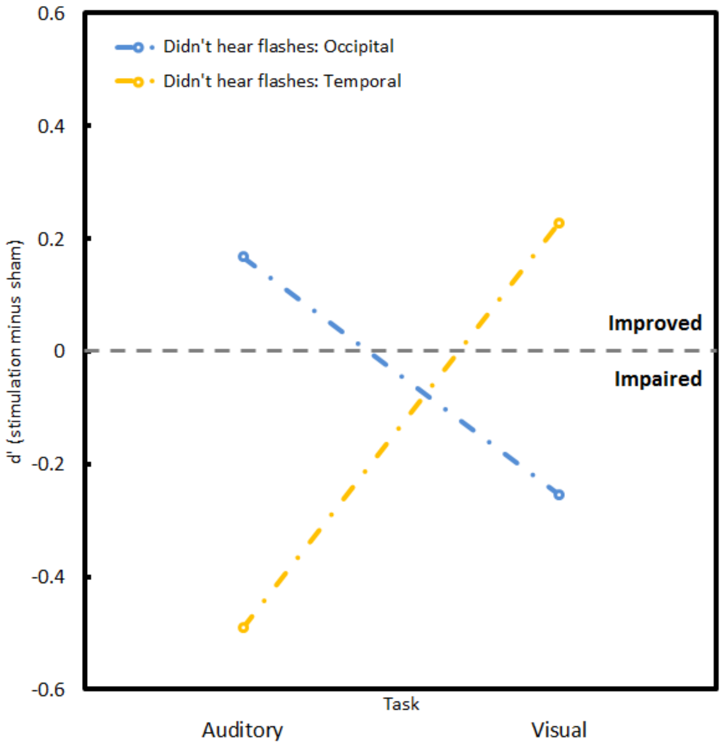


Figure 3-11: The effect of TACS in participants who did not experience vEAR

This suggests that in those who do not experience vEAR the visual and auditory cortices may mutually inhibit one another (Figure 3-12), with this inhibition being carried by alpha frequency oscillations (Klimesh, Sauseng, & Hanslmayr, 2007). Alpha is the dominant oscillatory frequency in the human brain (Linkenkaer-Hansen et al., 2004; Klimesch, 2012) and is thought to represent a mechanism of short-range cortical inhibition, (Jensen and Mazaheri, 2010; Klimesh, Sauseng, & Hanslmayr, 2007; Mathewson et al., 2009; Sauseng et al., 2009), with a greater alpha amplitude meaning a greater level of local inhibition. In studies employing biased competition paradigms in which a subject must attend to one stimulus and ignore another, alpha amplitude is greater in cortical regions associated with the unattended stimuli (e.g. Foxe et al., 1998, Snyder and Foxe, 2010). By applying alpha band TACS we may then be biasing this inhibitory balance in favour of disinhibition of the unstimulated region (Figure 3-13), potentially feeding back to inhibit the stimulated region.

In participants who did experience vEAR there was a reduced effect of TACS, which suggests that any such cortical rivalry is weaker in those who are able to hear visual events. It may be that these individuals have less of this mutual inhibition (Figure 3-14) so the application of alpha band TACS has less effect (Figure 3-15). In this case, rather than compete for resources the auditory and visual cortices of people who experience vEAR instead cooperate rather than inhibit one another, leading to a reduced effect of TACS.

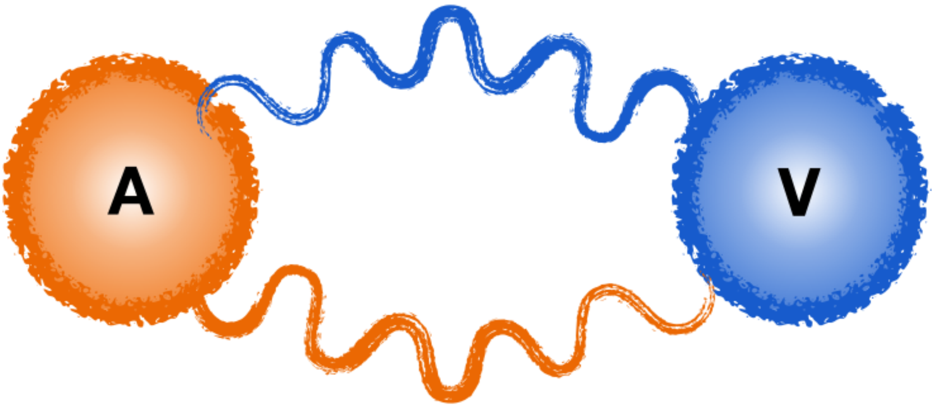


Figure 3-12: Mutual inhibition of the visual and auditory cortices carried by alpha Oscillations in participants who do not experience vEAR.

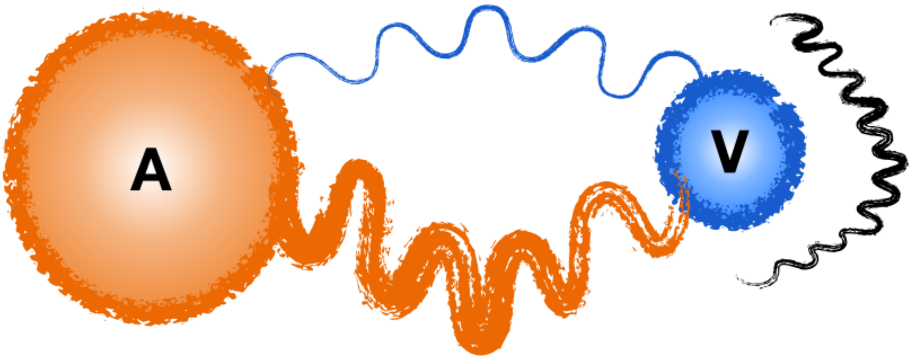


Figure 3-13: Alpha-band TACS (black line) applied to one cortex biases these inhibition in favour of the unstimulated region of cortex in participants who do not experience vEAR.

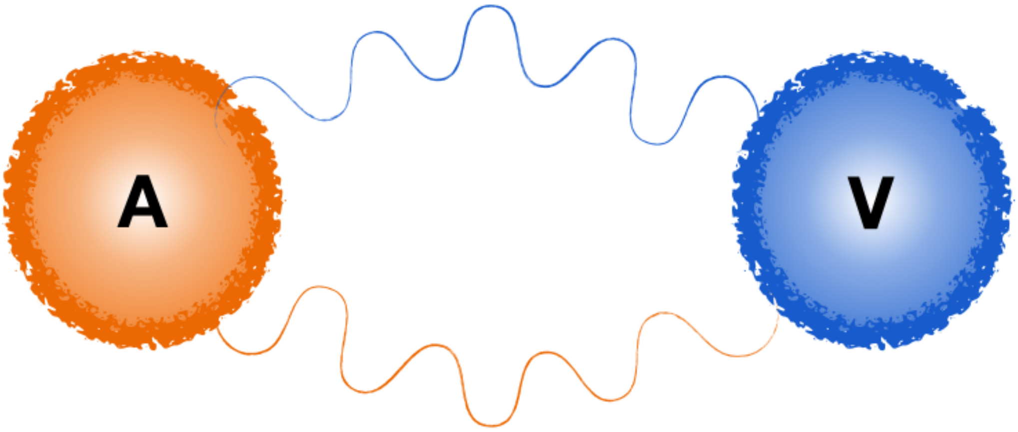


Figure 3-14: Less inhibition between cortices in participants who experience vEAR.

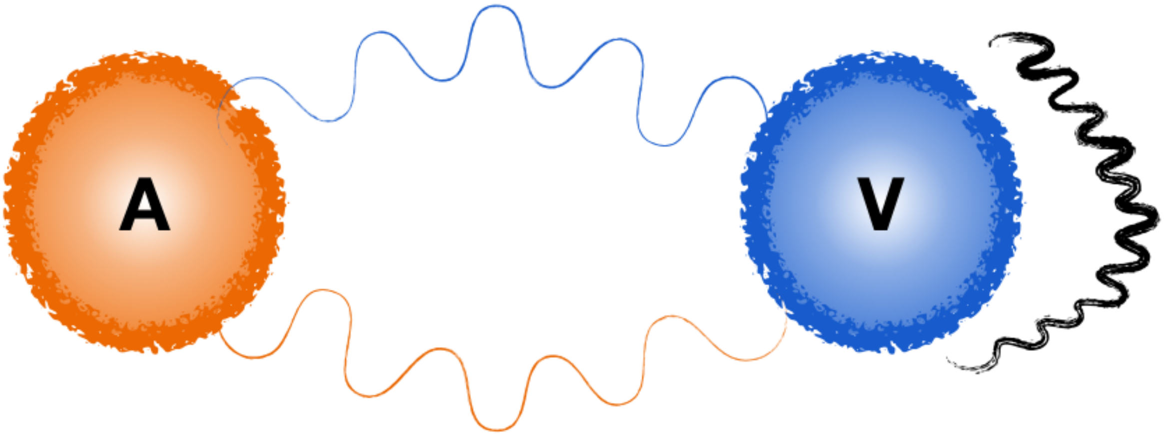


Figure 3-15: Alpha-band TACS applied to one cortex has little effect as there is less intercortical inhibition in those who experience vEAR.

As well as explaining the TACS results described above, this lack of inhibition may also explain why these individuals report hearing visual events in the first place. With less mutual inhibition it is likely that overall auditory cortical excitability is generally higher in these individuals, meaning it takes comparatively little signal from the visual areas to sufficiently excite the auditory cortex to generate faint auditory qualia.

One interesting finding in our results was the significantly higher rates of vEAR experienced by musicians compared to non-musicians. This may be because musicians, and highly-trained classical musicians in particular, are especially adept at anticipating and predicting musical cues from their fellow players' bodily movements; for example, the sudden downward motion of a percussion players arms signifies an imminent crash, while the outward jutting of a violinist's elbow will synchronize with the sound of their instruments. The cumulative effective of years of attending to these cross-modal cues, often from childhood, may be to reinforce cortical connections between auditory and visual regions. In non-musicians too, this strong association between sight and sound may explain findings that our perception of the quality of a musical performance is driven largely by the visual spectacle rather than simply what we hear (Tsay, 2013), or the popularity of multimedia displays (lights, lasers, dancers etc.) at musical events. The increased prevalence of vEAR in musicians may be the result of the disinhibition of connections between the visual and auditory cortices outlined above. As a result, musicians may be more prone to experience vEAR if the baseline excitability of the auditory cortex is higher as a result of this disinhibition.

There are limitations to this study insofar as we have not yet tested any other frequencies of TACS, such as Gamma, meaning our hypothesis about the mutual-inhibition in those without vEAR being carried my alpha oscillations remains tentative. We present compelling evidence that alpha TACS is able to disrupt this pattern of inhibition, but we cannot say that it is not the effect of stimulation *per se* that is disrupting cortical processing. However, as we did not observe an effect of TACS in participants who experience vEAR this seems unlikely to be the case, as any generic cortical 'jamming' should otherwise effect both vEAR and non-vEAR participants equally. We can therefore say that we present evidence that the brains of vEAR and non-vEAR participants do indeed appear to differ in the extent to which their auditory versus visual cortices cooperate versus compete respectively, although until we examine a range of TACS frequencies our claims about specific mechanisms must remain tentative (although see chapter 5 for some preliminary findings with Gamma TACS).

A second limitation comes in the lack of generalizability of our findings from musicians to the wider population. Although we presented an a priori rationalisation of why we predicted greater levels of vEAR in highly-trained musicians, in which these individuals have a greater than average exposure to highly synchronised sound and motion (and are thus the extreme end of normal) it is possible that musicians differ from the general public in other ways we had not anticipated. For example, musicians' brains display reduced inhibition between hemispheres (Ridding, Brouwer, & Nordstrom, 2000) potentially due an enlarged anterior corpus callosum (Schlaug et al., 1995) compared to controls. If there is a confound it may be in the experience of these participants to translate visuospatial symbols (in the form of musical notation) into both sound (both internally in the form of imagery and externally via their instrument) and the necessary motor actions to perform this translation. This process requires a sequencing skill absent in non-musicians that has neurological correlates in regions such as Broca's area (Sluming et al., 2007) and wider motor and auditory areas (Gaser & Schlaug, 2003). Therefore it may be the case that our musicians are better at translating from one modality into another, particularly in the temporal domain, rather than experiencing vEAR in a spontaneous way. However, this would not explain why a greater proportion of our musicians did indeed report being able to hear the flashes in the visual Sequence Discrimination Task, compared to controls.

3.5 Summary

In summary, in individuals who did not report hearing the visual flashes, TACS disrupted sequence discrimination ability in the modality relevant to the stimulated region of cortex, while actually improving performance in the opposing modality. We propose that this is due to a naturally-occurring mutual inhibition between the visual and auditory cortices, in which each must compete for attentional resources by inhibiting the other (e.g. Lurilli et al., 2012) possibly carried via alpha frequency oscillations (Foxe & Snyder, 2011; Frey et al., 2014; Klimesch, Sauseng, & Hanslmayr, 2007; Strauß, Wöstmann, & Obleser, 2014). This mutual inhibition might facilitate selective attention to one modality over the other. The effect of TACS was absent in the participants who do hear the visual flashes, indicating that this mutual inhibition is less dominant in those who experience vEAR. This suggests that one key difference between these two populations is whether their visual and auditory cortices compete or cooperate for resources; specifically they can use either modality to perform the task, so they employ the cortices in unison. In those who experience vEAR, for example, TACS may have had less effect overall because there is less alpha-mediated competition between cortices, and thus TACS cannot effectively interfere with such competition and bias the balance of dominance between vision and audition (see Figure 3-14 and Figure 3-15). We had hypothesised that vEAR may be explained by excess levels of connectivity between the visual and auditory cortices in those who experience the phenomenon, and/or by an atypical disinhibition of the otherwise normally-occurring connections between the cortices.

As predicted we saw significantly higher levels of vEAR in our classical musicians than in other subjects. We suggest that this may be because this population have been particularly exposed to years of correlation between movement, such as the conductor’s baton or the body movements of other players, and the sound of the orchestra. This is in turn consistent with our assertion that a degree of learning is involved in the development of vEAR, and the higher prevalence of vEAR which is characterised by highly consistent pairings in the natural world, compared to some canonical synaesthesias.

These findings build on those reported in Chapter 2 in several ways. First, they begin to examine the physiological differences between those who do versus do not report experiencing vEAR. In addition we gain support for our hypothesis that vEAR may be more prevalent than canonical synaesthesia due to the increased co-occurrence of movement and sound in the natural world by demonstrating that individuals who are particularly exposed to this co-occurrence are more susceptible to experiencing vEAR. Finally, we begin to tentatively construct the profile of who experiences vEAR by demonstrating the aforementioned relationship with musicianship. In the following chapter we further explore the demographic and trait predictors of vEAR, as well as the specific properties of visual stimuli that best evoke vEAR in respondents to a large online survey.

Experiment 3. Who hears visual motion, and what looks loudest? A large-scale online survey

Chapter 4: [2]

This chapter explores the types of visual stimuli that most effectively evoke vEAR, and those who tend to experience it. Here we describe the results of a large online survey in which respondents were asked to rate the amount of auditory sensation evoked by a series of randomly presented silent videos depicting a range of motion types. These varied from biological motion, such as dancers, to impacts, such as a hammer hitting a nail, as well as more abstract computer-generated imagery. Respondents were also asked other demographic questions relating to their auditory perception. A Principle Component Analysis was performed on the data, with motion energy and predictiveness of sound were identified as the two major components contributing to ratings. Predictiveness was shown to influence the video ratings in all participants but motion energy was specific to those who experience vEAR. Other characteristic traits that predict higher included the frequency an individual experiences musical imagery in their head, or whether they have other types of synaesthesia. Results indicate that predictiveness of sound is a common contributor to video ratings across all respondents, while motion energy is an additional factor that influences ratings specifically in those who experience vEAR.

---

[2] This chapter is based on data that has previously been published in Fassnidge, C., & Freeman, E.D. (2018). Sounds from seeing silent motion: Who hears them, and what looks loudest? Cortex, (in press).

4.1 Introduction

In the present study we aimed to gain a fuller understanding of the types of visual stimuli that evoke high ratings of vEAR in terms of intensity of sensation, and of the kinds of individuals who experience these sensations (i.e. other demographic and trait predictors of vEAR experience). We also explore whether the auditory sensations are generated via low-level or high-level mechanisms. To address these questions we devised an online tool that required participants to view a series of short, silent video clips depicting an assortment of different movement types, ranging from biological motion, such as dancers and people walking, to 'high impact' movements such as a hammer hitting a nail or a box striking a punch bag, as well as videos with a learned expectation of an accompanying sound, such as fireworks going off or a face screaming. Other more abstract videos had little predictiveness but contained high levels of motion energy, such as twinkling lights or LED displays. For each of these stimuli participants were asked to numerically rate them from 0 to 5 for the amount of internal auditory sensation they experienced when viewing the content.

Another aim of the present study is to establish an estimate of prevalence in a larger sample than our random laboratory sample presented in chapter 2. Our findings thus far appear to support the hypothesis that vEAR is a normal phenomenon that is relatively widespread in the population, certainly compared to canonical synaesthesias. In our results from chapter 2 we report that 20% of our experimental sample reported when prompted that they were aware of hearing some kind of auditory sensation accompanying the presentation of visual flashes presented on a computer monitor as part of a visual Sequence Discrimination Task, and 11% reported that they were aware of experiencing this in daily life. We outlined in chapter 1 why we believe that this prevalence may be so markedly higher than other estimates of synaesthesia occurrence. In summary, if a degree of early years learning is involved in the development of synaesthetic pairings, then we would expect stimuli that co-occur frequently in the natural environment to be more robustly reinforced through such learning and thus these pairings will be more common than some of the more unusual pairings, such as between graphemes and odours for example, which are not frequently reinforced.

If, as our data suggest, the prevalence of vEAR is indeed substantially higher than existing estimates of synaesthesia, then we may wish to examine our classification of the phenomenon in order to ensure that we are comparing like with like. In chapter 1 we state that we maintain an 'agnostic' stance on whether vEAR meets all necessary criteria to be considered a true synaesthesia sub-type. One of the key questions we wish to address is whether or not vEAR reflects genuine sensory crosstalk between brain areas processing basic visual and auditory information (Schroeder & Foxe, 2005) that is driven by primarily bottom-up processes. Alternatively, vEAR might instead be a kind of reflective top-down form of imagery, generated consciously by expectations and predictions. This type of analysis is often problematic as such associations can be acquired between stimuli that are both high-level and cultural in origin, such as letters or words evoking colours (Bor, Rothen, Schwartzman, Clayton, & Seth, 2014; Witthoft, Winawer, & Eagleman, 2015). If vEAR were exclusively the result of high-level associations between stimuli we might expect participants to give the highest ratings on our video questionnaire to items that depict events which they have learnt to be highly predictive of an accompanying sound in the real world, such as explosions, impacts, or lip movements, whereas items depicting motion that does not predict sound, such as flashing neon lights or abstract patterns would evoke lower scores. Conversely, if vEAR is predominantly the result of low-level crosstalk between early visual and auditory areas then we might expect item scoring to also be driven by more basic stimulus properties, such as the 'motion energy' contained within the scene (Adelson & Bergen, 1985), regardless of their meaning or any learned association.

It is feasible that these two paths to vEAR, high and low-level that are differentially expressed in different individuals, in which case we aim to establish how these two cohorts differ and what other factors and traits correlate with which each type of vEAR. Gaining this understanding may help inform an on-going debate concerning the extent to which the brains of people with synaesthesia are fundamentally unique in their architecture, or whether they are structurally normal but differ from others in that the levels of inhibition between different brain regions. This is discussed in length in chapters 1 and 3, in which we present the cross-activation versus the disinhibition hypotheses of synaesthesia. Our findings in chapter 3 in particular seem to support the concept that vEAR is characterised by a disinhibition of crosstalk between visual and auditory areas. In the present study we hypothesise that if vEAR does indeed reflect systemic cross-modal disinhibition, then video ratings for vEAR might correlate with other traits potentially associated with greater cortical excitability or reduced inhibition, such as the frequency with which one experiences musical ‘earworms' (Kumar et al., 2014) (also known as Involuntary Musical Imagery (INMI)), tinnitus (Kaltenbach, 2011), and the hypnogogic auditory-evoked visual sensations (phosphenes perceived when one is falling asleep, usually in darkness, and awoken by a sudden sound). This phenomenon is little-studied but often reported anecdotally, although similar experiences have previously been documented in patients with pathologically reduced visual input (Afra et al., 2012; Jacobs et al., 1981; Lessell & Cohen, 1979). This phenomenon might arise due to enhanced excitability of visual cortex during light deprivation which may unmask input from connections from outside visual areas (Boroojerdi et al., 2000).

4.2 Hypotheses

Our first set of hypotheses concerned the characteristics of the visual stimuli that evoke vEAR. Primarily we were interested in testing two hypotheses that may elucidate the mechanisms that underlie vEAR; firstly, that vEAR is strongest when there is a learned semantic association between a visual stimulus and a particular sound, as would be the case with lip movements and speech, or fireworks. In these cases it is the learned expectation of a sound that may generate mental representations of that sounds, irrespective of the amount of movement depicted. For example, a video of a person screaming might be relatively static, but the learnt facial expression and specific positioning of the mouth is strongly evocative of a loud sound.

Our second set of hypotheses concern which stimulus features predict higher ratings from respondents. First, it may be that the strength of the auditory sensation depends only on low-level properties of the stimulus, such as the amount of motion energy in the visual display. This information can be extracted and objectively measured using existing algorithms (following Adelson & Bergen, 1985) that model the spatiotemporal receptive field properties of neurons in areas such as V1 and V5, and therefore represent a biologically plausible method by which such strictly bottom-up stimulus properties may be extracted from the visual scene (e.g. Challinor &

Mather, 2010; Emerson, Bergen, & Adelson, 1992; Heeger, 1993; Ringach, 2002; Watson & Ahumada, 1985). Alternatively, it may be that videos which are more predictive of an accompanying real-world sound are rated as evoking more vEAR. By comparing ratings for vEAR according to these properties we are able to provide evidence for the mechanisms underlying the vEAR phenomenon.

These two hypotheses represent two potential methods by which mental ‘sound’ could be extracted from a given visual scene, with a potential high-level route via learned associations, semantics and top-down prediction, or a lower-level route via more direct intercortical connections. We also proposed a third hypothesis, which suggests that these two routes to vEAR might be differently expressed in people who say they hear movement and those who do not, with those who do not being more influenced by learned associations and visual imagery when rating the videos, while those who do experience VEAR are in addition sensitive to low-level influences.

Building on our hypothesis of reduced inhibition in those who experience vEAR, we hypothesised that a number of other traits which also reflect reduced inhibition of the auditory and neighbouring cortices may predict vEAR. As such we included a series of questions asking about the respondents’ experience of phenomena such as tinnitus, involuntary musical imagery, and hypnagogic sound-induced phosphenes. We hypothesised that these traits may also reflect reduced cortical inhibition, and as such would be observed in greater numbers in those who experience vEAR relative to those who do not.

4.3 Methods

4.3.1 Materials and Stimuli

We devised and administered a video questionnaire using Qualtrics, an online portal for creating and circulating surveys and questionnaires via the Internet. Our stimuli consisted of 24 copyright-free video clips downloaded from www.videoblocks.com that were chosen to reflect a range of different motion types which were and were edited to 5 seconds in duration. Examples included a television being smashed with a sledgehammer (Figure 4-1), a bouncing tennis ball (Figure 4-2), and a ballet dancer performing a pirouette (Figure 4-3), as well as more abstract digital animations such as flickering lights (Figure 4-4) (visit the following URL to view the visual stimuli: goo.gl/xARxPE or to view the survey itself visit www.tinyurl.com/vEARsurvey).

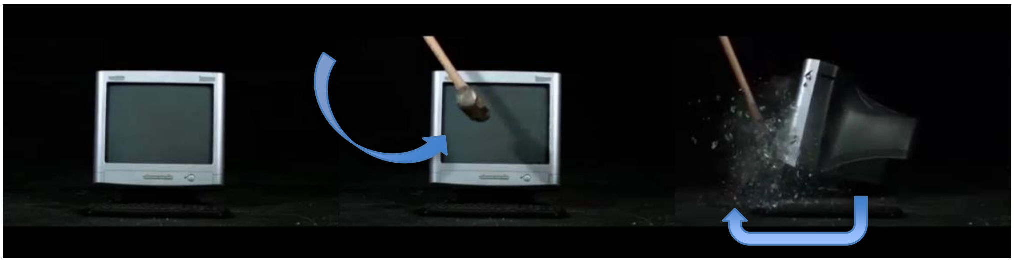


Figure 4-1: Hammer swings, strikes TV Screen, Screen smashes and spins right with the impact.

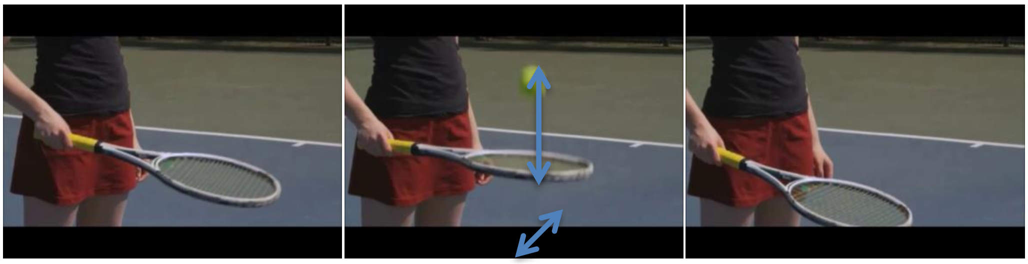


Figure 4-2: Tennis ball bounces on racket, racket moves up and down with impact.

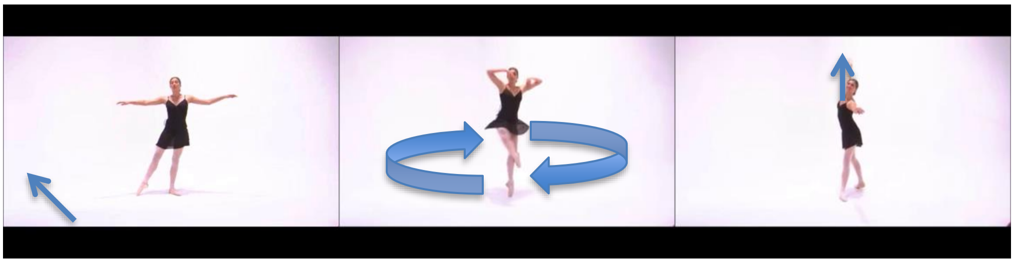


Figure 4-3: Ballet dancer performs pirouette (leg is elevated, dancer spins, arm is lifted)

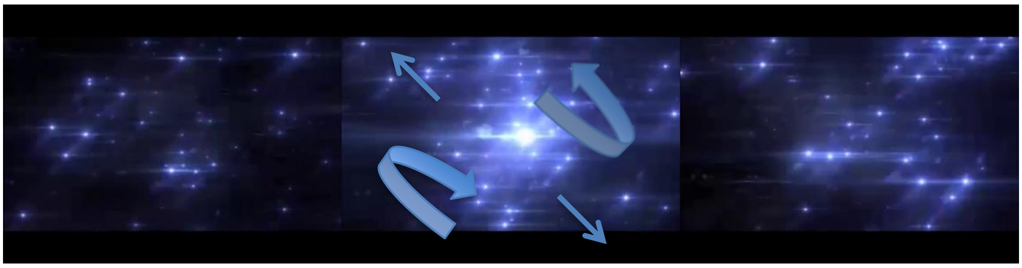


Figure 4-4: Blue lights flicker and move in a ripple formation

4.3.2 Participants

Participants were recruited in two ways. 17 participants took part in our laboratory after completing one of the experimental tasks outlined in previous chapters. We also made the questionnaire publicly available via the internet, with invitations to participate placed on several Facebook groups for people who experience or have an interest in synaesthesia such as:

Synaesthesia Research:

https://www.facebook.com/groups/synaesthesiaresearch/

UK Synaesthesia Association Facebook Page:

https://www.facebook.com/groups/128219365930/

I'm not a freak, I'm a synaesthete:

https://www.facebook.com/groups/2226778430/

In addition, considerable traffic was driven to the questionnaire after a URL was included in coverage in the popular press of Fassnidge, Cecconi-Marcotti and Freeman (2016), which reports the findings presented here in chapter 1. The articles presented a summary of our findings, as well as an accessibly layman yet accurate explanation of vEAR. These articles are available at the below URLs:

http://www.dailymail.co.uk/sciencetech/article-4129468/Take-test-reveals-HEAR-flashes-light.html

https://www.theguardian.com/science/2017/jan/17/listen-with-your-eyes-one-in-five-of-us-may-hear-flashes-of-light-synaesthesia

A total of 32,947 individuals consented to take part, however only 4,061 of these completed every item on the questionnaire. We set a cut off criteria for inclusion of participants who left no more than 2 of the 24 videos unrated, which equalled 4,516 people. Of these, 2,333 were male, 1,888 female, 50 identified with another gender identity, and 245 declined to disclose their gender (See Table 4-1 and Figure 4-5). The Qualtrics software records the geographic location of all respondents who complete the questionnaire in full. Respondents were located all over the world, with most concentrated in North America and Western Europe (See Figure 4-6).

Demographic	N	Percentage
Age		
18 - 24	900	19.9
25 - 34	1219	27.0
35 - 44	912	20.2
45 - 54	660	14.6
55 - 64	397	8.8
65+	184	4.1
Undisclosed	244	5.4
Gender		
Female	1888	41.8
Male	2333	51.7
Other	50	1.1
Undisclosed	245	5.4

Table 4-1: Age and gender breakdown of questionnaire respondents

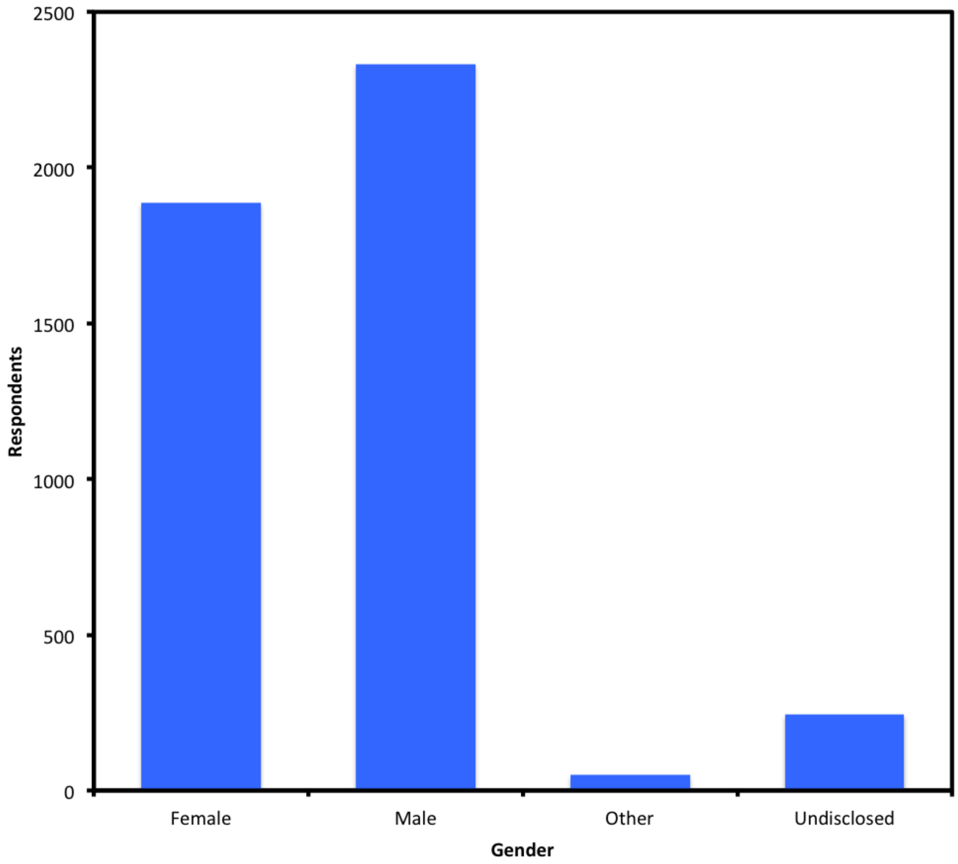


Figure 4-5: Bar chart depicting gender breakdown of respondents.

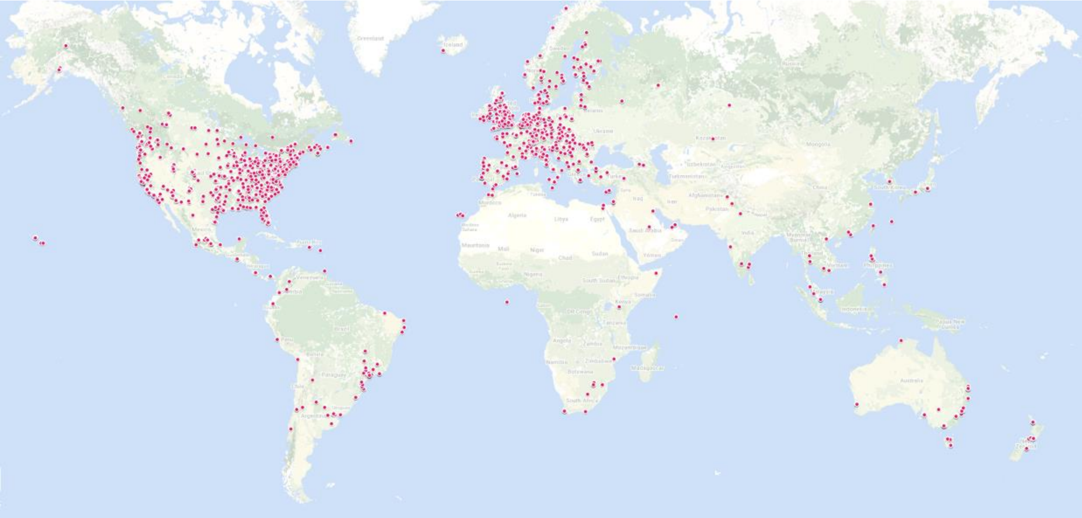


Figure 4-6: Geographical location of questionnaire respondents who completed all items.

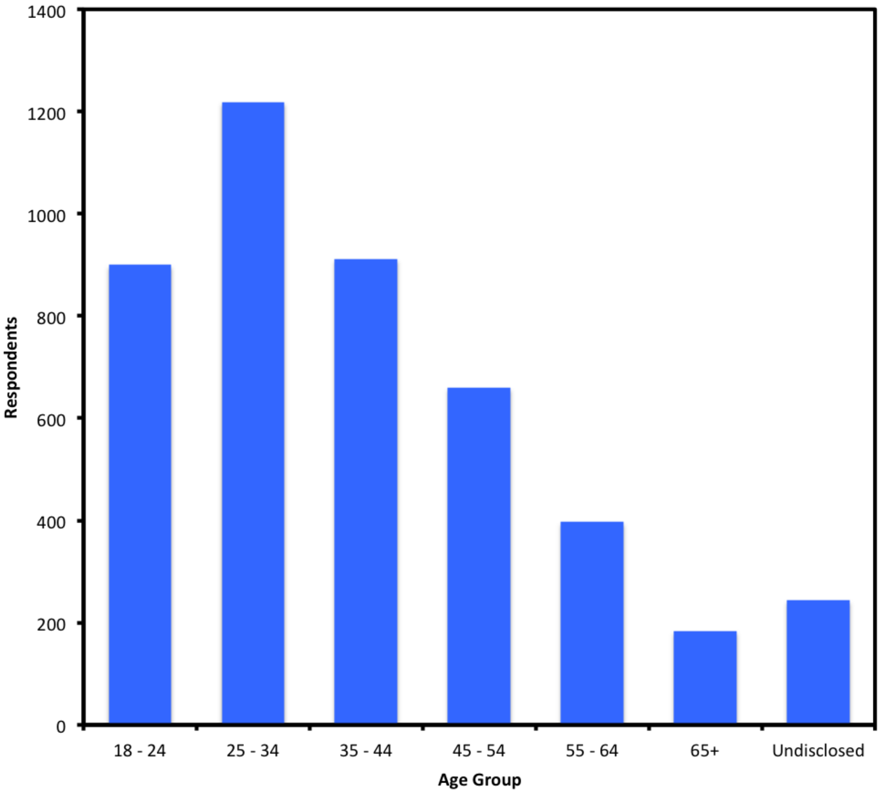


Figure 4-7: Bar graph age breakdown of respondents.

4.3.3 Procedure

Each participant was first provided with an on-screen information sheet setting out the task requirements and explaining their right to withdraw. They were also presented with a consent form to which they were required to affirmatively respond before the questionnaire would begin. Participants were then shown a briefing screen clearly explaining the nature of the vEAR sensation and how it differs from real-world hearing, and asked whether they believed they had ever experienced a phenomenon like vEAR in the past:

> *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]*

Participants were then presented with a brief explanation of the Likert rating scale for the questionnaire items and a description of the sensation to which each end of the scale corresponded. They were also instructed to participate in as silent an environment as possible (see below).

> *As you watch each of the following video clips we would like you to rate them for how intense any associated auditory sensation is. The sensation may be very faint, so you will have to listen carefully. Please try and complete this questionnaire in a silent room, or with as little noise as possible. You may watch the videos as many times as you like before rating them.*
>
> *Please rate the clips from 0 (no auditory sensation at all) to 5 (very vivid and definite auditory sensation).*

The first of 24 randomly presented video clips then appeared on the screen. Under the embedded video was a 6-point scale ranging from 0 to 5. Participants were asked via onscreen text “on a scale of 0 to 5, how much auditory sensation do you experience when viewing this video?” (See Figure 4-8). The clip only began when the participant pressed play, and participants were free to view each clip as many times as they wished. Once a rating had been selected the next screen with the following video clip automatically followed.

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

0	1	2	3	4	5
○	○	○	○	○	○

Figure 4-8: Stimulus rating screen with presentation window and Likert scale

After all 24 videos were rated we asked for some basic demographic information such as age, gender and whether the participants considered themselves to be a synaesthete, according to the explanatory criteria provided:

Synaesthesia is a rare condition where sensation in one sense can cause you to experience sensation in another sense. Examples might include seeing colours when you hear music, always seeing particular letters and numbers in specific colours, or experiencing tastes/smells when you hear or read particular words.

Do you consider yourself to be a synaesthete (somebody who has synaesthesia)?

[Yes / No / Not Sure]

An additional set of questions were appended to the questionnaire after the unexpectedly large sample size presented the opportunity to introduce a number of extra measures in which we had become interested based on the findings of the previous chapter. These were designed specifically to examine whether any other traits that may be characterised by a particularly excitable auditory cortex correlate with item ratings, and in the case of the question relating to hypnagogic phosphenes cortical excitability more generally. These questions can be seen below. As well as adding these new items to the open questionnaire an invitation to complete these additional questions was sent to all participants who had previously taken part and had provided an email address with consent to be contacted about future research. A total of 997 participants completed this extended version of the questionnaire featuring the supplementary questions

Do you suffer from tinnitus? (ringing in your ears)

[Yes / No / Not Sure]

When in the dark or falling asleep do you ever see flashes of light triggered by sudden sounds?

[Yes / No / Not Sure]

Do you ever 'hear' music in your head?

[Never / Rarely / Occasionally / Frequently / Very Frequently]

In everyday life are you ever aware of hearing sounds when you see flashing lights or movement? (E.g. shop displays, car indicators, or people walking?)

[Yes / No / Not Sure]

Do you ever associate certain colours with particular letters or numbers, or with music, or tastes with certain sounds?

[Yes / No / Not Sure]

If you answered yes to the last question, please give more detail.

The survey takes approximately 5 minutes to complete in one sitting with a single viewing of each video. However, participants were free to view each video as many times as they wished, and as they completed the survey in their own home they were free to complete the survey at their leisure, leaving and returning to it as they wished. After 24 hours of inactivity the Qualtrics software logged their results and terminated the session. Completion times ranged from 2 minutes to 23 hours and 26 minutes, with a mean completion time of 56 minutes (SD = .34). The mode completion time was 7 minutes, with 80% of respondents completing in under 14 minutes and 95% completing in under 45 minutes. Histogram presenting completion times in minutes, hours and days are presented in figures 4-9, 4-10, and 4-11.

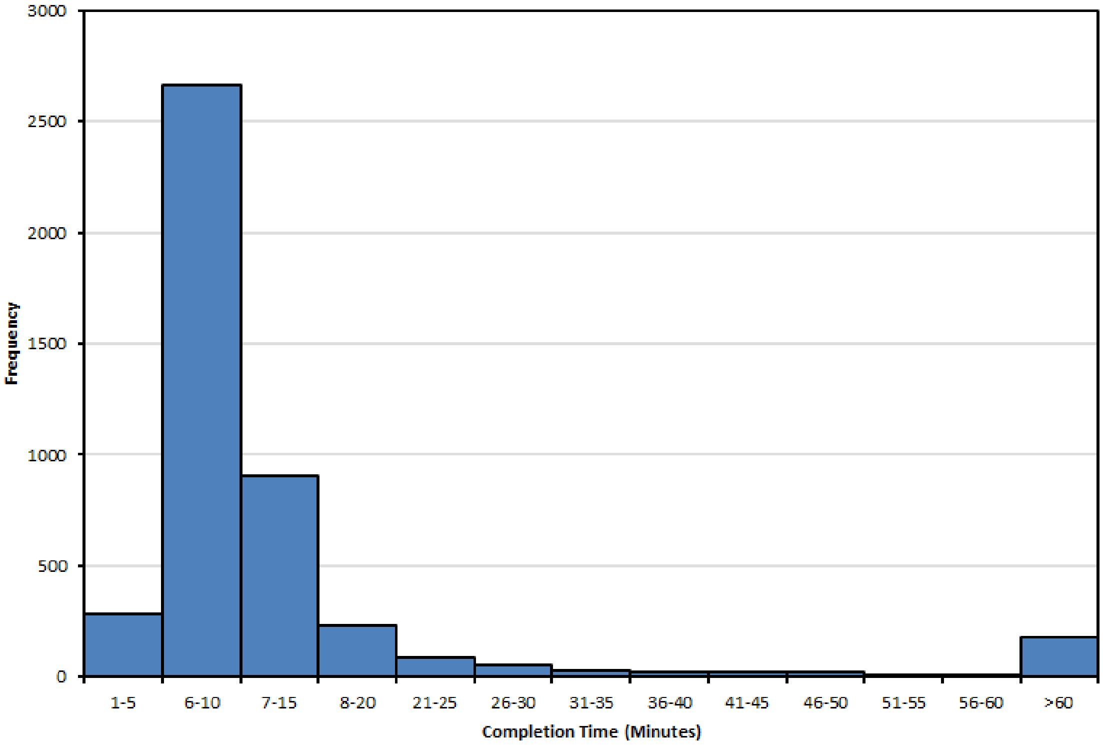


Figure 4-9: Breakdown of Questionnaire Completion Times in Minutes

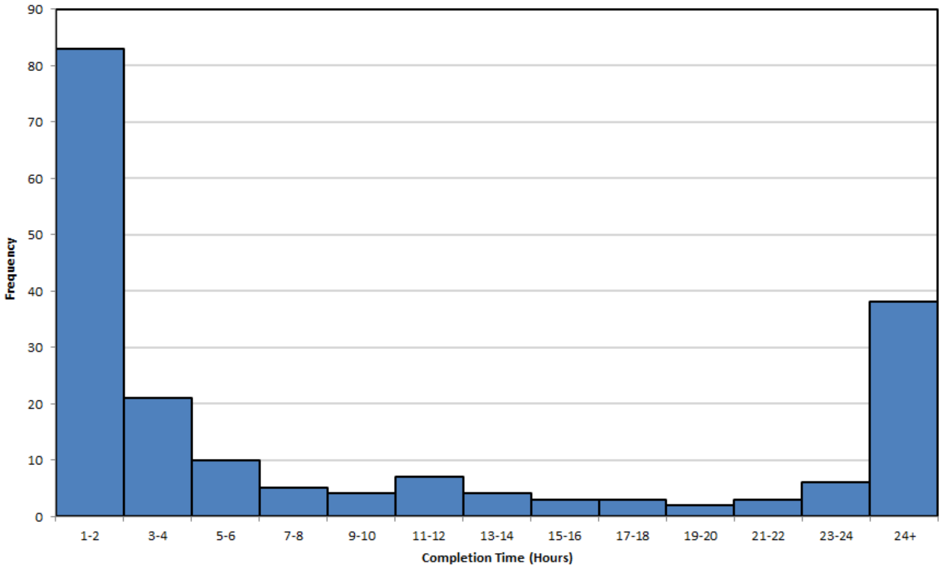


Figure 4-10: Breakdown of Questionnaire Completion Times in Hours

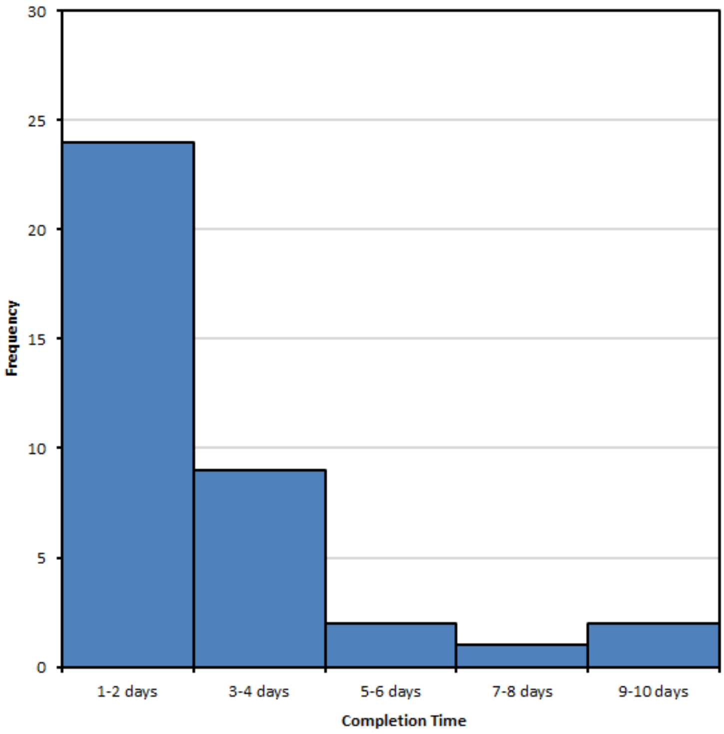


Figure 4-11: Breakdown of Questionnaire Completion Times in Days

4.4 Results

4.4.1 Demographics

In total 478 participants (10.6%) when asked identified as synaesthetes, 3,786 (83.8%) said that they did not or that they were unsure. An additional 252 (5.6%) did not answer this question. These should not be taken as representative of the true population prevalence of synaesthesia, as we specifically targeted synaesthete populations via social media, and press coverage of the topic may have drawn synaesthetes disproportionately to our survey. We did not ask participants their exact age, instead asking them to select which age group they belonged to (see Table 4-1 for full demographic break down).

19.9% of respondents reported before participating that they had previously been aware of hearing internal sounds accompanying visual events consistent with our definition of vEAR. This prevalence is similar to the 22% vEAR prevalence reported in chapter 2, although in that sample participants had not generally been aware of the sensation prior to participation. The more comparable statistic from chapter 2 is the 11% of participants who reported that they had previously been aware of experiencing vEAR. It is therefore likely that the prevalence reported here may be inflated due to self-selection sampling bias. When asked after participating, 10.6% of our sample reported identifying as synaesthetes, which is higher than the typically reported prevalence of 2-4% of the population, although our figure is also likely to be inflated due to a biased sample. A full break down of these two questionnaire items is presented in Table 4-2 and Figure 4-12 and Figure 4-13:.

Demographic	N	Percentage
Synaesthete*		
No/Unsure	3786	83.8
Yes	478	10.6
Undisclosed	252	5.6
Prior awareness of hearing visual events		
No	1761	39.0
Not Sure	1850	41.0
Yes	897	19.9
Undisclosed	8	0.2
Total	4,516	

*By Self-Report

Table 4-2: Percentages of respondents who identify as syneasthetes and/or were previously aware of experiencing vEAR.

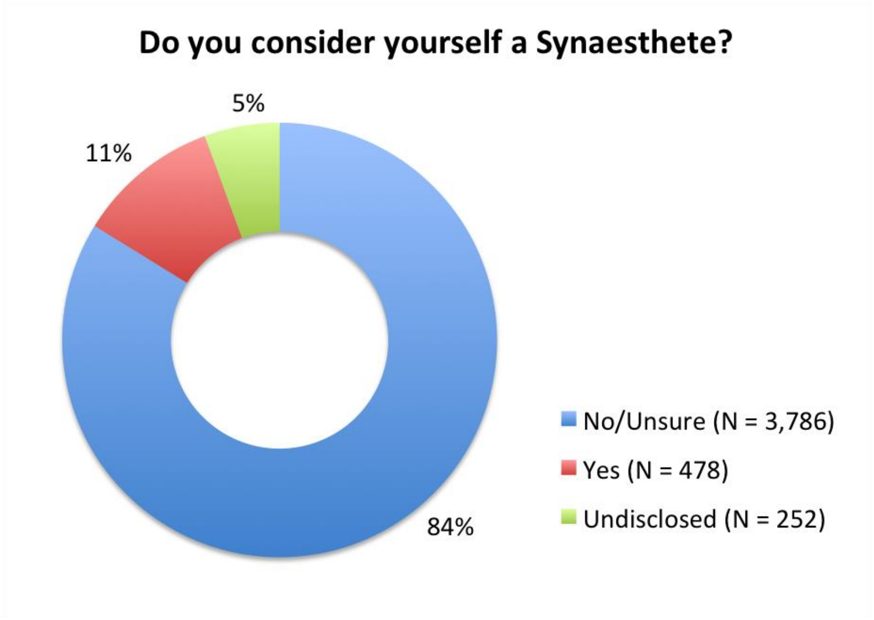


Figure 4-12: Percentages of respondents who identified as synaesthetes.

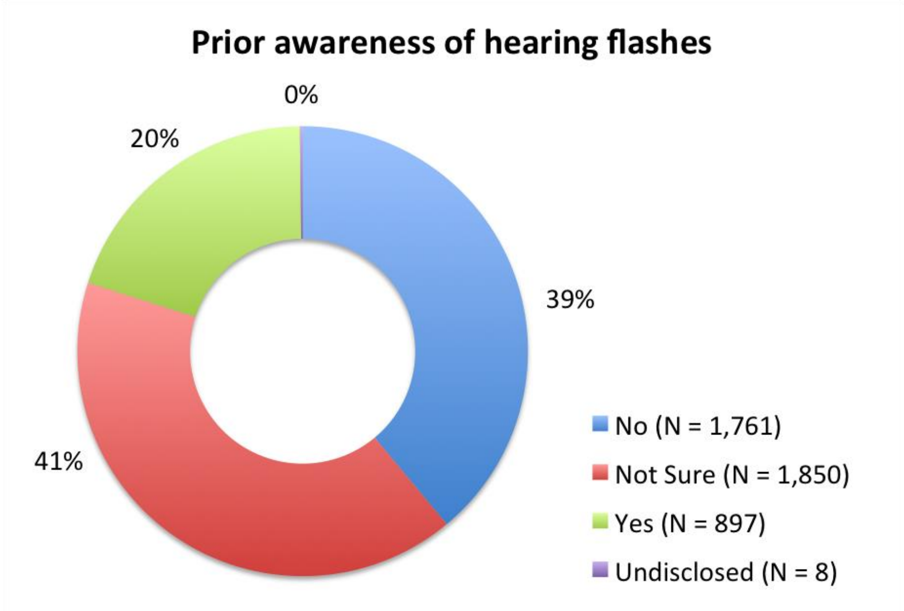


Figure 4-13: Percentages of respondents who were previously aware of hearing visual events.

4.4.2 Crosstabulations

First we performed a serious of crosstabulations to see the relationship between respondents' previous awareness of experiencing vEAR and the other traits potentially predicted by vEAR, which might relate to increased cortical excitability. There was a significant positive association between previous awareness and how frequently one experienced involuntary musical imagery [$\chi^2(8) = 93.79$, $p<0.001$] (see Figure 4-14). There was also a significant association between previous awareness and hypnagogic phosphenes [$\chi^2(4) = 90.89$, $p<0.001$] (see Figure 4-15). There was a significant association between previous awareness and experiencing tinnitus [$\chi^2(4) = 19.95$, $p<0.001$] (see Figure 4-16). Finally, there was a significant association between previous awareness of vEAR and whether a participant experienced synaesthetic associations between the senses [$\chi^2(4) = 245.73$, $p<0.001$] (see Figure 4-17).

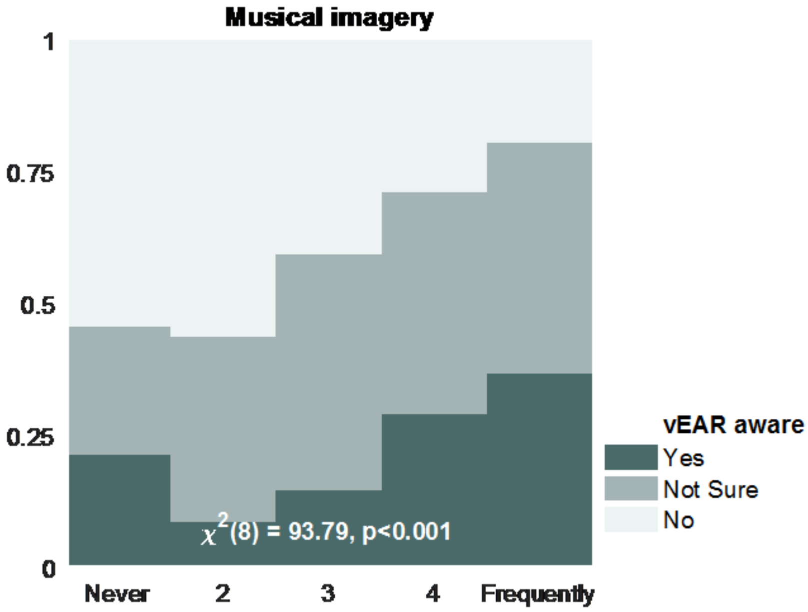


Figure 4-14: Proportions of respondents who report previous awareness of vEAR split by regularity of musical imagery experience.

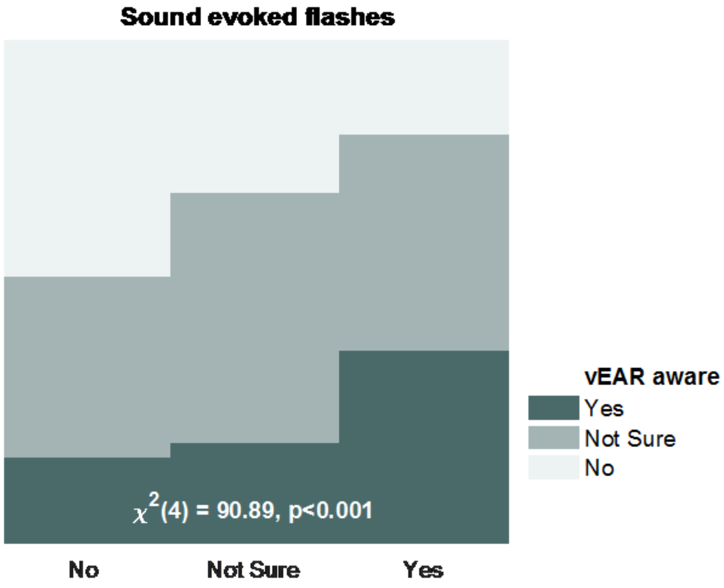


Figure 4-15: Proportions of respondents who report previous awareness of vEAR split by prior experience of hypnagogic phosphenes.

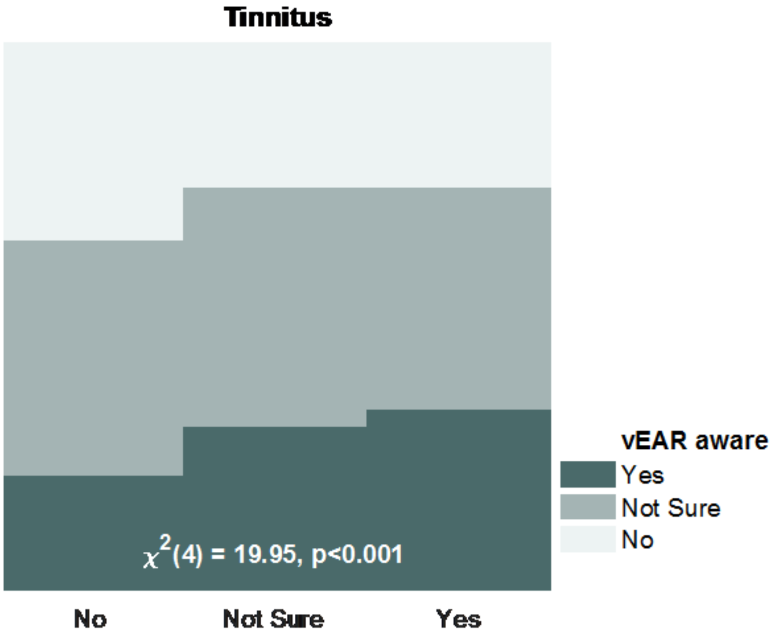


Figure 4-16: Proportions of respondents who report previous awareness of vEAR split by experience of tinnitus.

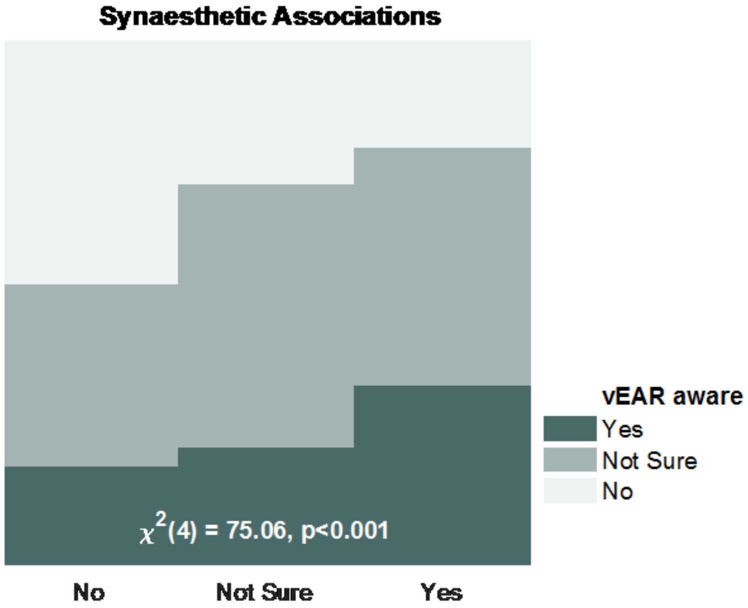


Figure 4-17: Proportions of respondents who report previous awareness of vEAR split by experience of synaesthetic associations

4.4.3 Visual Stimulus Ratings

The mean rating for each video stimulus, based on participants' rating of evoked internal sound from 0 (no awareness of any mental sound) to 5 (a vivid internal auditory sensation), can be seen in Table 4-3 along with measure of standard deviation, and in Figure 4-18. These data are calculated from the ratings of all participants who completed the full questionnaire or who failed to rate no more than 2 items (4,516 participants met these criteria). Data are presented in order of stimuli that evoked the least to the highest scores of internal mental sound (vEAR). The internal consistency of questionnaire items was extremely high with a Cronbach's alpha of 0.97, meaning that our items are positively correlated with one another, suggesting that they are indeed measuring the same phenomenon or experience.

We then broke down the item ratings according to whether or not participants said in response to the first question that they had previously been aware of hearing visual events in the past (see Figure 4-19). A One-way ANOVA revealed that there was a significant effect of previous awareness of experiencing vEAR, with those who had not rating items on average the lowest [M = 0.839], those who were unsure rating them higher [M = 1.555], and those who responded that they had been aware of experiencing vEAR rating items the highest on average [M = 2.323][F(2,4508) = 616.6, p = < .001] (see Figure 4-22). A Tukey HSD Post-hoc Test all three response groups differed significantly from each other at the <0.001 level [No vs Not Sure: Diff=0.7159, 95%CI=0.63 to 0.8, p = <0.001, No vs Yes: Diff=1.48, 95%CI=1.38 to 1.59, p = <0.001, Not Sure vs Yes: Diff=0.77, 95%CI=0.67 to 0.87, p = <0.001]. Pairwise correlations in ratings were all highly significant, (see

Table 4-4). This means that some videos are generally rated higher than others regardless of subjective awareness of vEAR, which suggests that either we all experience some effect of vEAR, but in some individuals the accompanying sounds can be subliminal, or alternatively that ratings are based on imagery and the predictiveness between each video and an accompanying sound. A Pearson product-moment correlation coefficient was calculated to assess the relationship between mean item scores between those who did versus did not report previous awareness of hearing visual events. There was a highly significant correlation between ratings of the two groups [$r = 0.979$, $p = <0.001$].

Stimulus	M	SD
Ballerina	0.710	1.225
Spinning Dot Globe	0.850	1.317
Random Moving Dots	0.853	1.294
Riviera	0.892	1.346
Bumper Cars	1.061	1.406
Disco Lights	1.123	1.452
Tennis Ball Bounce (Slow motion)	1.190	1.441
LED Squares	1.211	1.509
Orange Twinkling Lights	1.283	1.535
Police car Lights & Passing Traffic	1.289	1.520
Multiple Clocks	1.321	1.583
Blue Twinkling Lights	1.344	1.522
Footsteps Beach	1.436	1.531
Golfer	1.497	1.594
Police car Lights	1.501	1.589
Rollercoaster	1.564	1.633
Bouncing Tennis Ball	1.602	1.524
Bouncing Black Balls	1.716	1.633
TV Smash	1.795	1.671
Punch bag Rapid Punches	1.889	1.644
Silent Scream	1.974	1.729
Newton's Cradle	2.035	1.697
Fireworks	2.038	1.755
Hammer Hitting Nail	2.152	1.706
	Overall Mean	SE
	1.430	0.017

Table 4-3: Mean rating (out of a possible 0-5) and standard deviation for each questionnaire item

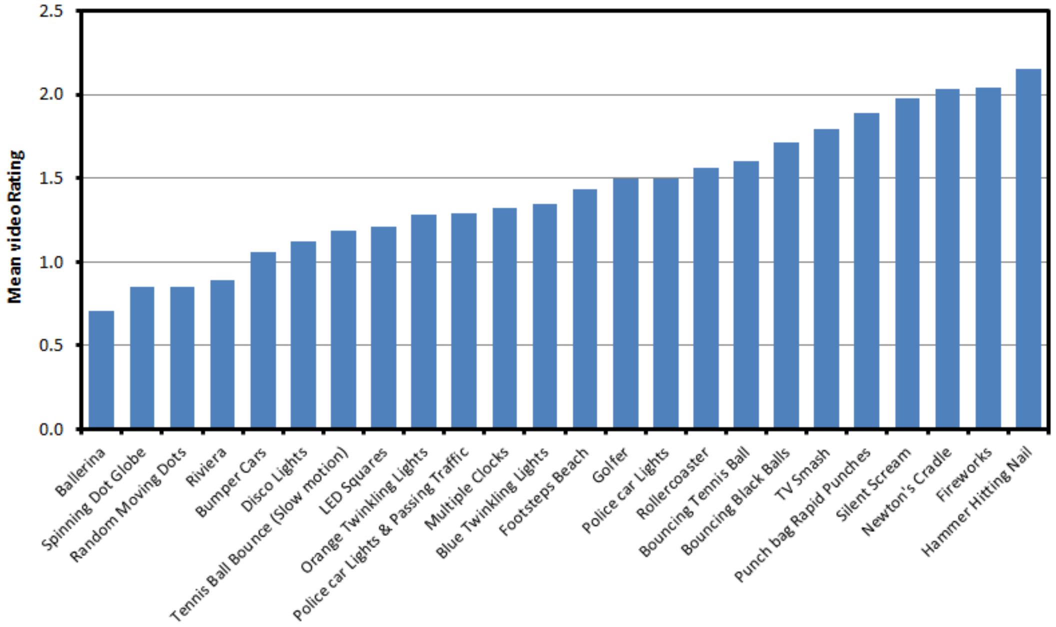


Figure 4-18: Bar chart depicting the mean scores of all participants ranked from lowest to highest rating

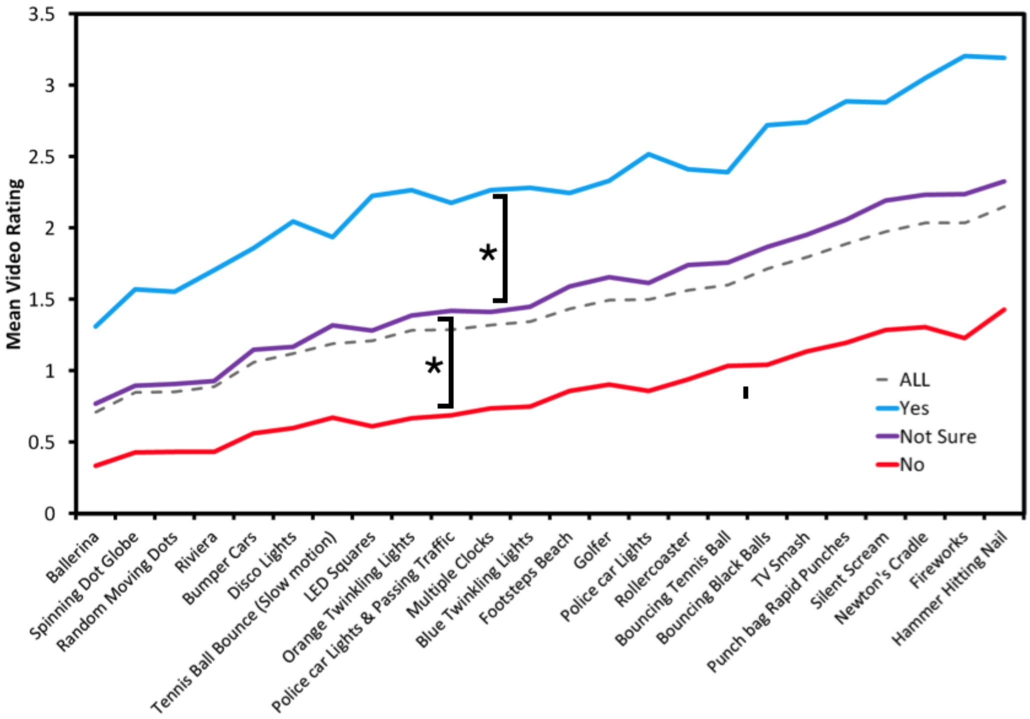


Figure 4-19: Mean rating (out of a possible 0-5) for each video stimulus broken down by previous awareness of hearing visual events, sorted from low to high vEAR rating.

Previously Aware?	Yes	Not Sure	No
Yes	1		
Not Sure	0.978*	1	
No	0.962*	0.994*	1

* Correlation is significant at the .001 level (2-tailed)

Table 4-4: Correlation matrix for mean ratings for each stimulus, split by participant group (previous awareness of experiencing vEAR).

4.4.3.1 Demographic influences on item ratings

We next examined the relationship between the various other traits captured by our questionnaire and item ratings. We first broke down the mean item rating by a number of demographic criteria to explore variations in vEAR within our sample, beginning with a series of t-tests. The first of these revealed that on average female subjects rated videos [M = 1.60] significantly higher than males [M = 1.32], [t(4219) = 7.86, p<0.00, Cohen’s D = 0.24] (see Figure 4-20). Next we explored the effect of synaesthesia on item ratings, with participants who identified as synaesthetes rating videos significantly higher [M = 2.39] than those who did not or were unsure [M = 1.34] [t(4262) = 19.02, p<0.00, Cohen’s D = -0.88] (see Figure 4-19). We then divided participants into those who had previously been aware of experiencing vEAR versus those who had not or were unsure Ratings were significantly higher in participants reporting previous awareness of hearing visual motion [M = 2.36], compared to those said they had no previous awareness [M = 1.07], [t(586) = 14.40, p<0.00, Cohen's D = 1.19] or were unsure [M = 1.77], [t(702) = 6.92, p<0.00, Cohen's D = 0.53]. Participants who were not sure if they had previously awareness of hearing motion rated items higher [M = 1.77] than those had no previous awareness [t(694) = 8.36, p< 0.00, Cohen's D = 0.65] (See Figure 4-19].

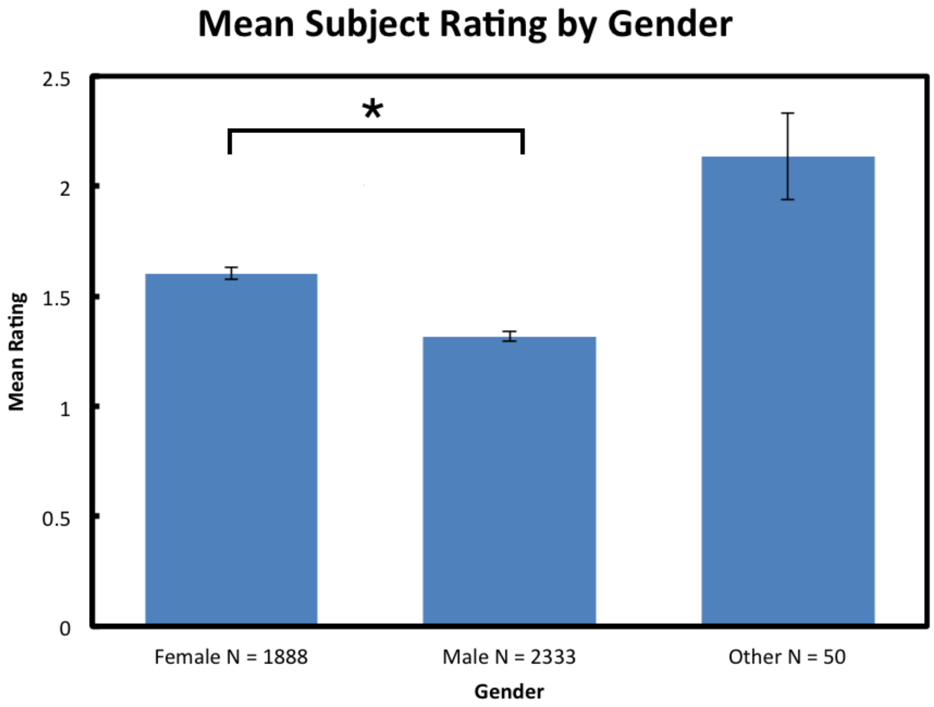


Figure 4-20: Mean item rating by gender with standard error bars.

Female participants rated videos higher on average than male participants. N.B. participants who identified as another gender (e.g. transgender, non-binary) were excluded from this analysis on the basis of the much smaller sample size.

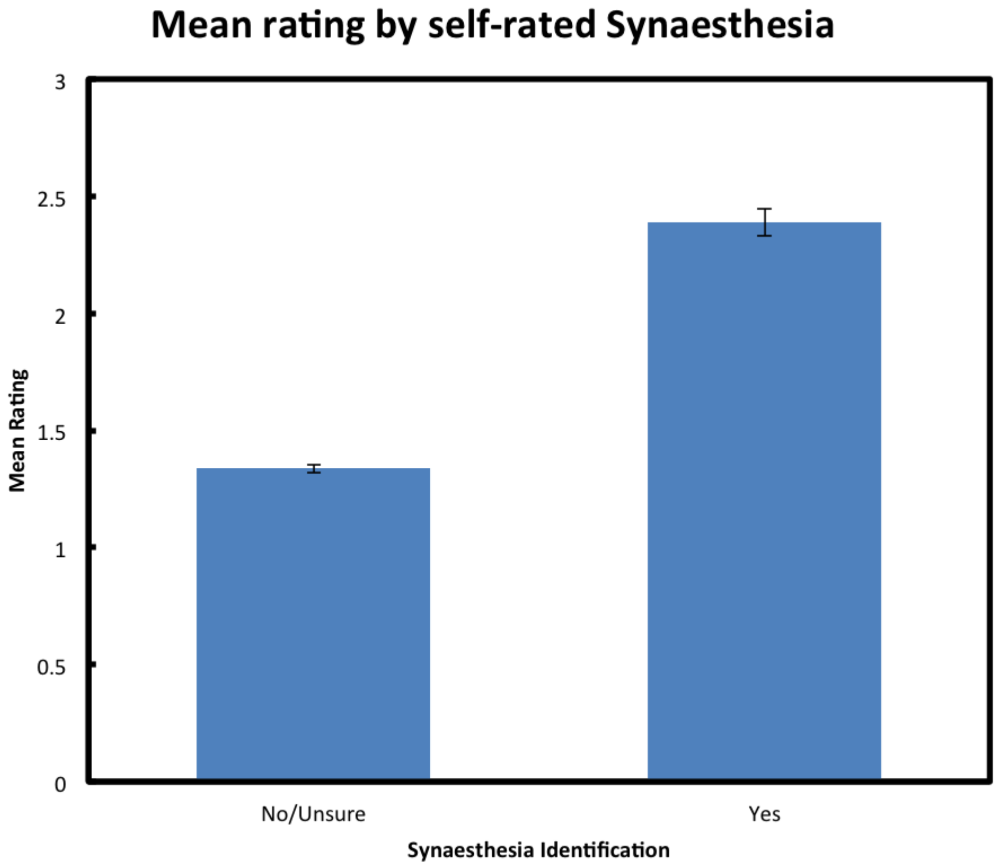


Figure 4-21: Mean video rating by identification as a synaesthete.

Participants who identified as synaesthetes (according to the explanatory criteria presented, see methods) rated videos as evoking significantly higher levels of vEAR on average than those who did not identify as synaesthetes.

There was a significant effect of age group on mean item rating [F(6,4487) = 15.457, p = < 0.001], with a consistent reduction in the mean score of each group. A post-hoc Tukey HSD test revealed that each mean rating declined with each increasing age bracket (outlined in Figure 4-23) and there was no interaction between age and awareness of vEAR [F(12,4487) = .924, p = > .05].

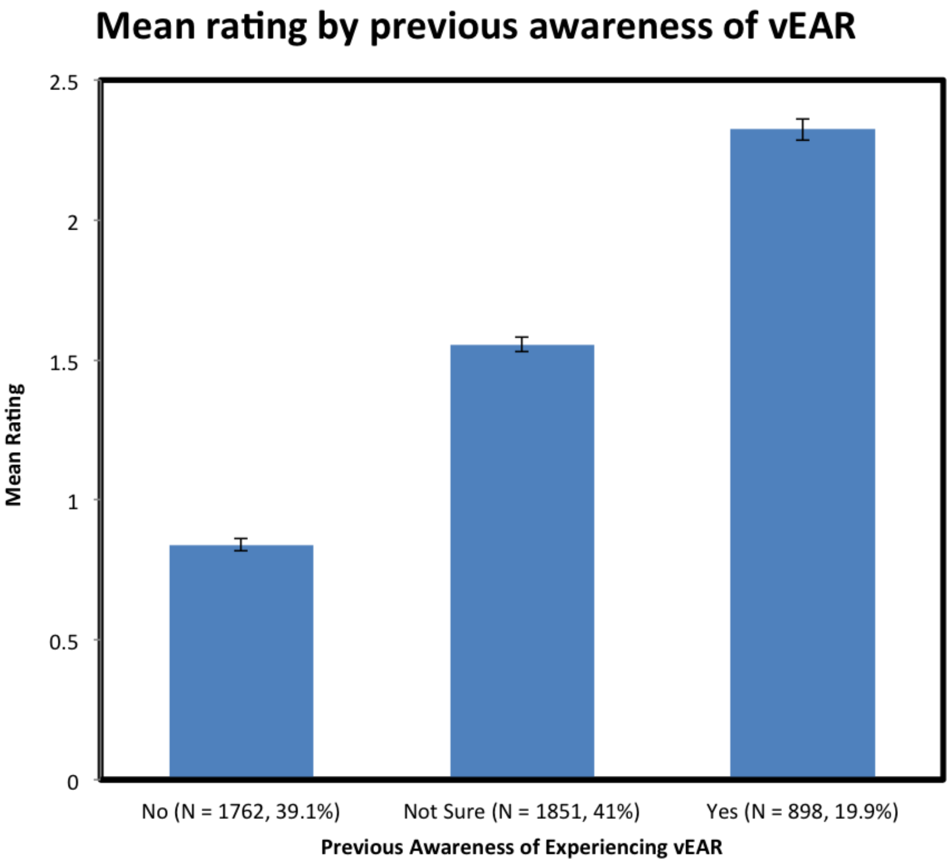


Figure 4-22: Participants rated the videos as evoking more vEAR in proportion to the certainty with which they reported previous awareness of experiencing vEAR. Those who responded ‘yes’ rated the videos higher, meaning greater vEAR, compared to those who responded ‘no’ experiencing the least.

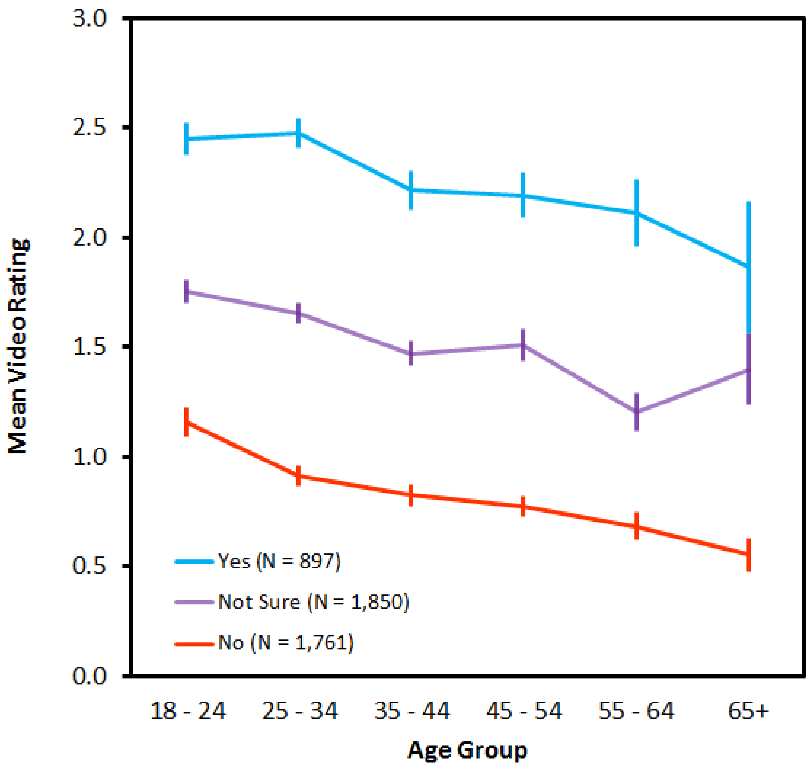


Figure 4-23: Mean ratings declined with age irrespective of whether participants were previously aware of hearing visual events, but the between group differences persisted across age groups.

4.4.4 Supplementary Questions

Next we examined other factors that may be predicted by vEAR. Here we specifically explored areas where we hypothesised that a particularly excitable auditory cortex, or interconnectivity between auditory and visual cortical regions, may mean that sound and vision perception may be modulated by one another. Note that these additional items were added to the questionnaire after an unexpectedly large number of respondents took part in the initial survey, presenting the opportunity to gain more data, hence the following sample size is smaller than the above analyses.

We performed a series of ANOVAs to explore the relationships between our supplementary questions presented in section 4.3.3 and participants' ratings of the vEAR evoked by the video stimuli. The first of these was *'In everyday life are you ever aware of hearing sounds when you see flashing lights or movement? (e.g. shop displays, car indicators, or people walking?)'*. This was included to establish whether for these individuals vEAR is regularly occurring phenomenon, as distinct from the related earlier question. There was a significant difference in mean video rating dependent on participants' answers [F(2,991) = 101.51, p = < 0.001]. A Tukey HSD Post-hoc Test revealed that ratings were significantly lower in those who responded 'no' versus those who responded 'not sure' [Diff=0.6978, 95%CI=0.4996 to 0.8960, p=<0.001] and those who responded 'No' vs those who responded 'Yes' [Diff=1.2900, 95%CI=1.0774 to 1.5027, p= <0.001], and between those who responded 'Not Sure' versus who responded 'Yes' [Diff=0.5922, 95%CI=0.3956 to 0.7889, p= <0.001]. These results demonstrate that mean item ratings increase with participants' certainty about experiencing vEAR in daily life.

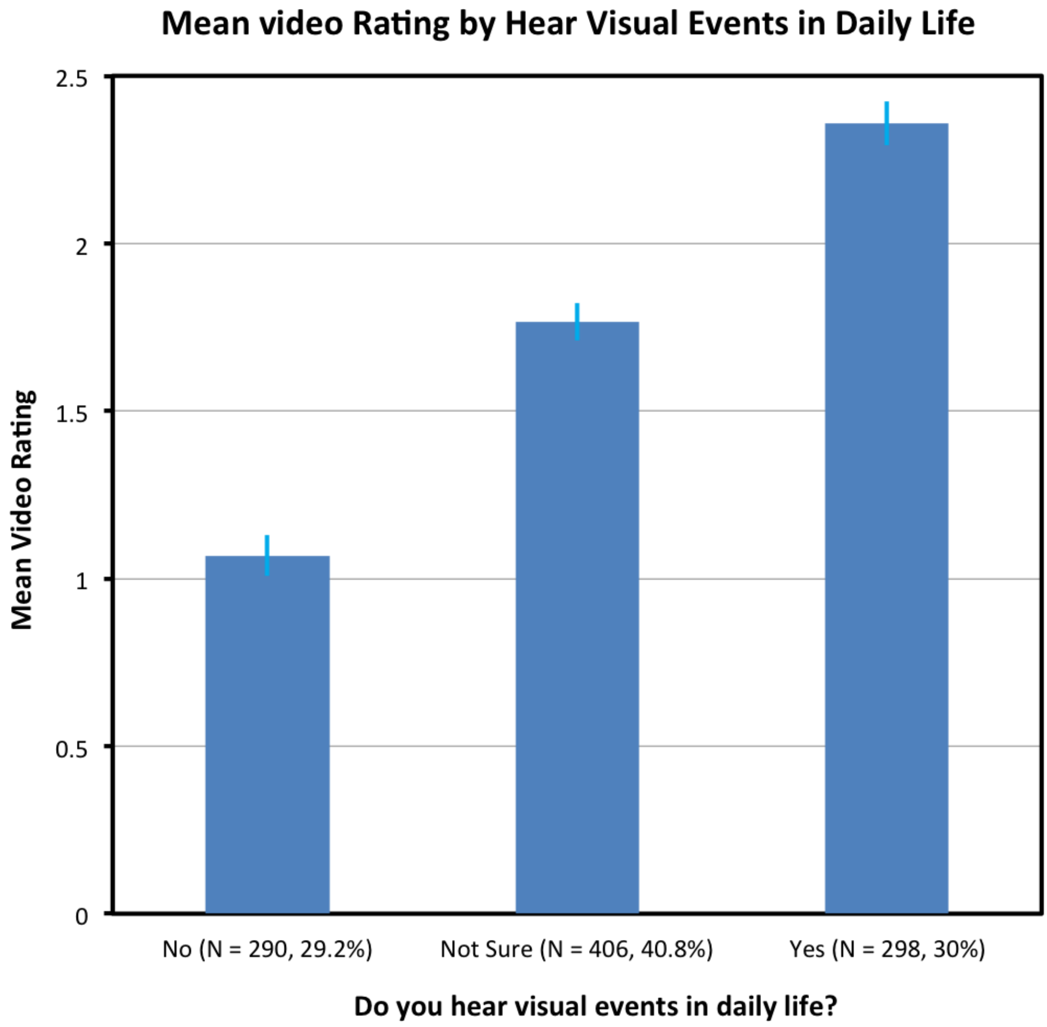


Figure 4-24: Mean item ratings broken down by participants' report of hearing visual events in daily life. Ratings increase significantly in proportion to their certainty of experiencing daily vEAR-like sensations.

In response to the question 'Do you ever associate certain colours with particular letters or numbers, or with music, or tastes with certain sounds?' there was a significant difference in mean video ratings depending on participants' response [F(2,986) = 36.26, p = < 0.001]. A Tukey HSD Post-hoc Test revealed that there was a significant increase in vEAR ratings between those who responded ‘no’ and those who responded ‘not sure’ [Diff=0.3818, 95%CI=0.1364 to 0.6272, p=0.0008], between those who responded ‘no’ versus ‘yes’ [Diff=0.7050, 95%CI=0.5107 to 0.8993, p=0.0000], and those who responded ‘Not Sure’ versus those who responded ‘Yes’ [Diff=0.3232, 95%CI=0.0851 to 0.5614, p=0.0042]. This demonstrates that as participants’ certainty about experiencing sensory pairings typical of synaesthesia increases, so do their ratings of vEAR evoked by the video stimuli.

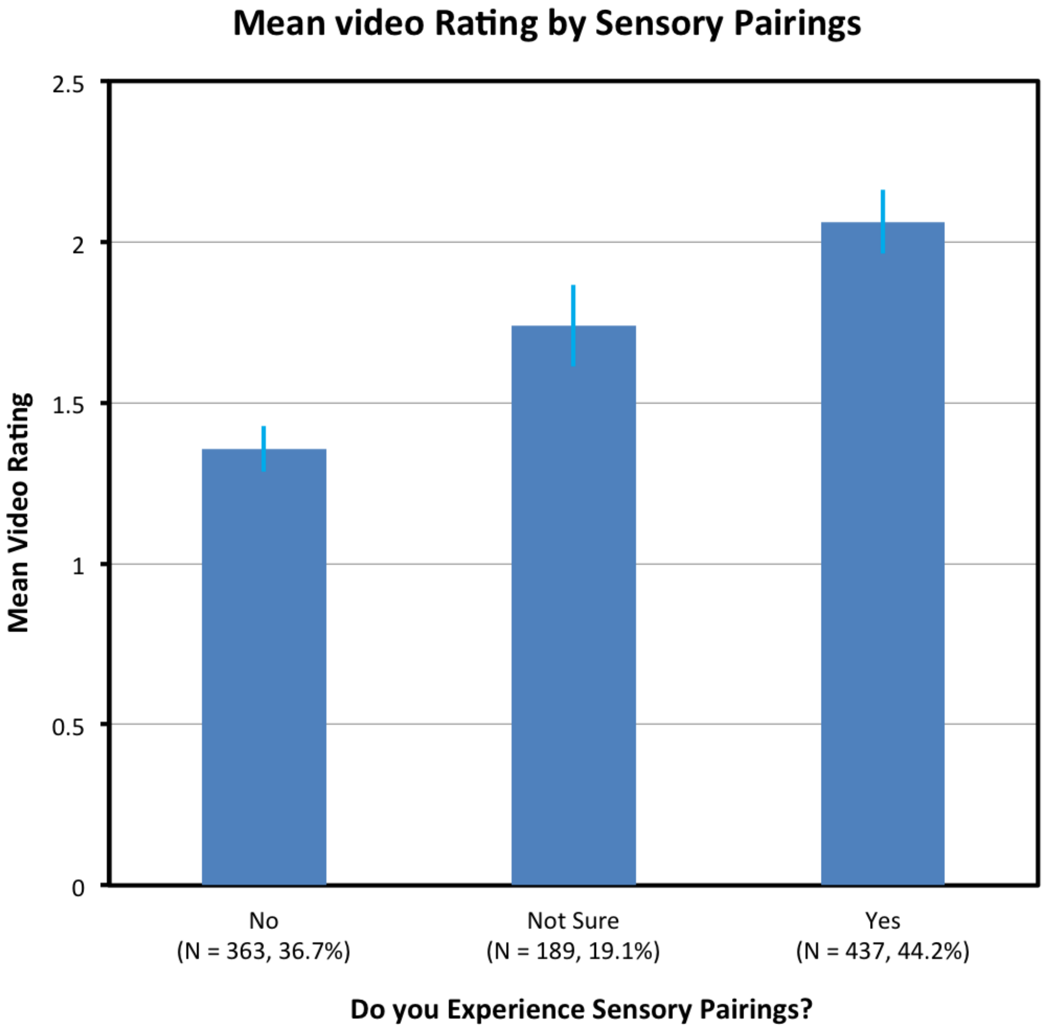


Figure 4-25: Mean item ratings broken down by participants' experience of sensory pairings typical of synaesthesia. Ratings increase significantly in proportion to participants' experience of cross-modal sensory pairings.

In response to the question 'When in the dark or falling asleep do you ever see flashes of light triggered by sudden sounds?' there was a significant effect of the participant's response on their mean item rating [F(2,992) = 36.04, p = < 0.001]. Tukey HSD post-hoc tests revealed that those who responded 'no' scored items on average lower than those who responded 'not sure'

[Diff=0.3722, 95%CI=0.1520 to 0.5925, p= <0.001] or those who responded ‘yes’ [Diff=0.7471, 95%CI=0.5405 to 0.9537, p= <0.001]. Participants who responded ‘Not sure’ rated items lower than those who responded ‘yes’ [Diff=0.3749, 95%CI=0.1613 to 0.5884, p=0.001]. These results demonstrate that as certainty about perceiving nocturnal sound-induced flashes increases so do ratings of vEAR evoked by the videos used in the questionnaire. This suggests that the relationship between sound and flashes may be in part bidirectional.

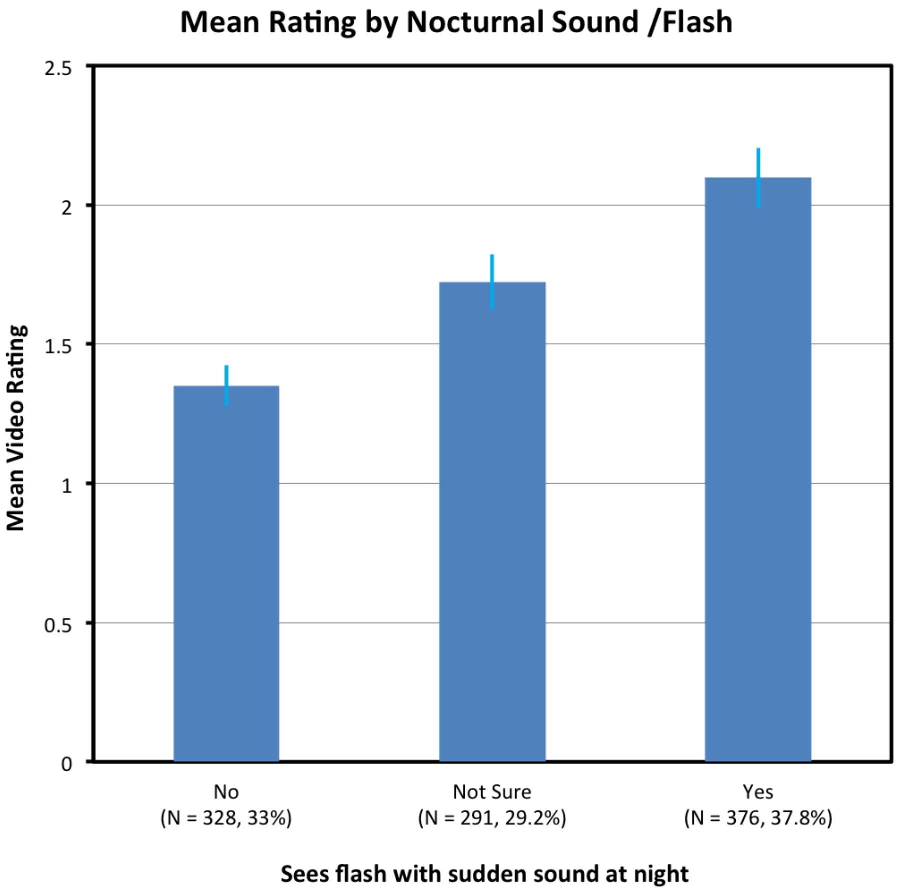


Figure 4-26: Mean video ratings by experience of hypnagogic auditory-induced phosphenes.

There was a significant effect of participants' response to the question 'Do you suffer from tinnitus? (ringing in your ears)' on their mean item rating [F(2,994) = 5.69, p = < 0.05], with a Tukey HSD post-hoc test revealing that while there was no difference between those who responded 'no' relative to those who responded 'yes' [Diff=0.0871, 95%CI=-0.1186 to 0.2929, p=0.5810], there was between those who responded 'No' versus those who responded 'Not sure' [Diff=0.3407, 95%CI=0.1037 to 0.5777, p=0.0022] and between those who responded 'Not sure' versus those who responded 'Yes' [Diff=-0.2536, 95%CI=-0.5125 to 0.0053, p=0.0564]. These results demonstrate that ambiguity over whether or not a participant experiences tinnitus predicts higher reports of vEAR evoked by the video stimuli. This may be because the participant is unsure how to categorise the vEAR sensation, yet they are aware of some kind of unusual auditory percept to which they cannot provide a label.

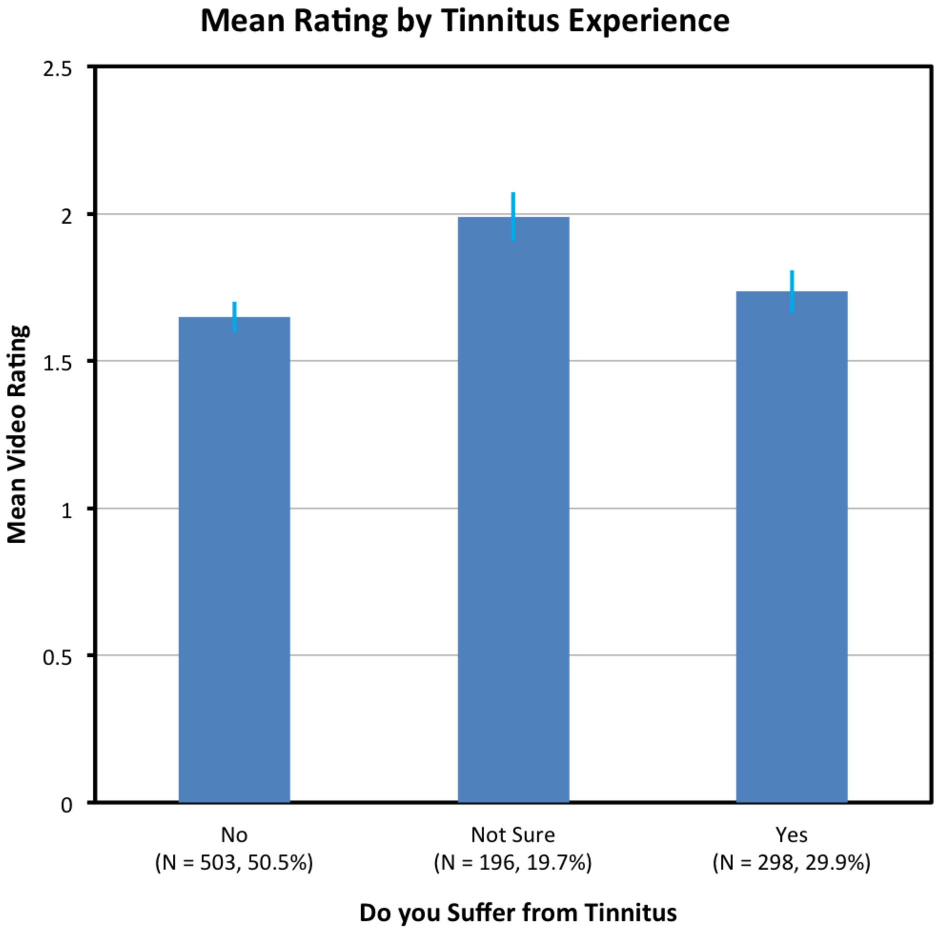


Figure 4-27: Mean video rating by experience of tinnitus

We then performed an ANOVA for responses to the question ‘Do you ever ‘hear’ music in your head?’, a phenomenon more commonly known as experiencing an ‘earworm’ or subsequently here as ‘musical imagery’. There was a significant effect of participants’ answers on their ratings of vEAR evoked by the video stimuli [F(4,991) = 42.43, p = < 0.001]. Tukey HSD post-hoc tests demonstrated that mean ratings increased significantly with reported frequency of experienced musical imagery, with the exception of between those who responded ‘never’ and those who responded ‘rarely’ [Diff=0.3760, 95%CI=-0.2258 to 0.9778, p=0.4302].

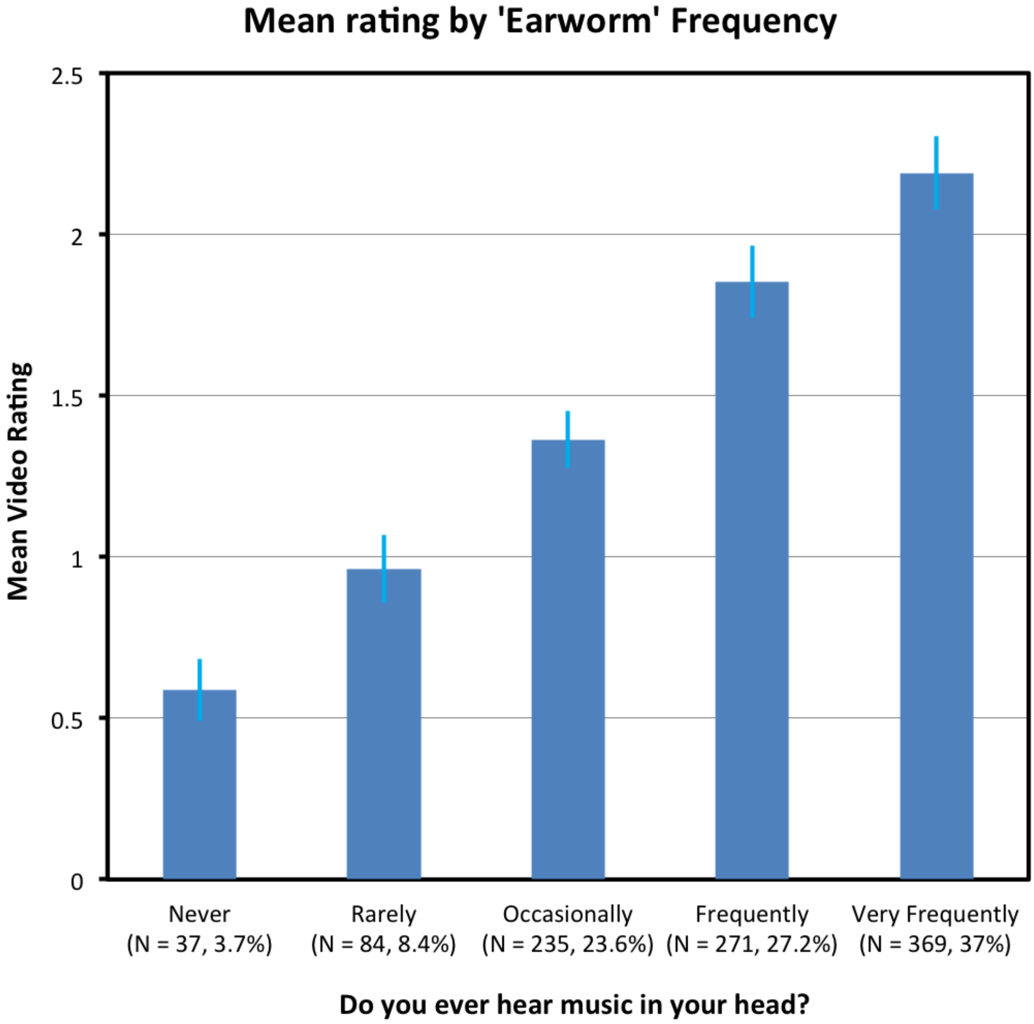


Figure 4-28: Mean video rating by frequency of musical imagery experience

4.4.5 Stimulus Properties

We had proposed two potential properties of visual stimuli that might influence the ratings each video. The first of these was a learned expectation of an accompanying sound, either through learning or some form of semantic connection. Examples of these might be the face of a person screaming. The second was the amount of raw motion energy (ME) contained within a stimulus, regardless of any meaningful content in the video. We predicted that these factors should each affect performance independently, as they relate to more cognitive versus more perceptual processes respectively. We tested this hypothesis using a principle components analysis (PCA) of the average rating for each video.

4.4.5.1 Principle Component Analysis

Principle Component Analysis (PCA) is a statistical tool that is used to transform a dataset comprised of several potentially correlated variables into a group of linearly uncorrelated factors known as principal components. The assumption underlying this process is that these correlated items reflect a set of common broader factors that each independently drive variability within the sample, referred to as the principle components. Here we use PCA to address our two hypotheses for what properties of the visual stimulus predict higher scores for auditory sensation evoked by the visual stimuli: net motion energy or the predictiveness of a sound (through a semantic or learned association). The PCA identified two main PCs consistent with our prediction. The first and second PCs explained 61% and 6.9% of the data respectively (See Figure 4-29).

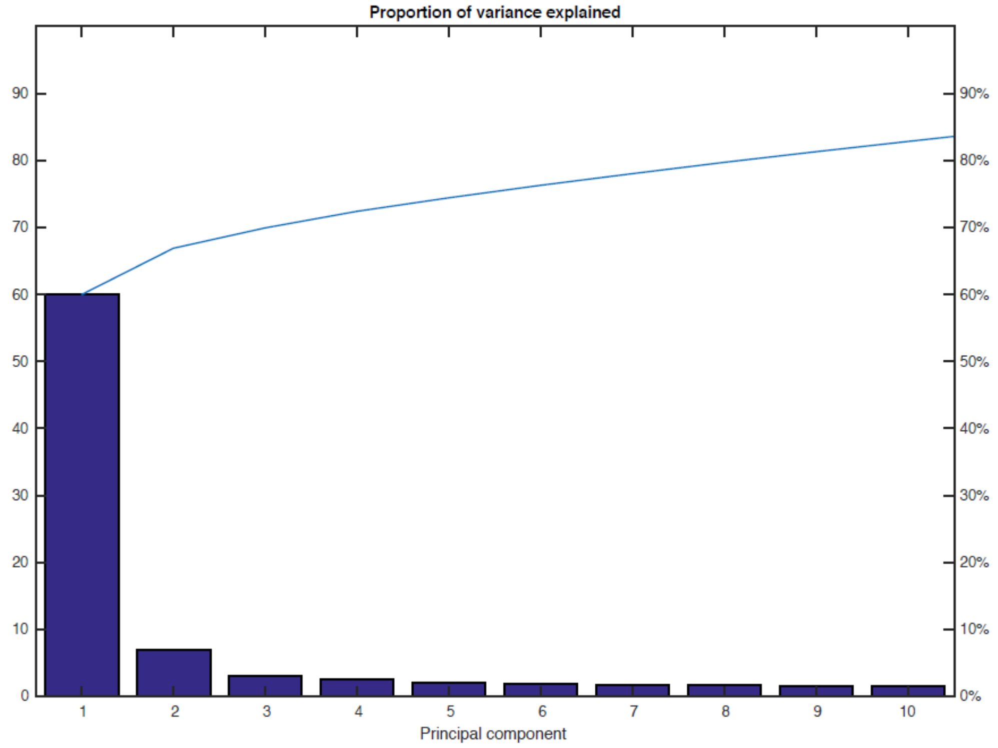


Figure 4-29: Pareto plot depicting results of PCA. Bars displays a scree plot indicating two major PCs. Line indicate proportion of the data explained by each PC. The scree flattens out after third PC.

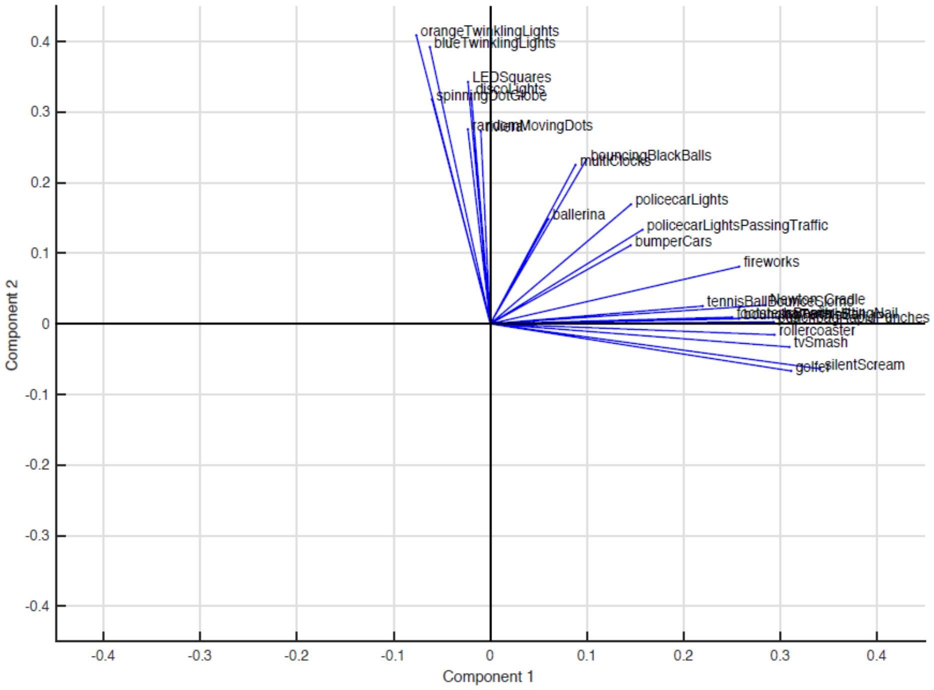


Figure 4-30: Each video plotted by coefficient of PCs 1 and 2.

In Figure 4-30 we can see each video plotted by their coefficient according to each of the PC dimensions. We can see that the horizontal dimension appears to represent those videos depicting motion that is predictive of sound (screams, impacts etc.), while the vertical dimension appears to represent motion energy (flashing LEDs, twinkling lights etc.).We next ranked the videos according to their coefficient score for each PC (see Figure 4-31 and Figure 4-32). Inspecting these rankings we observed that items that scored high on PC1 tended to be those with meaningful content that was highly predictive of an accompanying sound, such as a screaming face, fireworks exploding, and impacts such as a hammer hitting a nail. Items that scored high for PC2 conversely tended to be more abstract moving patterns, such as digital animations depicting flickering lights or flashing neon signs. Videos ranked lower on both PCs tended to depict real-world scenes in rapid motion, which might also be associated with sounds, such as dancing, police car lights, and fireworks.

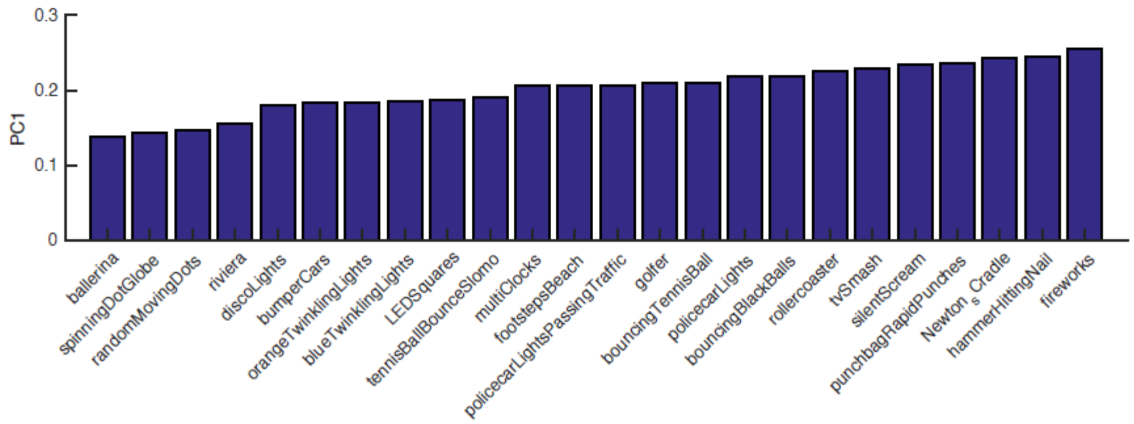


Figure 4-31: Questionnaire items ranked by coefficient for Principle Component 1.

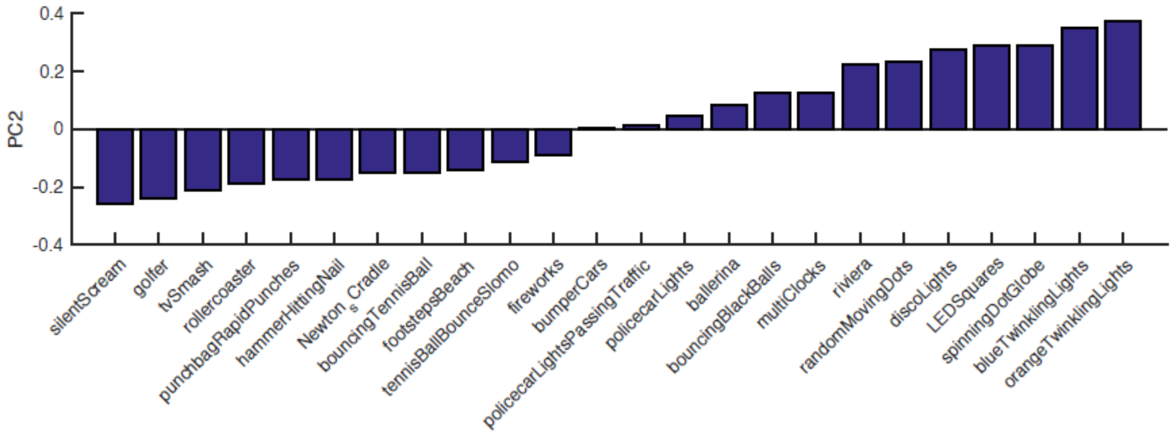


Figure 4-32: Questionnaire items ranked by coefficient for Principle Component 2.

4.4.5.2 Motion Energy Analysis

In order to verify that PCs did indeed tend to depict high contrast items depicting high levels of motion energy (ME) we ran an analysis to produce an objective measure of the ME contained within each video stimulus. To achieve this we used a publically available Matlab script (Mather, 2013), used to model the motion processing of complex cells within the visual system, modified from Adelson & Bergen (1985). For each image sequence, motion energy is calculated separately for left-right motion and up-down motion, and the results are averaged. First, grey-level values are first averaged across one spatial direction (e.g. vertical), to create a two-dimensional space-time matrix. This matrix is then convolved with four linear filters based on Gabor functions that extract two speeds of motion, in two opposite directions (e.g. left vs right). Each filter output is rectified and normalised relative to the other filters, and opposite motion directions are then subtracted from each other. The analysis then is repeated for the same image sequence after first averaging across the other spatial dimension (e.g. horizontal). The result of these two analyses is then averaged to produce a single number which summarizes how much motion energy there is, in any direction, in the whole image sequence. The individual ME scores for each video stimulus can be seen in Appendix 1.

To test whether people who report experiencing vEAR tend to give higher ratings to videos that contain high levels of ME we next performed a median split on the videos according to the ME each contained, dividing them into high versus low ME groupings. We divided the sample by whether they reported experience vEAR or not, performed a 2x2 ANOVA on the data. Results showed a significant interaction between awareness and ME [F(1,4123) = 73.09, p<0.0001, $\eta^2_{partial}$ = 0.049], as well as significant main effects of both awareness of vEAR [F(1,4123) = 720.94, p<0.0001, $\eta^2_{partial}$ = 0.149], and Motion Energy [F(1,4123) = 213.15, p=0.0001, $\eta^2_{partial}$ < 0.001], confirming that all participants regardless of vEAR rated high ME videos higher than low ME videos, and that this effect was significantly larger in those respondents who had previously been aware of experiencing vEAR (see Figure 4-33).

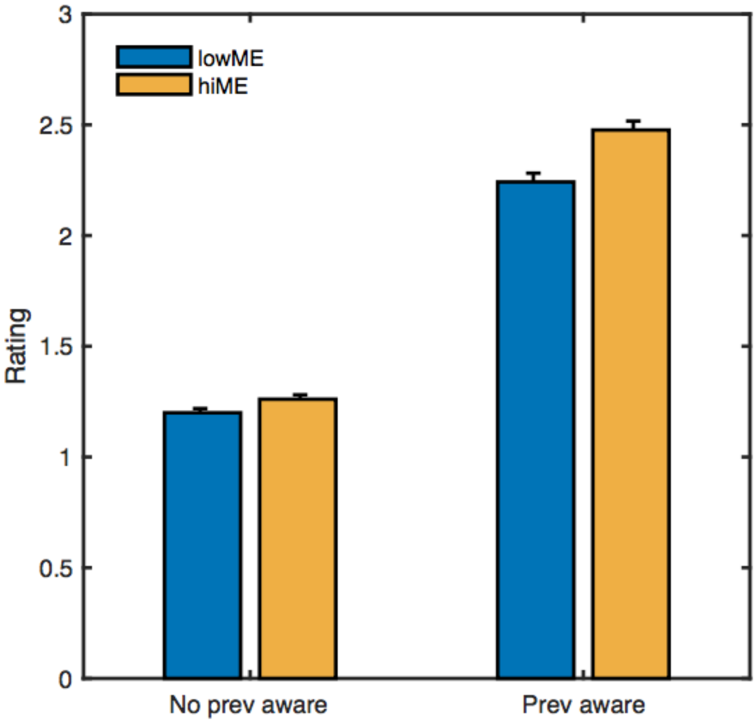


Figure 4-33: Mean video ratings split by high/low ME and previous awareness of vEAR (yes/no)

To complement this analysis we next used the results of the PCA results to predict the ratings for each video stimulus as a function of ME and also of previous awareness of hearing visual events. We reconstructed ratings for each video based on PC2 and higher components, but crucially we eliminated PC1, and split these data by awareness group, before averaging over participants. Reconstructed ratings averaged for each stimulus correlated strongly and positively with ME for the 'Yes' awareness group [r(21) = .57, p = <.005], but more weakly and negatively with the No [r(21) = -.41, p = <.052] and Not Sure [r(21) = -.43, p = <.040] groups (see Figure 4-34). As seen in Figure 4-35, no significant correlations were observed when ratings were reconstructed from PC1 instead (excluding PC2).

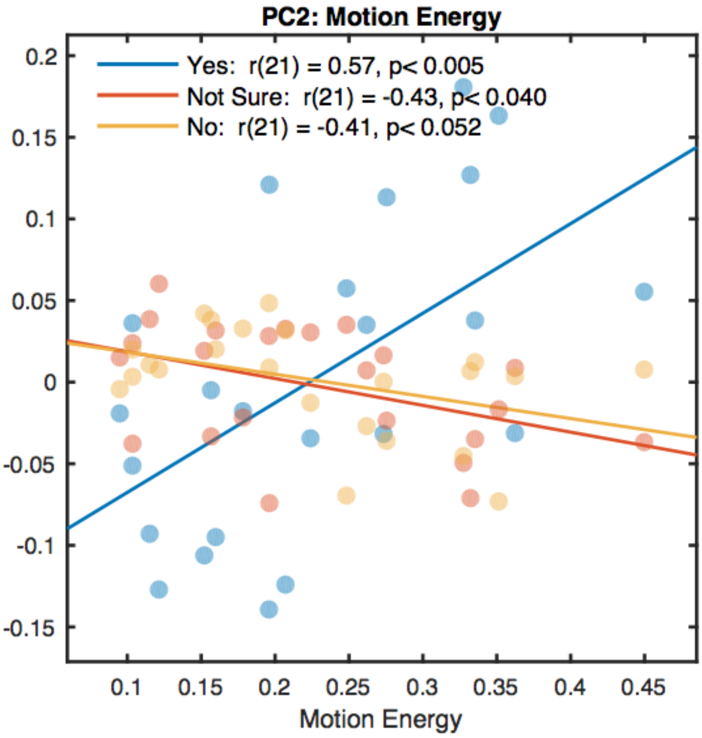


Figure 4-34: Averaged ratings for each video reconstructed from PC2: Only strong correlation is with those who report previous awareness of hearing visual events.

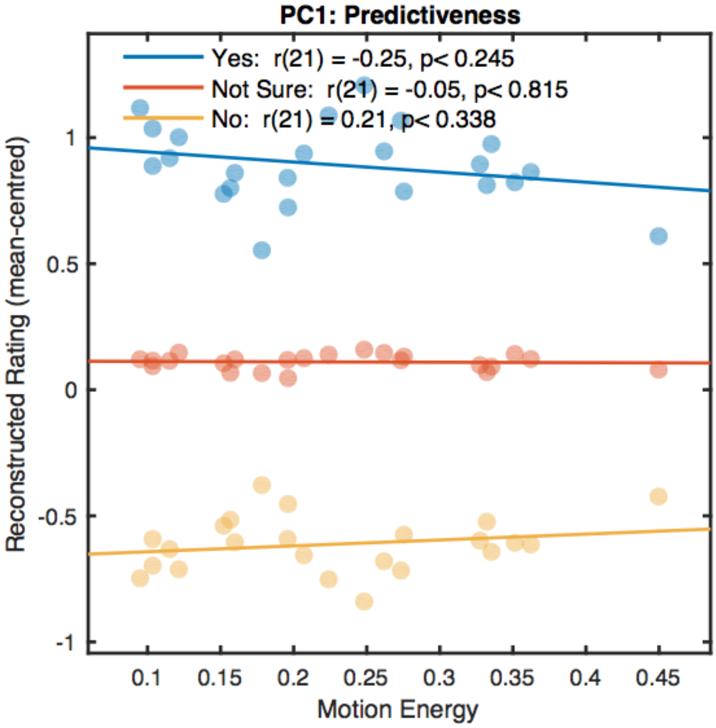


Figure 4-35: Averaged ratings for each video reconstructed from PC1: No correlations with any awareness group were observed.

4.5 Discussion

These findings represent the first ever large-scale exploration of vEAR in the general population, adding considerably to our understanding of a previously little-studied phenomenon in a number of ways. Firstly, we can now begin to examine the individual trait factors of persons who experience particularly vivid experiences of vEAR, compared to those who experience very little or none. Secondly, having moved away from rudimentary psychophysical stimuli such as beeps and flashes to more 'real-world' stimuli we can now begin to better understand specifically which properties of a visual stimulus best evokes the sensation of vEAR, allowing us to build on the tentative speculation about the underlying mechanisms that we began to explore in chapter 3.

In chapter 2 we reported that 22% of our experimental sample, when asked in an unexpected debrief questionnaire post-task completion, that they had been aware of hearing some faint auditory sensations in their head when viewing flashes of light presented on a monitor as part of a visual Sequence Discrimination Task. In the present study, 21% of questionnaire respondents responded ‘Yes’ to the question asking whether they had previously been aware of hearing internal sounds evoked by visual motion (asked prior to completing the questionnaire). As an estimate of vEAR prevalence in the general population, these results must obviously be interpreted with caution, given that participants were self-selected after following a URL in an online article in the popular press about visually-evoked auditory sensations. We therefore may have a sample biased toward finding out more about a mysterious condition they already know themselves to have. It should also be noted that in our previous random sample only 11% of participants had been aware of experiencing these visually-evoked auditory sensations prior to their attention being drawn to them in a introspectively focused manner under laboratory conditions, while the online questionnaire respondents indicated that they had indeed been aware of the experience before after being presented with a prompt outlining a detailed description of the vEAR phenomenon.

The figure of 21% prevalence of vEAR found in the present study is based only upon those respondents who completed the questionnaire in full, although because the question inquiring about previous experience of vEAR was asked at the outset of the questionnaire we are able to report the prevalence of the overall respondent population (N = 33,504) regardless of completion status. Of this larger sample, 16% reported that they had previously been aware of hearing visually-evoked auditory sensations. Including these non-completing respondents may help to counteract the potential for a sampling bias described above, because it is possible that a proportion of those who failed to complete the questionnaire did so because it became apparent that they were not experiencing any such auditory sensations. Other reasons for non-completion might include technical issues such as poor internet connectivity or compatibility with the flash video. We therefore suggest that a truer prevalence of vEAR in the sample may lie between the 11-21% estimates.

Our analysis of the visual properties of the video stimuli shows that ratings were generally higher in videos which depict events that are predictive of naturally associated sounds, such as objects colliding with each other. However, as we predicted, item ratings were also independently influenced by the objectively measured motion energy contained within the visual stimulus, which is present even in abstract moving patterns that are not predictive of a real-world sound. This influence of motion energy as a factor was significantly greater in participants who reported that they had previously been aware of hearing visual events. This suggests that in these individuals the experience of vEAR reflects a low-level cross-sensory experience in addition to simply a top-down form of learned auditory imagery.

We are now able to make some first tentative conclusions about the types of people who experience vEAR and certain common traits they might share. We added a series of follow-up questions to our initial battery that included questions about respondents' experience of tinnitus, musical imagery (i.e. earworms) and hypnagogic synaesthesia (a little-studied phenomenon whereby individuals report that sudden unexpected loud noises when falling asleep can trigger visual phosphenes). These items were added as we were developing a hypothesis that vEAR may be due in part to an unusually excitable auditory cortex, or conversely reduced inhibition in cortical areas responding to sensory inputs, and these other traits may share this aetiology.

There is evidence suggesting that earworms, or songs that we 'hear in our head', might arise from spontaneous activity in auditory cortical areas, particularly following hearing loss (Kumar et al., 2014), and some cases of tinnitus may be caused by disinhibition of auditory cortex (Kaltenbach, 2011). Our results show positive associations between these traits and the video stimulus ratings. Interestingly the high proportion of our sample (39%) who reported experiencing hypnogogic sound-evoked visual phosphenes is to our knowledge the first report of the prevalence of this experience outside of a clinical population (for examples in individuals with visual impairment see Afra et al., 2012; Jacobs, Karpik, Bozian, & Gøthgen, 1981). It was also associated with awareness of vEAR, and thus suggests that there may be a degree of bidirectionality involved, as those who hear sounds when they see flashes were more inclined, under certain circumstances, to see flashes when they hear sounds. This phenomenon might arise due to enhanced excitability of visual cortex during light deprivation which may unmask input from connections from outside visual areas (Boroojerdi et al., 2000). Taken together these associated traits suggest that individual differences in levels of inhibition of sensory areas may represent a common framework for explaining the phenomenon of vEAR (e.g. Grossenbacher & Lovelace, 2001; Neufeld et al., 2012), and its association with this variety of other traits.

Based on these results we conclude that vEAR can evoked by low-level objective stimulus properties such as motion energy but also is influenced by higher-level experience-based imagery and expectation. Following our findings in chapter 3, we suggested that disinhibition of crosstalk between the auditory and visual cortices may result in the experience of vEAR. Based on the present findings we suggest that this lack of inhibition represents a higher baseline level of auditory cortical activity which in those individuals who experience vEAR may also result in a variety of conscious experiences such as hearing music or seeing flashes evoked by sound. These findings are in keeping with theories that have been proposed to explain phenomena such as grapheme-colour synaesthesia (Terhune et al., 2011) and mirror-touch synaesthesia (Banissy & Ward, 2013; Bolognini et al., 2013; Ward & Banissy, 2015) that may be characterised by an unusually excitable visual and somatosensory cortex respectively. In the latter case, TDCS when applied to somatosensory cortex has even been able to increase cortical excitability to a sufficient level to induce mirror-touch synaesthesia in non-synaesthetes (Bolognini et al., 2013).

4.6 Summary

In summary, in the present chapter we have confirmed via a large online survey that there are indeed a substantial number of people who believe that they do experience vEAR, with the caveats discussed above regarding an inherent sampling bias in our sample. We have demonstrated that vEAR may be slightly more prevalent in females than males, as is often asserted to be the case with canonical synaesthesias (although see Simner & Carmichael, 2015) and experience declines with age. Our results suggest that other personal characteristics which predict experience of vEAR include synaesthesia, experience of hypnagogic sound-induced phosphenes, and frequent experience of involuntary musical imagery. We have also identified two factors that seem to be driving higher ratings of vEAR evoked by our stimuli, the first of these being a learnt predictiveness of sound, which explained 61% of our data, and a second which was the motion energy (ME) contained within a stimulus, which explained 6.9%. We also demonstrated that while all participants tended to rate videos containing high levels of ME higher than those with low ME, this tendency was greater in respondents who were previously aware of experiencing vEAR. These results suggest that ratings in all respondents are driven largely by the predictiveness of a sound, a higher-level phenomenon, while those who experience vEAR also receive low-level sensory crossover at a pre-attentive level that specifically influences their video ratings. In the following chapter we will collate the findings from our last three chapters and outline how they may inform some of the debates outlined in chapter 1, as well as propose a model of vEAR that encompasses these findings, and identify opportunities for future research.

General Discussion

Chapter 5:

5.1 Overview

In this thesis we have explored the Visually-Evoked Auditory Response (vEAR), a phenomenon characterised by the ability of certain individuals to hear mental sounds accompanying dynamic visual stimuli. In this final chapter, we first summarise the experimental findings reported in Chapters 2 to 4. We then return to the contemporary arguments in synaesthesia research presented in Chapter 1, before identifying themes emerging from this thesis as a whole and how they may inform these debates. Discussions include whether vEAR itself meets the necessary criteria to be considered a form of synaesthesia. Finally, we then identify any questions that remain unanswered or that have come to light because of the findings reported in this thesis.

5.1.1 Chapter 2 Summary

Chapter 2 began by outlining a 2008 study by Saenz and Koch that described a hitherto unreported sub-type of synaesthesia in which individuals hear synaesthetic sounds when viewing visual motion. Both the synaesthete and control groups showed relatively strong sequence discrimination ability in the auditory domain, due to our better temporal acuity in this modality, but the synaesthete cohort significantly outperformed controls on discrimination ability in the visual domain. The authors present this as evidence for the veracity of the synaesthetes claims to hear visual events, as any synchronous auditory concurrents would effectively render the visual trials bimodal, with the increased temporal acuity in the auditory compared to the visual domain accounting for their better performance compared to controls. By using a randomly selected group of participants we are able to provide a first tentative estimate at the prevalence of vEAR at around one in five. Like Saenz and Koch, we reported that those who were able to hear the visual events significantly outperformed those who did not on a measure of visual sequence discrimination ability. This ability was negatively associated with auditory detection ability when the auditory target co-occurred with an irrelevant visual stimulus, following Lovelace, Stein and Wallace (2003). This suggests that the visually-evoked auditory sensations disrupted detection of real-world auditory signals. Scores on both tasks were independent of a third task, assessing visual dominance over audition following Colavita (1974), which suggests this phenomenon is not explained by an attentional bias toward vision.

5.1.2 Chapter 3 Summary

In chapter 3 we began to explore the physiological differences between those who do versus do not experience vEAR. To achieve this we employed Transcranial Alternating Current Stimulation (TACS) to disrupt normal cortical processing in the temporal versus occipital lobes while participants performed the Sequence Discrimination Task outlined in Chapter 2. By repeating the debrief questionnaire in this new sample we were able to divide our participants into those who do versus do not experience vEAR, and thus examine how TACS differentially affected task performance across these two groups. Our results demonstrated that in those individuals who did not experience vEAR, TACS impaired sequence discrimination ability in the modality relevant to the site to which stimulation was delivered (i.e. visual sequences with occipital TACS, auditory with temporal TACS), but improved performance in the other modality.

5.1.3 Chapter 4 Summary

In Chapter 4 we moved away from rudimentary psychophysical stimuli to explore a range of more naturalistic visual stimuli, and some more abstract forms of motion, to examine what specific types of visual stimulus best evoke vEAR. The motivation behind this was to test whether it was primarily a higher level process that drives this phenomenon, such as a learned expectation and prediction of an accompanying sound, and/or an objective physical property of the stimulus, such as motion energy. This in turn allows us to ascertain whether vEAR is likely to be a low-level phenomenon that occurs early in the sensory processing hierarchy, or a higher-level process featuring a degree of executive functioning. We reported the results of our large online survey that confirms first that there is indeed a substantial portion of the population that believes that they do experience vEAR, although our attempts to extrapolate from this an estimate of the prevalence of vEAR in the population must necessarily remain cautious due to the likelihood of a sampling bias. We also asked respondents about a number of demographic and trait details that we predicted might correlate with vEAR, as they might share a possible characteristic aetiology in an increased cortical excitability, such as propensity to experience involuntary musical imagery, or tinnitus.

Our Principle Component Analysis (PCA) identified two factors that seem to be driving ratings of vEAR evoked by our stimuli, the first of these being a predictiveness of sound, which explained

61% of our data, and a second which was the motion energy (ME) contained within a stimulus, which explained 6.9%. We demonstrated that predictiveness of a sound, a higher-level influence, was a factor driving ratings in all participants irrespective of vEAR, while the low-level stimulus-driven factor of ME was specific to those respondents who were previously aware of experiencing vEAR.

5.2 Returning to the Debates from Chapter 1

We began Chapter 1 by introducing the concept of synaesthesia, a condition in which sensory experience in one modality or cognitive domain triggers illusory percepts in another sensory modality or cognitive domain. In particular we focused on a little-known synaesthetic subtype in which individuals report being able to hear illusory sounds accompanying visual motion (Guttman, Gilroy & Blake, 2005; Saenz & Koch, 2008). Although this has been referred to as hearing-motion synaesthesia in the past (Saenz and Koch, 2008) we remain agnostic about whether this truly represent a type of synaesthesia until we have a better understanding of this phenomenon. We then outlined several contemporary debates from the field of synaesthesia research that may potentially be informed by the findings of this thesis. We now present some of these key points below.

5.2.1 Neurophysiological Diagnostic Criteria

The first of these debates surrounds whether we need to adopt a better understanding of the neurophysiology that underlies the condition into our defining characteristics of synaesthesia, which some have argued are excessively focused on behavioural markers of the condition (e.g. Simner, 2012). Such potential physiological underpinnings can broadly be divided into two categories; those that focus on the amount of physical connectivity between cortical regions (e.g. Bargary & Mitchell, 2008; Leeuwen, den Ouden, & Hagoort, 2011; Ramachandran and Hubbard, 2001; Rouw, 2013; Tomson et al., 2011) and those that focus on the balance of excitability and inhibition of the regions involved in the qualia associated with synaesthesia (e.g. Cohen Kadosh et al., 2009; Cohen Kadosh & Walsh, 2006; Grossenbacher & Lovelace, 2001; Neufeld et al., 2012).

5.2.1.1 Neural Basis of vEAR

Our results from Chapter 3 indicate that the brains of those who do versus do not experience vEAR are indeed physiologically distinct. We proposed that the TACS effect we observed in those who do not experience vEAR is indicative of a naturally-occurring mutual inhibition between the visual and auditory cortices that might be carried by alpha frequency oscillations (Jensen and Mazaheri, 2010; Klimesh, Sauseng, & Hanslmayr, 2007; Mathewson et al., 2009; Sauseng et al., 2009). Alpha is the dominant oscillatory frequency in the human brain (Linkenkaer-Hansen et al., 2004; Klimesch, 2012) and is thought to represent a mechanism of local cortical inhibition, (Jensen and Mazaheri, 2010; Klimesh, Sauseng, & Hanslmayr, 2007; Mathewson et al., 2009; Sauseng et al., 2009), with a greater alpha amplitude meaning a greater level of local inhibition. By applying alpha band TACS to the visual versus auditory cortex we appear to bias this inhibitory balance in favour of a greater disinhibition of the unstimulated region. This shifting of the balance of inhibition might allow the individual to attend to one modality and ignore the other. By applying alpha frequency TACS we appear to have disrupted the inhibitory signal from the stimulated region, leading to an improvement of performance that might be related to a disinhibition in the unstimulated region of cortex. This could explain why performance in the first modality deteriorates and in the second modality it is enhanced, if the disinhibited region is then able to further inhibit the stimulated region.

In those who do experience vEAR there was a considerably reduced effect of TACS, which suggests that any such cortical rivalry is substantially weaker in those who report being able to hear visual events. It may be that these individuals have less of the mutual audiovisual cortical inhibition described above, so the application of alpha band TACS consequently has less effect. In this case, rather than compete for resources the auditory and visual cortices of people who experience vEAR cooperate rather than inhibit one another. This means that no matter whether TACS is applied to the visual or the auditory cortex, they are able to perform the Sequence Discrimination Task in either modality with little effect of TACS. This key difference between our two groups suggests that people who experience vEAR may have less inhibition of signals from the visual to the auditory cortex. As well as explaining the TACS results described above, this lack of inhibition of the auditory cortex by the visual cortex may also explain why these individuals report hearing visual events in the first place. With less mutual inhibition it is likely that overall auditory cortical excitability is generally higher in these individuals, meaning it takes comparatively little signal from the visual areas to sufficiently excite the auditory cortex to generate faint auditory qualia. This greater level of baseline cortical activity may explain in part why we observe a greater prevalence of vEAR in individuals with tinnitus, those who frequently experience auditory imagery, and those who experience hypnagogic sound-induced phosphenes, as it would require less cross-modal input for their already excitable cortices to generate this range of unusual qualia.

5.2.2 Consistency of Inducer-Concurrent Pairings

We also discussed whether or not the consistency of inducer-concurrent pairings is necessarily a mandatory component of any working definition of synaesthesia. Asking synaesthetes to report the properties of their concurrents (often colour) evoked by a range of inducers has often been used to test the veracity of synaesthetes claims to perceive the world in the way they do (e.g. Asher et al., 2006; Baron-Cohen et al., 1996). However, Eagleman (2012) notes that in the extensive dataset collected via his online synaesthesia battery (synesthete.org; Eagleman et al., 2007), they do not find a bimodal distribution of scores but rather some highly consistent scores and a long tail representing less consistent scores (Eagleman, 2012). Simner (2012) reports that a large number of people claim to experience synaesthesia but routinely fail tests of consistency. We suggest that an insistence that inducer-concurrents must be consistent over time neglects the possibility that transitory synaesthesia-like states may exist, as well as excluding phenomena such as vEAR in which no specific concurrent can be pinpointed (rather, a generic ‘whooshing’ or ‘beeping’ sound is reported). Therefore, while one could argue that there is a degree of consistency, in that motion consistently leads to a generic auditory sensation, we are not able to map a range of specific visual stimuli to a diverse array of sound types in the way that we can with, for example, tone-colour synaesthesia.

Our results suggest that the association between visual motion and an accompanying auditory sensation is consistent, although not as specific as canonical synaesthesias such as grapheme-colour or tone-colour synaesthesia. If vEAR was only present when it was beneficial to a particular task, then we would not expect to see vEAR interrupt auditory detection, as we reported in chapter 2. This inability to turn vEAR on and off as and when it is required suggests it is indeed a consistent concurrent accompanying visual stimulation. This could be further tested by repeating our video questionnaire described in chapter 4 over time. If vEAR is indeed a consistent phenomenon then we would expect it to be consistently evoked by the same types of visual stimuli, meaning ratings for each video should remain consistent over time.

5.2.3 Is Synaesthesia Dichotomous or Continuous?

Synaesthesia is typically presented as a rare and unusual phenomenon that only affects 2-4% of the population (Simner at al., 2006; Ward, 2013). The assumption made here is that synaesthesia is a dichotomous phenomenon that a small number of people have while the rest of us do not. Alternatively, synaesthesia may manifest along a continuum, with a range of graded experiences experienced throughout the population. Proponents of the former account may suggest that synaesthesia is the result of a rare genetic variant (Asher, 2009; Brang & Ramachandran, 2011; Tomson et al., 2011), although cases have been reported in which only one of two monozygotic twins experiences synaesthesia (Smilek et al., 2002; Smilek, Dixon, & Merikle, 2005) which suggests that environmental factors may influence the development of the condition (see following section). Evidence for some form of synaesthesia spectrum can be seen in the individual differences in synesthetic percepts, such as the distinction between associator and projector synaesthetes, and in the fact that many objective tests of synaesthesia do not result in a bimodal distribution of scores (Eagleman, 2012; Rouw & Scholte, 2007) as we would predict for a truly dichotomous phenomenon. We also suggested that many studies may risk exhibiting a sampling bias by recruiting both synaesthetes and controls exclusively through self-selection, whereas it may be that some of the control participants actually experience low-level synaesthetic pairings of which they may not be aware. This would then lead to a sampling bias in which only the extreme ends of the spectrum are captured and thought to be unrelated samples rather than opposing poles.

5.2.3.1 Is vEAR Dichotomous or Continuous

We may ask whether the experience of vEAR is one that runs throughout the population in a graded manner, or whether it is a binary ability that one either has or does not have. If it is characterised by such a dichotomy, then vEAR might be the same phenomenon as the 'hearing-motion synaesthesia' outlined by Saenz and Koch, or alternatively these individuals may be at the top of an ongoing spectrum? The synaesthete participants in the original Saenz and Koch (2008) study had prior, conscious awareness of their concurrents. In our original random sample outlined in chapter 2 however, we saw a discrepancy between those who reported a prior awareness of this ability (11%) and those who reported a retrospective awareness of hearing the visual flashes when questioned in our lab (22%). It may therefore be the case that the experience of vEAR is a weaker form of nascent synaesthesia that appears to be more common in the population than standard estimates of more overt synaesthesias, which typically sit at around 2-4% (Simner at al., 2006; Ward, 2013). We may also address this question of dichotomy versus continuity by exploring whether the results of our behavioural measures display a bimodal or a continuous distribution. In chapter 2 we employed a mix of objective psychophysical tasks (e.g. Sequence Discrimination, Auditory Detection) and more subjective questionnaire-based measures. Of these, only when we asked our participants whether they were aware of hearing the sounds on the Sequence Discrimination Task did we get a dichotomous response distribution (necessarily so with a binary outcome question). However, although those responded that they could hear the flashes were significantly better at visual sequence discrimination, the distribution of our objective measures appears smooth rather than discontinuous. This continuity is unlikely to indicate a distinct group of individuals who experience visual stimulation in a uniquely bimodal way, rather it is more probably that this data reflects individual differences in normal audiovisual cortical connectivity. Furthermore, the smooth trait associations found with video ratings reported in chapter 4 suggest vEAR reflects systemic physiological variables, rather than being restricted to one particular group.

Our data from Chapter 4 appear to suggest that vEAR may indeed be spectral in nature. When asked to rate a series of silent videos for the amount of auditory sensation the evoked, there was a high level of agreement over which videos evoked the most auditory sensation, even between those who do versus do not report experiencing vEAR. This indicates that vEAR may be tapping into a normal cross-modal phenomenon, perhaps akin to cross-modal correspondences (Spence, 2011), that in some individuals crosses a threshold into a consciously perceived sound. The results of our Principle Component Analysis suggest that while the predictiveness of a sound is a common driving factor in stimulus ratings of vEAR, the additional component of motion energy is an additional factor only influencing ratings in those who report a previous awareness of vEAR. This may be a candidate for an additional factor that raises vEAR into conscious perceptibility.

5.2.4 Are Synaesthetic Pairings Environmentally Influenced?

We also discussed the extent to which the specific pairings experienced by a synaesthete are learned from their environment. Taking the example of grapheme-colour synaesthesia, Rich, Bradshaw & Mattingley (2005) reported considerable consistency between the pairings of different synaesthetes, with the letter 'Y' eliciting the colour yellow in 45% of their sample, and the letter 'D' evoking brown for 47%. Further, we can chart the developmental trajectory of these pairings, with child grapheme-colour synaesthetes having concurrent colours for approximately 35% or letters by the age of 7, rising to 70% of letters by the age of 11 (Simner et al., 2009; Simner and Bain, 2013). It is not known where in the environment these letter-colour pairings originate, but some have suggested that coloured letter fridge magnet toys may be one candidate (Witthoft, Winawer, and Eagleman, 2015). Studies of twins with coloured-sequence synaesthesia (CSS) have shown a greater concordance of inducer-colour pairings in monozygotic twins (73.9%) than in dizygotic twins (36.4%) (Bosley & Eagleman, 2015), indicating that while genetics clearly seems to play a role in determining these pairings, environmental exposure too must play a significant role.

5.2.4.1 vEAR and Environmental Exposure

Working from this premise we proposed that synaesthesia may be more common for sensory pairings that co-occur more frequently in the environment compared to more unusual and bizarre pairings, such as words and tastes. We suggested that sound and visual motion may be one such pairing that is highly correlated in our environment, with examples including lip-movements and speech, or motion-to-impact sounds (e.g. clapping hands or footsteps). These pairings have the further advantage that they do not rely on an infant to be sufficiently cognitively developed to possess concepts such as numbers and letters, and are reinforced from an earlier age prior to the critical period of development and the subsequent synaptic pruning. This may mean any synaesthetic pairings between movement and sound are more likely to persist into adulthood due to a stronger association formed via Hebbian learning. This is supported by our higher prevalence estimates of vEAR compared to canonical synaesthesia, which is consistent with our environmental exposure hypothesis.

5.2.4.2 Prevalence of vEAR

We reported two measures of prevalence of vEAR in Chapter 2, the first of these is the 22% of participants who reported being able to hear an auditory sensation accompanying the flashes on the Sequence Discrimination Task under laboratory conditions. The second, more conservative figure is the 11% of participants who stated that they had previously been aware of being able to hear visual events prior to participating. This discrepancy could be explained in a number of ways. First, although efforts were made not to lead the participants in any way, it is possible that our 22% figure is inflated by demand characteristics, although we were conservative in our interpretation of answers on the debrief questionnaire, so for example if a participant was asked whether they were aware of hearing mental sounds accompanying the flashes and answered 'maybe' or 'I think so' then this was interpreted as a no. Another explanation for the discrepancy in prevalence scores might be that some people who experience vEAR had genuinely not noticed the phenomenon until their attention was first drawn to it in a silent and introspective environment and their attention was then drawn to it by the questions of the experimenter. This is consistent with precisely the type of 'weak and strong' vEAR that we would predict in the phenomenon manifests as a continuum, as discussed above. Therefore it may be the case that those participants who had previously been aware of hearing visual events are those for whom the visually-evoked sounds are the most vivid, enough to enter conscious awareness even with background noise, while some of those who became aware during the experiment are experiencing a subtler form of vEAR. One way in which this could be tested empirically would be to invite these two groups to complete our questionnaire from Chapter 4, on which we might predict those with the ‘new’ experience of vEAR to score items as evoking less sound than would those with a previous awareness of experiencing the condition, leading to lower overall ratings for the video stimuli.

In Chapter 4 we report that 21% of our survey respondents from a main sample of 4128 (fully-completing respondents) reported that they have previously experienced sounds evoked by visual motion. This is a considerably larger proportion than the 11% who reported prior awareness of vEAR in Chapter 2, but this latter estimate is likely subject to self-selection bias. Survey respondents volunteered to participate by following a hyperlink on an online article in the popular press about visually-evoked auditory sensations. If we include the respondents who did not complete the survey in full (a much larger sample of 33,504) the frequency of ‘yes’ responses here was slightly less at 16%. Including these respondents may somewhat diminish the sampling bias, as presumably many of these declined to complete the survey because they realised they did not experience any visually-evoked sensations, with the more motivated respondents who did complete the survey being more likely to experience vEAR. It is likely therefore that the true prevalence of individuals with a prior awareness of experiencing visually-evoked sounds is somewhere between this 11-16%

It should be noted that even our more conservative estimate of prevalence (11%) is still considerably higher than the 2-4.4% prevalence reported for canonical synaesthesia (Simner at al., 2006; Ward, 2013). The increased prevalence of vEAR is consistent with our hypothesis outlined in section 1.4.4 that some synaesthetic sub-types might occur more frequently than others due to the inducing and concurrent sensations being more frequently co-occurring in nature. This was based on evidence that suggests some synaesthetic associations may result from exposure to recurrently paired sensory features (Bor et al., 2014; Witthoft et al., 2015). Consequently, we predicted that if an individual had a predisposition toward developing synaesthetic pairings they would be more likely to develop an association between pairings that are consistently reinforced in the world around them than they would for pairings that are extremely rare in nature, such as between graphemes and colours, or between words and smells, for example.

In Chapter 3 we also examined the potential differences between the general population and classically-trained musicians, having hypothesised that this population will have substantially more exposure to synchronous motion (e.g. the conductor's baton, the movement of other orchestra members) and the accompanying music. As predicted we saw significantly higher levels of vEAR in our classical musicians than in other subjects. This is in turn consistent with our assertion that a degree of learning is involved in the development of vEAR, combined with the higher prevalence of vEAR that results from frequently-occurring pairings in the natural world, compared to some canonical synaesthesias. However, this claim could be contested on the grounds that we are not able to establish causality; it may be that musicians are born with atypical audiovisual connectivity which affords them such prodigious talents in the first place.

If vEAR is as common as we suggest, then why is there so little awareness of the phenomenon? The dearth of publications on the topic, and the discrepancy in our participants between those with new and existing awareness of their own vEAR, would attest to the fact that it is not a condition about which there is a great deal of knowledge. Somewhat counter intuitively, this lack of awareness may well be precisely because of the high levels of co-occurrence of visual motion and sound in the natural world. Because visual motion often co-occurs with sound, when an individual perceives a faint auditory sensation that accompanies a moving object they may be more likely to ignore the sound as it is in no way novel or unexpected. They may even mistake the visually-evoked sound for a genuine sound made by the moving object. For example, a faint sensation that accompanies the stride of an individual walking ahead of us could easily be mistaken for the sounds of their feet hitting the pavement. This is considerable less likely to occur with unusual sensory pairings; we would not expect an individual who perceives a taste when they hear a particular word to make the same error. The common association of movement and sound may make experiences like vEAR all the less remarkable when they occur. Unlike unusual sensory pairings, those which are so frequently experienced together may be so ubiquitous as to simply be ‘part of the package’, as with taste and the scent of our food. These two senses are so closely entwined in the culinary experience that frequent correspondence in this domain (e.g. the ‘sweet’ smell of chocolate (Stevenson & Tomiczek, 2007; van Campen, 2008) scarcely register as being cross-modal even in a metaphorical sense. This propensity to discount frequently co-occurring cross-modal sensations may explain in part why in Chapter 2 we identified 22% of participants who were able to hear the visual flashes and yet only 11% report previously being aware of experiencing this phenomenon. This raises the intriguing possibility that we may be able to train individuals, not necessarily to experience vEAR, but to learn to attend to it.

5.2.5 Learning to Attend to vEAR

Over the course of writing my thesis, the question I have been asked the most is whether a person can be trained to experience synaesthesia. I can certainly attest that I personally have become aware of such sensations over the course of my research that I had not been attuned to before. In Chapter 2 I recount becoming aware of hearing a fluttering sound evoked by the shadows of roadside trees as I drove along one dusky evening. Another notable inducer that I have become aware of is the flashing of cyclists' lights at night. It is possible that I have become aware of a latent ability that I had not previously been attuned to, or that my countless hours in the lab watching participants perform the Sequence Discrimination Task outlined in Chapters 2 and 3 have become bound through some combination of Hebbian learning and top-down expectation. What is striking is that both of these examples of my own awareness of vEAR occurred in low-light situations, potentially with the effect of boosting the signal-to-noise ratio of my own latent vEAR. My own experience raises a question that is frequently raised in the field of synaesthesia research; can we train people to possess this ability? There have been numerous attempts to achieve this empirically, with limited success (e.g. Bor et al., 2014; Colizoli, Murre, & Rouw, 2012; Meier & Rothern, 2009; Rothern & Meier, 2014).

How then are we able to explain my own limited experience of vEAR to date? Although I had heard of synaesthesia prior to becoming actively engaged in this research, I had not heard of hearing-motion synaesthesia, and we had not yet coined the term vEAR. Nor had I ever been aware, that I can recall, of hearing any visually-evoked auditory sensations whether I had a term to describe them or not. Have I therefore 'trained' myself to hear auditory sensations when viewing visual flashes, perhaps as the result of many hours sat in a dark lab watching participants complete the

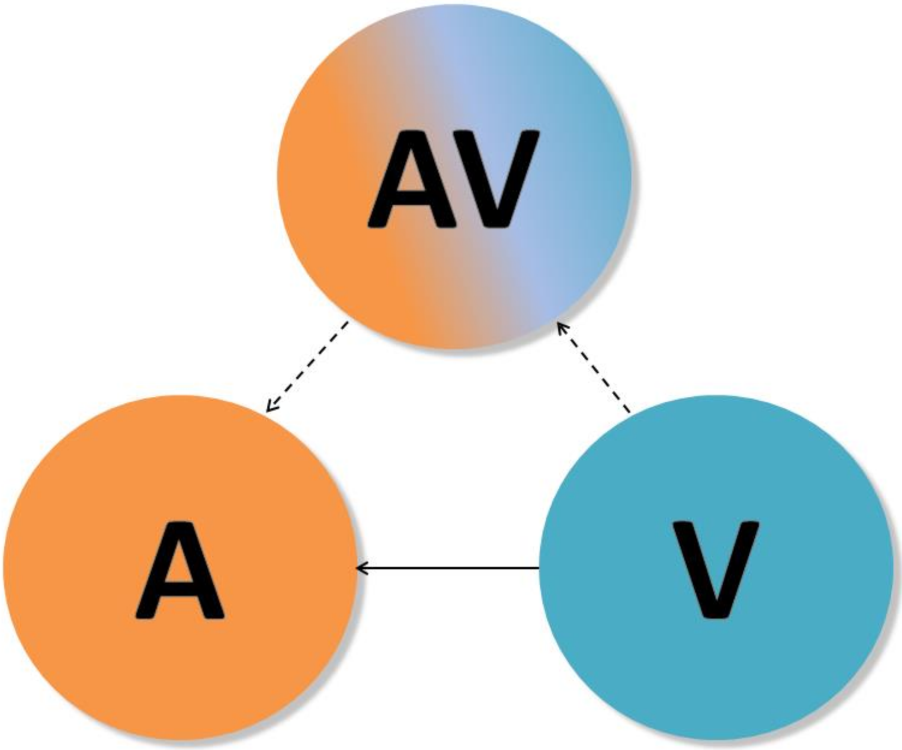
Sequence Discrimination TASK? This seems unlikely. If we reflect on the two different prevalence estimates that we reported in Chapter 2, these could be divided into those who reported a previous awareness of experiencing vEAR (11%) and those who reported a retrospective awareness of hearing the visual flashes under laboratory conditions (22%). We outlined above how, if vEAR manifests along a spectrum, there may be a portion of the population who experience faint auditory sensations that have not entered into conscious awareness, either because they tend to be masked by real sounds or because they have learned not to attend to these sensations. Therefore we suggest that it may not be possible to train an individual to experience vEAR, but instead to become aware of their nascent abilities and to introspectively focus upon them. It may therefore be possible under these circumstances even to improve visual sequence discrimination ability under an inwardly attentive, almost meditative-like state. This could involve repeating the Sequence Discrimination Task after the participant had become attuned to their own vEAR, and monitoring any subsequent improvement in visual sequence discrimination ability. To control for practice effects this could then be compared with subsequent experimental sessions with participants who did not report hearing the visual sequences.

5.3 A Model to Explain vEAR

We are now able to collate these findings into a model which we think conveys what may be the underlying mechanisms that give rise to vEAR. In Chapter 1 we outlined two potential mechanisms that may explain vEAR, that of increased connectivity between the brain regions processing the inducer stimulus modality and the concurrent sensation qualia, or alternatively a greater level of disinhibition between otherwise normal connectivity between these regions. In Chapter 3 we proposed that the mutual-inhibition between the visual and auditory cortices that occurs in the general population (Mattingley et al., 1997) and may be led by alpha oscillations (Klimesch, Sauseng, & Hanslmayr, 2007), is reduced in the participants who experience vEAR. This was based on our observation that alpha frequency TACS modulates task performance only in those who do not experience vEAR, presumably because if alpha disrupts this pattern of mutual inhibition then its effect will be diminished in individuals with little inhibition to disrupt.

We are not the first to identify disinhibition as a candidate mechanism to explain synaesthesia (or in the case of vEAR, a synaesthesia-like phenomenon). Neufeld (2012) compared the brains of auditory-visual synaesthetes to controls using fMRI to perform a functional connectivity analysis to determine how different cortical regions interact in synaesthetes versus controls during the perception of auditory stimuli. While they found no differences in the anatomical connections between the auditory and visual cortices, synaesthetes displayed greater levels of functional connectivity of the left inferior parietal cortex with both the left primary auditory cortex and right primary visual cortex. This suggests that in these synaesthetes their brains may not necessarily differ structurally from controls in how their visual and auditory cortices are physically connected, instead the synaesthetes are receiving feedback from higher level multimodal areas to early visual and auditory cortices that are usually inhibited in the typical population.

Although our focus in Chapter 3 was on early sensory cortices, in Chapter 4 we explored the relative contribution to vEAR of low-level, bottom-up stimulus features, as well as top-down higher processes based on expectation and prediction. We presented respondents to a large internet survey with a series of short silent video clips depicting a range of different types of visual motion, and explored the factors that influenced ratings. The two principle components (PCs) that identified in Chapter 4 suggest that visual scenes are rated more highly for the amount of vEAR evoked if they depict an event that is predictive of an accompanying sound (this PC accounted for 61% of the observed scores), or to a lesser degree whether they contain high levels of motion energy (this PC accounted for 6.9% of the data). We suggest that the former represents a higher-level executive process built on learning and expectation, as a result of the high levels of co-occurrence of visual motion and sound in the natural world. This effect should be present throughout the population, irrespective of whether one experiences vEAR or not, because we all inhabit a world in which visual motion and accompanying sounds co-occur extremely frequently. The latter component that drives ratings of vEAR in visual scenes is likely to be a lower-level process that is preattentive and carried by direct connections between early visual and auditory cortical areas. Our findings suggest that this component was only predictive of scores in individuals who do experience vEAR, and thus is likely to be a feature that distinguishes those who do experience the phenomenon from those who do not.



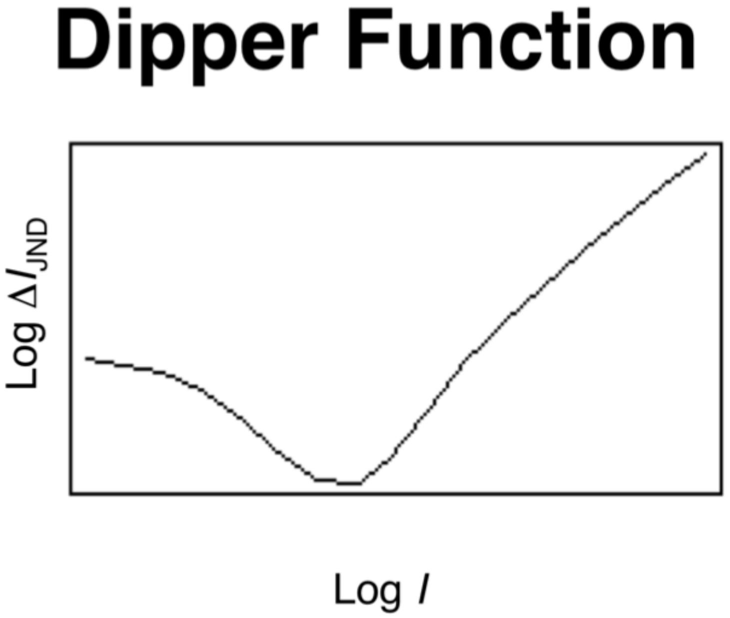
5-1: A model of vEAR. In this model dynamic visual stimuli are processed both via the normal higher-level route. Dashed line represents common sensory cross-talk that is largely ubiquitous, while solid line may be unique to those who experience vEAR.

5.4 Is vEAR a type of Synaesthesia?

In chapter 2 we presented evidence to support the assertion that these visually-evoked sounds are perceptually 'real' to the perceiver, following Lovelace, Stein & Wallace (2003) who reported that an irrelevant visual flash can aid detection of a faint auditory target. We first measured individual auditory thresholds for each subject, before running a single interval forced choice task in which participants had to report whether or not they detected the auditory target, which was present on 50% of trials. The target was either presented alone or with an accompanying dynamic visual stimulus that was designed to evoke vEAR. Results showed that for some participants the presence of the visual stimulus improved auditory detection (as in the original Lovelace, Stein & Wallace report, 2003), in others auditory detection was impaired by the presence of the visual stimulus. Crucially there was a negative correlation between the scores on this measure and on the visual Sequence Discrimination Task, meaning those who scored high on our measure of vEAR showed worse performance on the Auditory Detection Task when the auditory target was present with an irrelevant visual stimulus. This is consistent with our conceptualisation of vEAR as being a true auditory phenomenon that is 'heard' rather than willed or imagined, a claim which is supported by the lack of correlation with our measure of visual bias following Colavita (1974). We can therefore claim that vEAR meets the criteria of being a sensory experience that is evoked by stimulation in another sensory modality. We discuss above the distinction between consistency and specificity. Our results from chapter 2 indicate that the auditory sensation evoked in vEAR cannot be ignored when it is inconvenient, as seen in the poorer auditory detection in the presence of a visual stimulus observed in our participants who experience vEAR. We can therefore suggest that vEAR meets the criteria of consistency, although there is no specific mapping between particular sets of stimuli and distinct concurrent percepts. Perhaps the area in which vEAR cannot be considered a type of synaesthesia as we currently understand the condition is in how widespread the phenomenon appears to be. Our data throughout this thesis consistently suggest that vEAR is at least twice as prevalent as other types of synaesthesia, with a conservative estimate of 11% of the population experiencing some form of visually-evoked auditory sensation. We therefore refer back to the discussion above in which we debate whether vEAR is continuous or dichotomous in the population. If this is indeed a normal phenomenon then it by definition cannot be a synaesthesia, which is usually characterised as being an anomalous sensory experience. We might then suggest that those at the top end of the scale are the 'true' synaesthetes, as they live with an overt vEAR sensation that is far more ubiquitous than the tentative claims of some of our random sample from chapter 2. Ultimately however we are left with the ambiguity of knowing at what point in the continuum these sensations pass the threshold into synaesthesia, reducing the debate to an academic exercise in semantics.

5.5 Directions for Future Research

In Chapter 2 we present evidence to suggest that the visually-evoked auditory sensations that characterise vEAR are perceptually real enough to modulate detection of real-world sounds. Following Lovelace, Stein and Wallace (2003) we presented auditory targets at threshold in the presence versus absence of an irrelevant dynamic visual stimulus. Some participants auditory detection was improved by the presence of the visual stimulus, as in the original report by Lovelace and colleagues, but our results demonstrated that for other participants the presence of the visual stimulus was detrimental to auditory target detection. Crucially, scores on this task were negatively correlated with performance on the visual element of the Sequence discrimination Task (Saenz & Koch, 2008), our measure of vEAR traits. Our interpretation of this finding is that in those participants who experience vEAR, the presence of the visual stimulus on the Auditory Detection Task introduces extra noise to the auditory system, reducing the signal-to-noise ratio and thus impairing target detection. One method in which future research could verify this interpretation is by analysing whether the visually-evoked auditory sensation is able to act as a pedestal and a psychometric ‘dipper’ function observed. This is a phenomenon in which the introduction of a small amount of additional noise in a low-contrast discrimination task has the somewhat counterintuitive effect of first actually reducing the just-noticeable different (JND) in signal increments before they increase again (for a comprehensive review of the Dipper Function, see Solomon, 2009). This technique has been used to demonstrate the nonlinearity of both the visual system (e.g. Morgan, Chubb, & Solomon, 2008) as well as the auditory system (Raab, Osman, & Rich, 1963). If, as we suggest, vEAR is a true faint auditory signal then one should be able to observe the dipper function on an auditory detection task in which the pedestal is a dynamic visual stimulus, and the presence of the dip in the psychometric curve should only be seen in participants who either report hearing flashes, or who score highly on other subjective measures of vEAR.



5-2: An example psychometric function depicting responses on an orientation discrimination task. A very small anticlockwise tilt in the stimulus leads to the characteristic ‘dip’.

Another avenue for future research could test our assertion in Chapter 3 that the normal inhibitory mechanism between the visual and auditory cortices, that we suggest is diminished in vEAR, is carried by alpha frequency neural oscillations. We present evidence in Chapter 3 that we believe supports this hypothesis, namely the absence of any effect of alpha TACS in our participants who report being able to hear flashes. Conversely, in those who do not experience vEAR we report that alpha frequency TACS impaired sequence discrimination ability in the modality relevant to the stimulation site (i.e. occipital stimulation impairs visual sequence discrimination, temporal stimulation impairs auditory sequence discrimination) but aids sequence discrimination in the converse modality. If alpha-mediated inhibition is present in those who do not experience vEAR, then we suggest that alpha TACS would bias this balance of inhibition by dampening the inhibitory signal from the stimulation site, leading to less inhibition in the unstimulated cortical site. This hypothesis could be tested using electroencephalography (EEG) by comparing alpha power in those who do versus do not experience vEAR while performing the Sequence Discrimination Task outlined in Chapter 2. EEG could thus be used to test whether alpha power is greater in those who do not experience vEAR compared to those who do, or potentially to demonstrate whether alpha oscillations in the visual and auditory cortices are asynchronous in participants who experience vEAR. An alternative test of our disinhibition hypothesis would be to repeat our TACS experiment from Chapter 3 using gamma rather than alpha frequency TACS. Our lab has now begun conducting this experiment and we expect to have the results shortly. Early findings indicate that both temporal and occipital gamma TACS impair performance on the visual Sequence Discrimination Task only in those who show default high visual performance relative to auditory sequence discrimination ability, but not in those with default low performance.

If vEAR is indeed the result of reduced inhibition then we would expect this to be reflected in a variety of other effects, such as further traits and abilities including extroversion, response times, anxiety (Wasserman et al., 2001). It is also possible to raise auditory cortical excitability using TMS and explore how this effects the ability to experience vEAR. In addition, drugs that modulate cortical excitability could affect performance on our behavioural tasks. Existing pharmacological evidence suggests that synaesthesia and visual phosphene thresholds can be modulated by taking drugs that selectively modulate the action of serotonin (Brang & Ramchandran, 2008; Brogaard, 2013; Luke & Terhune, 2013; Oliveri, 2003). This neurotransmitter may have complex effects on cortical excitability by acting on glutamate- and GABA-mediated transmission (Ciranna, 2006). It is therefore possible that vEAR and associated sensory phenomena may also be associated with generally raised serotonin levels, which could also be assessed experimentally.

5.6 Conclusion

To conclude, this thesis represents the first extensive exploration of the vEAR phenomenon, and other than the Saenz and Koch (2008) report it is to our knowledge the only examination of a visual-motion to sound synaesthesia. We have contributed to several continuing debates that surround synaesthesia as well as generating testable hypotheses for future research into vEAR, how it manifests, who experiences it and what evokes it. We have described a multisensory phenomenon in which certain individuals hear auditory sensations when viewing dynamic visual stimuli. We call this the Visually-Evoked Auditory Response (vEAR), and argue that this shares some of the canonical features of synaesthesia, most specifically the hearing-motion synaesthesia described by Saenz and Koch (2008). Where it diverges from the standard defining features of synaesthesia, such as in the lack of a specific inducer-to-concurrent mapping (instead being characterised by a generic ‘white-noise’ sensation), it may help to inform a reconceptualization of synaesthesia that some authors have proposed (e.g. Eagleman, 2012; Simner, 2012). We have presented data to demonstrate that vEAR is sufficiently tangible to impair detection of real-world sounds, and that it cannot be explained by an attentional bias to vision over audition. Using neurostimulation to disrupt task performance in those who do versus do not experience vEAR we have argued that the phenomenon appears to be the result of a disinhibition of connections between the auditory and visual cortices, meaning that individual with vEAR is able to use either cortical region to perform the task. Finally, we examined the types of visual stimulus that most evoke vEAR. Using a series of silent video clips that had been rated according to how much of an auditory sensation they evoke we report that in those who experience vEAR the highest rated visual stimuli are characterised by greater levels of motion energy (i.e. the amount of movement present in the scene), rather than those that predict the expectation of sound from learned experience. We also show that other factors that predict an individual’s propensity toward vEAR include particular traits that may also be associated with an unusually excitable or disinhibited auditory cortex, including experience of tinnitus, musical imagery (earworms) and experiencing hypnagogic sound-induced phosphenes. This appears to be consistent with vEAR reflecting a systemic excitability and/or disinhibition of auditory cortical and other neighbouring regions, and their inputs from visual cortical regions.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America,* 2(2), 284–99.

Afra P, Anderson, J., Funke, M., Johnson, M., Matsuo, F., Constantino, T., & Warner, J. (2012) Neurophysiological investigation of idiopathic acquired auditory–visual synesthesia. *Neurocase 18*(4):323–329.

Ahmadi, J., Keshtkar, M. & Pridmore, S. (2011). Methamphetamine induced synesthesia: a case report. *American Journal of Addiction, 20*, 306.

Anderson, H. P., Seth, A. K., Dienes, Z., & Ward, J. (2014). Can grapheme-color synesthesia be induced by hypnosis? *Frontiers in Human Neuroscience, 8*, 220. Published online 2014 Apr 28. doi: 10.3389/fnhum.2014.00220. PMCID: PMC4009433.

Asher, J. E., Lamb, J. A., Brocklebank, D., Cazier, J., Maestrini, E., Addis, L., Sen, M., Baron-Cohen, S., & Monaco, A. P. (2009). A Whole-Genome Scan and Fine-Mapping Linkage Study of Auditory-Visual Synesthesia Reveals Evidence of Linkage to Chromosomes 2q24, 5q33, 6p12, and 12p12. *American Journal of Human Genetics, 84*(2), 279-285).

Auvray, M., and Deroy, O. (in press). “How do synesthetes experience the world,” in *Oxford Handbook of Philosophy of Perception*, ed M. Matthen (New York, NY: Oxford University Press).

Banissy, M. J., & Ward, J. (2013). Mechanisms of self-other representations and vicarious experiences of touch in mirror-touch synesthesia *Frontiers in Human Neuroscience, 7*: 112. Published online 2013 Apr 3. doi: 10.3389/fnhum.2013.00112

Bargary, G., & Mitchell, K. J. (2008). Synaesthesia and cortical connectivity. *Trends in Neuroscience, 31*(7):335-42. doi: 10.1016/j.tins.2008.03.007. Epub 2008 Jun 10.

Baron-Cohen, S. (1996). Is There a Normal Phase of Synaesthesia in Development? *PSYCHE, 2*(27); available http://psyche.cs.monash.edu.au/v2/psyche-2-27-baron cohen.html

Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J., & Bolton, P. (1996) Synaesthesia: prevalence and familiality. *Perception, 25*, 1073–1079.

Baron-Cohen, S., Wyke, M. A., & Binnie, C. (1987). Hearing words and seeing colours: an experimental investigation of a case of synaesthesia. *Perception, 16*(6):761-7.

Bedny, M., Richardson, H., & Saxe, R. (2015) "Visual" Cortex Responds to Spoken Language in Blind Children. Journal of Neuroscience, 35(33), 11674-11681. DOI: https://doi.org/10.1523/JNEUROSCI.0634-15.2015

Bezzola, Ladina, Susan Mérillat, Christian Gaser, & Lutz Jäncke. (2011). Training-induced neural plasticity in golf novices. *The Journal of Neuroscience 31* (35): 12444–12448.

Bien, N., ten Oever, S., Goebel ,R., & Sack, A. T. (2012). The sound of size: cross-modal binding in pitch-size synesthesia: a combined TMS, EEG and psychophysics study. *Neuroimage, 59*(1), 663-72. doi: 10.1016/j.neuroimage.2011.06.095.

Block, L. (1983). Comparative Tone-Colour Responses of College Music Majors with Absolute Pitch and Good Relative Pitch. *Psychology of Music, 11*, 59-67.

Bolognini, N., Miniussi, C., Gallo, S., & Vallar, G. (2013). Induction of mirror-touch synaesthesia by increasing somatosensory cortical excitability. *Current Biology, 23(*10), R436-R437.

Bor, D., Rothen, N., Schwartzman, D. J., Clayton, S., & Seth, A. K. (2014). Adults Can Be Trained to Acquire Synesthetic Experiences. *Nature: Scientific Reports, 4*, retrieved from https://www.nature.com/articles/srep07089

Boroojerdi, B., Bushara, K. O., Corwell, B., Immisch, I., Battaglia, F., Muellbacher, W., & Cohen, L. G. (2000) Enhanced excitability of the human visual cortex induced by short-term light deprivation. *Cereb Cortex 10*(5):529–534.

Bosley, H. G., & Eagleman, D. M. (2015). Synesthesia in twins: Incomplete concordance in monozygotes suggests extragenic factors. *Behavioural Brain Research, 286*(1), 93-96.

Brang, D., & Ramachandran, V. S. (2011). Survival of the Synesthesia Gene: Why Do People Hear Colors and Taste Words? *PLoS Biology 9*(11): e1001205. doi: 10.1371/journal.pbio.1001205.

Brogaard, B. (2013). Serotonergic hyperactivity as a potential factor in developmental, acquired and drug-induced synesthesia. *Frontiers in Human Neuroscience, 7*(657). doi: 10.3389/fnhum.2013.00657. eCollection 2013.

Bueti, D., and Macaluso, E. (2010). Auditory temporal expectations modulate activity in visual cortex. *Neuroimage, 51*, 1168–1183.

Calkins M., W. (1895). Synaesthesia. *The American Journal of Psychology. 7*, 90–107. doi: 10.2307/1412040.

Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience, 22*, 2886–2902.

Cecere, R., Rees, G., & Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception. *Current Biology, 25*(2):231-5.

Challinor, K. L., & Mather, G. (2010). A motion-energy model predicts the direction discrimination and MAE duration of two-stroke apparent motion at high and low retinal illuminance. *Vision Research, 50*(12), 1109-1116.

Chiou, R., Stelter, M., & Rich, A. N. (2012). Beyond colour perception: Auditory–visual synaesthesia induces experiences of geometric objects in specific locations. *Cortex, 49*(6), 1750-1763.

Ciranna L (2006) Serotonin as a Modulator of Glutamate- and GABA-Mediated Neurotransmission: Implications in Physiological Functions and in Pathology. *Current Neuropharmacology, 4*(2), 101–114.

Clemo, H. R., Sharma, G. K.,Allman, B. L., & Meredith, M .A. (2008). Auditory projections to extrastriate visual cortex: connectional basis for multisensory processing in ‘unimodal’ visual neurons. *Experimental Brain Research, 191*(1), 37–47.

Cohen, Laurent, Stanislas Dehaene, Lionel Naccache, Stephan Lehericy, Ghislaine Dehaene-Lambertz, Marie-Anne Henaff, & Francois Michel. (2000). The visual word form area—Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123*, 291–307.

Cohen J. (2017). Synesthetic perception as continuous with ordinary perception, or: we're all synesthetes now, in Sensory Blendings: On Synaesthesia and Related Phenomena, ed Deroy O., editor. (New York, NY: Oxford University Press), 1–28.

Cohen, L. G., Celnik, P., Pascual-Leone, A., Corwell, B., Falz, L., Dambrosia, J., Honda, M., Sadato, N., Gerloff, C., Catalá, M. D., & Hallett, M. (1997). Functional relevance of cross-modal plasticity in blind humans. *Nature, 389*(6647), 180-3.

Cohen Kadosh R., Henik A., Catena A., Walsh V., Fuentes L. J. (2009). Induced cross-modal synaesthetic experience without abnormal neural connections. *Psychological Science, 20*, 258–265 10.1111/j.1467-9280.2009.02286.x

Cohen Kadosh, R., & Henik, A. (2006). When a line is a number: color yields magnitude information in a digit-color synesthete. *Neuroscience, 137*(1), 3-5. doi: 10.1016/j.neuroscience.2005.08.057.

Cohen Kadosh, R., Henik, A., & Walsh, V. (2009). Synaesthesia: learned or lost? *Developmental Science, 12*(3), 484-91. doi: 10.1111/j.1467-7687.2008.00798.x.

Cohen Kadosh, R., & Terhune, D. B. (2012). Redefining synaesthesia? *British Journal of Psychology, 103*(1), 20–23.

Cohen Kadosh, R., & Walsh, V. (2006). Cognitive neuroscience: rewired or crosswired brains? *Current Biology, 16*(22), R962–3. doi:10.1016/j.cub.2006.10.017.

Cohen Kadosh, R., Cohen Kadosh, K., & Henik, A. (2007). The neuronal correlate of bidirectional synesthesia: a combined event-related potential and functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience, 19*(12), 2050-2059. doi: 10.1162/jocn.2007.19.12.2050.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics,16*(2), 409-412.

Colizoli, O., Murre, J. M. J., & Rouw, R. (2012). Pseudo-Synesthesia through Reading Books with Colored Letters. *PLOS One,* https://doi.org/10.1371/journal.pone.0039799.

Craika, F., & Bialystokb, E. (2006, March). Cognition through the lifespan: mechanisms of change. *Trends in Cognitive Sciences, 10*(3), 131-138.

Cytowic, R. E. (2002), *Synesthesia: A Union of the Senses*, 2nd ed, Cambridge, MA: MIT Press, ISBN 0-262-03296-1.

Cytowic R. E., & Eagleman D. M. (2009). Wednesday is Indigo Blue: Discovering the Brain of Synesthesia. Cambridge: MIT Press – in Simner, J. (2012). Defining synaesthesia. *British Journal of Psychology, 103*, 1-15. doi: 10.1348/000712610x528305.

Cytowic, R. E., & Wood, F. B. (1982). Synesthesia. I. A review of major theories and their brain basis. *Brain and Cognition, 1*(1), 23-35.

Day, S. (2005). *Some demographic and socio-cultural aspects of synesthesia*. In: L. C. Robertson & N. Sagiv (Eds.), Synesthesia. Perspectives from cognitive neuroscience (pp. 11–33). New York: Oxford University Press.

Day, S. A. (2013). *Synesthesia: A first-person perspective*. In The Oxford Handbook of Synesthesia, chapter 44, eds Simner, J., & Hubbard, E. M., editors. (Oxford: University Press), 903–923.

De Meo, R., Murray, M. M., Clarke, S., & Matusz, P. J. (2015). Top-down control and early multisensory processes: chicken vs. egg. *Frontiers in Integrative Neuroscience*, *9*(17). http://doi.org/10.3389/fnint.2015.00017.

Deroy, O., & Spence, C. (2013a). Are we all born synaesthetic? Examining the neonatal synaesthesia hypothesis. *Neuroscience and Biobehavioral Reviews*, *37*(7), 1240–53.doi:10.1016/j.neubiorev.2013.04.001.

Deroy, O., & Spence, C. (2013b). Why we are not all synesthetes (not even weakly so). *Psychonomic Bulletin & Review*, *20*(4), 643–64. doi:10.3758/s13423-013-0387-2.

Dixon, M. J., Smilek, D., Duffy, P. L., Zanna, M. P., & Merikle, P. M. (2006). The role of meaning in grapheme-colour synaesthesia. *Cortex, 42*(2), 243-52.

Dixon M, J., Smilek, D., & Merikle, P. M. (2004). Not all synaesthetes are created equal: projector versus associator synaesthetes. *Cognitive, Affective & Behavioural Neuroscience, 4*(3), 335-343.

Drachman, D. (2005). Do we have brain to spare? *Neurology, 64*(12), 2004-2005. doi: http://dx.doi.org/10.1212/01.WNL.0000166914.38327.BB

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature,381*, 66–68.

Eagleman, D. M. (2009). The objectification of overlearned sequences: a new view of spatial sequence synesthesia. *Cortex, 45*(10), 1266-1277. doi: 10.1016/j.cortex.2009.06.012.

Eagleman, D. M. (2012). Synaesthesia in its protean guises. *British Journal of Psychology, 103*(1), 16-19. doi: 10.1111/j.2044-8295.2011.02020.x

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., & Sarma, A. K. (2007). A standardized test battery for the study of synaesthesia. *Journal of Neuroscience Methods, 159*(1), 139-145. ISSN 0165-0270, https://doi.org/10.1016/j.jneumeth.2006.07.012.

Emerson, R. C., Bergen, J. R., & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research, 32*(2), 203-218.

Falchier, A., Schroeder, C. E., Hackett, T. A., Lakatos, P., Nascimento-Silva, S., Ulbert,I., Karmos, G., & Smiley, J. F. (2009). Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. *Cerebral Cortex, 20*, 1529–1538.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1-47.

Fernay, L., Reby, D., & Ward, J. (2012). Visualized voices: a case study of audio-visual synaesthesia. *Neurocase, 18*(1), 50-6.

Foss-Feig, J. H.,Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar,H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research, 203*(2), 381–389.

Foxe J. J., Simpson G. V., Ahlfors S. P. (1998). Parieto-occipital approximately 10 Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport 9*, 3929–3933.

Foxe, J. J., & Snyder, A. C. (2011). The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. *Frontiers in Psychology, 2*, 154. http://doi.org/10.3389/fpsyg.2011.00154.

Frey, J. N., Mainy, N., Lachaux, J-P., Müller, N., Bertrand, O., & Weisz, N. (2014). Selective Modulation of Auditory Cortical Alpha Activity in an Audiovisual Spatial Attention Task. *The Journal of Neuroscience, 34*(19), 6634 – 6639.

Fröhlich, F., & McCormick, D. A. (2010). Endogenous electric fields may guide neocortical network activity. *Neuron, 67*(1):129-43. doi: 10.1016/j.neuron.2010.06.005.

Fulkerson, M. (2014). Rethinking the senses and their interactions: the case for sensory pluralism. *Frontiers in Psychology*, *5*, 1–14. doi:10.3389/fpsyg.2014.01426.

Galton, F. (1880). Visualized Numerals. *Nature, 21*(543), 494–5.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience, 23*, 9240–5.

Ghazanfar, A. A, & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–85. doi:10.1016/j.tics.2006.04.008.

Glenberg, A. M., Mann, S., Altman, L., Forman T., & Procise, S. (1989). Modality effects in the coding reproduction of rhythm. *Memory & Cognition, 17*(4), 373-383.

Goller, A. I., Otten, L. J., & Ward, J. (2009). Seeing sounds and hearing colors: an event-related potential study of auditory-visual synesthesia. *Journal of Cognitive Neuroscience, 21*(10):1869-81. doi: 10.1162/jocn.2009.21134.

Green, D. M., Swets, J. A. (1966) Signal Detection Theory and Psychophysics. New York: Wiley. (ISBN 0-471-32420-5).

Grossenbacher, P. G. (1997). *Perception and sensory information in synesthetic experience*. In Synesthesia: Classic and Contemporary Readings (Baron-Cohen, S. and Harrison, J., eds), pp. 148–172, Blackwell.

Grossenbacher, P. G., & Lovelace, C. T. (2001). Mechanisms of synesthesia: cognitive and physiological constraints. *Trends in Cognitive Sciences, 5*(1), 36-41.

Guttman, S. E., Gilroy, L. A., & Blake, R. (2006). Hearing What the Eyes See: Auditory Encoding of Visual Temporal Sequences. *Psychological Science, 16*(3), 228-235.

Hänggi, J., Beeli, G., Oechslin, M. S., & Jäncke, L. (2008). The multiple synaesthete E.S. — Neuroanatomical basis of interval-taste and tone-colour synaesthesia. *NeuroImage, 43*, 192-203.

Haß, K., Sinke, C., Reese, T., Roy, M., Wiswede, D., Dillo, W., Oranje, B., & Szycik, GR. (2017). Enlarged temporal integration window in schizophrenia indicated by the double-flash illusion. *Cognitive Neuropsychiatry, 22*(2), 145-158. doi: 10.1080/13546805.2017.1287693.

Heeger, D. J. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology, 70*(5), 1885-1898.

Helfrich, R. F., Schneider, T. R., Rach, S., Trautmann-Lengsfeld, S. A., Engel, A. K., & Herrmann C. S. (2014). Entrainment of brain oscillations by transcranial alternating current stimulation. *Current Biology, 24*(3):333-9. doi: 10.1016/j.cub.2013.12.041. Epub 2014 Jan 23.

Henschke, J. U.,Noesselt, T. , Scheich, H., & Budinger, E. (2015). Possible anatomical pathways for short-latency multisensory integration processes in primary sensory cortices. *Brain Structure and Function, 220*(4), 955–977.

Herrmann, C. S., Rach, S., Neuling, T., & Strüber, D. (2013). Transcranial alternating current stimulation: a review of the underlying mechanisms and modulation of cognitive processes. *Frontiers in Human Neuroscience, 7*(279). Published online 2013 Jun 14. doi: 10.3389/fnhum.2013.00279.

Hubbard, E. M. (2007). Neurophysiology of Synesthesia. *Current Psychiatry Reports, 9*, 193–199.

Hubbard, E. M., Arman, A. C., Ramachandran, V. S., & Boynton, G. M. (2005). Individual differences among grapheme-color synesthetes: brain-behavior correlations. *Neuron, 45*(6), 975-85.

Hubbard, E. M., Brang, D., & Ramachandran, V. S. The cross-activation theory at 10. *Journal of Neuropsychology, 5*(2):152-77. doi: 10.1111/j.1748-6653.2011.02014.x.

Iurilli, G. Ghezzi, D, Olcese, U, Lassi, G, Nazzaro, C, Tonini, R, Tucci, V, Benfenati, F, & Medini, P. (2012) Sound-driven synaptic inhibition in primary visual cortex. *Neuron*, *73*(4), 814–828. doi: 10.1016/j.neuron.2011.12.026.

Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., Tucci, V., Benfenati, F., & Medini, P. (2012). Sound-Driven Synaptic Inhibition in Primary Visual Cortex. *Neuron. 73*(4-2):, 814–828. doi: 10.1016/j.neuron.2011.12.026.

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

Jäncke, Lutz, Gian Beeli, Cornelia Eulig, & Jürgen Hänggi. (2009). The neuroanatomy of grapheme-color synesthesia. *European Journal of Neuroscience,* 29 (6), 1287–1293.

Jensen O., & Mazaheri A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Frontiers in Human Neuroscience, 4*, 186.

Kaltenbach, J. A. (2011) Tinnitus: Models and mechanisms. *Hearing Research,* 276(1–2), 52–60.

Kayser, C., & Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Structure and Function, 212*(2), 121-32.

Kerlin, J. R., & Shapiro, K. L. (2015). Multisensory Integration: How Sound Alters Sight. *Current Biology, 25*(2), R76-R77, ISSN 0960-9822, https://doi.org/10.1016/j.cub.2014.12.016.

King, A. J, & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research, 60*(3), 492–500. PMID 4076371. doi:10.1007/bf00236934.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences, 16*(12), 606-617.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Research Reviews, 53*(1), 63-88.

Knorre, K., Ernst, M. O., & Parise, C. V. (2013). On pitch-elevation mapping. Nature, nurture and behaviour. *Multisensory Research*, *26*, 190–190. doi:10.1163/22134808-000S0143.

Koppen, C., & Spence, C. (2007). Seeing the light: exploring the Colavita visual dominance effect. *Experimental Brain Research 180*,737–754.

Kumar, S., Sedley, W., Barnes, G. R., Teki, S., Friston, K. J., & Griffiths, T. D. (2013). A brain basis for musical hallucinations. *Cortex, 52*(100), 86-97. doi: 10.1016/j.cortex.2013.12.002.

Kupers, R., Fumal, A., de Nooedhout, A. M., Gjedde, A., Schoenen, J., & Ptito, M. (2006). Transcranial magnetic stimulation of the visual cortex induces somatotopically organized qualia in blind subjects. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 13256–13260.

Laramee, M. E., Rockland,K. S., Prince, S., Bronchti, G., & Boire, D. (2013). Principal component and cluster analysis of layer V pyramidal cells in visual and non-visual cortical areas projecting to the primary visual cortex of the mouse. *Cerebral Cortex, 23*(3), 714–728.

Larntz, K. (1978). Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *Journal of the American Statistical Association. 73*(362), 253–263. JSTOR 2286650. doi:10.2307/2286650.

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

Liang, M., Mouraux, A., Hu, L., & Iannetti, G. D. D. (2013). Primary sensory cortices contain distinguishable spatial patterns of activity for each sense. *Nature Communications*, *4*(May 2013), 1979. doi:10.1038/ncomms2979.

Likova, L. (2012). The spatiotopic 'visual' cortex of the blind. *Proceedings of the SPIE, 8291*, id. 82910L (2012). (SPIE Homepage) DOI: 10.1117/12.912257.

Linkenkaer-Hansen, K., Nikulin, V. V., Palva, S., Ilmoniemi, R. J., & Palva, J. M. (2004). Prestimulus oscillations enhance psychophysical performance in humans. *Journal of Neuroscience, 24*(45), 10186-90.

Loui, Psyche, Hui C. Charles Li, Anja Hohmann, & Gottfried Schlaug. 2011. Enhanced cortical connectivity in absolute pitch musicians: A model for local hyper-connectivity. *Journal of Cognitive Neuroscience 23* (4): 1015–1026.

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research*, *17*(2), 447–453.

Luke, D. P., & Terhune, D. B. (2013). The induction of synaesthesia with chemical agents: a systematic review. *Frontiers in Psychology, 4*, 753. Published online 2013 Oct 17. doi: 10.3389/fpsyg.2013.00753.

Mackey, Allyson P., Kirsty J. Whitaker, & Silvia A. Bunge. 2012. Experience-dependent plasticity in white matter microstructure: reasoning training alters structural connectivity. *Frontiers in Neuroanatomy* 6:32.

Mankin, J. L., & Simner, J. (2017). A Is for Apple: the Role of Letter–Word Associations in the Development of Grapheme–Colour Synaesthesia. *Multisensory Research, 30*(3-5), 409 – 446. DOI: 10.1163/22134808-00002554.

Marks, L. E. (1975). On colored-hearing synesthesia: cross-modal translations of sensory dimensions. *Psychological Bulletin, 82*(3), 303-31.

Martino, G., & Marks, L. E. (2000). Cross-modal interaction between vision and touch: the role of synesthetic correspondence. *Perception, 29*(6), 745- 754.

Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science, 10*, 61–65.

Mather, G. (2013) Matlab implementation of the Adelson-Bergen motion energy sensor. Available at: http://www.georgemather.com/Model.html [Accessed September 25, 2017].

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To See or Not to See: Pre-stimulus Alpha Phase Predicts Visual Awareness. *The Journal of Neuroscience, 29*(9), 2725–2732. http://doi.org/10.1523/JNEUROSCI.3963-08.2009.

Mattingley, J., Driver, J., Beschin, N., & Robertson, I. (1997) Attentional competition between modalities: extinction between touch and vision after right hemisphere damage. *Neuropsychologia*, 35(6), 867-880.

Maurer, D., & Maurer, C. (1988). The world of the newborn. New York: Basic Books.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1012311

McKeefry, D. J., & Zeki, S. (1997). The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain, 120*, 2229–2242.

Meier, B., and Rothen, N. (2007). When conditioned responses “fire back”: bidirectional cross-activation creates learning opportunities in synesthesia. *Neuroscience, 147*, 569–572. doi: 10.1016/j.neuroscience.2007.04.008.

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General, 116*(4), 323-336. http://dx.doi.org/10.1037/0096-3445.116.4.323

Meltzoff, A., & Borton, R. (1979) Intermodal matching by human neonates. *Nature, 282*, 403-404.

Meredith, M. A., & Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research, 365*(2), 350–4.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience, 7*(10), 3215–29.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience, 7*(10), 3215–29. PMID 3668625.

Moos, A., Simmons, D., Simner, J., & Smith, R. (2013). Color and texture associations in voice-induced synaesthesia. *Frontiers in Psychology, 4*, 568. Published online 2013 Sep 2. doi: 10.3389/fpsyg.2013.00568.

Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., & Matusz, P. J. (2016). The multisensory function of the human primary visual cortex. *Neuropsychologia*, *83*, 161-9. doi: 10.1016/j.neuropsychologia.2015.08.011. Epub 2015 Aug 11.

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*, 342-345

Naumer MJ, van den Bosch JJ (July 2009). "Touching sounds: thalamocortical plasticity and the neural basis of multisensory integration". *Journal of Neurophysiology, 102*(1): 7–8. PMID 19403745. doi:10.1152/jn.00209.2009.

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), 1471-7. doi: 10.1016/j.neuropsychologia.2012.02.032. Epub 2012 Mar 6.

O'Hanlon, E., Newell, F. N., & Mitchell, K. J. (2013). Combined structural and functional imaging reveals cortical deactivations in grapheme-color synaesthesia*. Frontiers in Psychology, 4*, 755. Published online 2013 Oct 30. doi: 10.3389/fpsyg.2013.00755.

Oliveri M (2003) Increased visual cortical excitability in ecstasy users: a transcranial magnetic stimulation study. *Journal of Neurology Neurosurgery and Psychiatry, 74*(8), 1136–1138.

Ozen S., Sirota A., Belluscio M. A., Anastassiou C. A., Stark E., Koch C., et al. (2010). Transcranial electric stimulation entrains cortical neuronal populations in rats. *Journal of Neuroscience, 30*, 11476–11485 10.1523/JNEUROSCI.5252-09.2010

Paulus, W., Nitsche, M. A., & Antal, A. (2016). Application of Transcranial Electric Stimulation (tDCS, tACS, tRNS):From Motor-Evoked Potentials Towards Modulation of Behaviour. *European Psychologist, 21*, 4-14. https://doi.org/10.1027/1016-9040/a000242.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review, 83*, 157-171.

Posner, M.I., Nissen, M.J., & Klein, M. (1976). Visual Dominance: An Information Processing Account of its Origins and Significance. *Psychological Review, 83*(2), 157-171.

Priori A. (2003). Brain polarization in humans: a reappraisal of an old tool for prolonged non-invasive modulation of brain excitability. *Clinical Neurophysiology, 114*(4), 589–95.

Rader, C. M., & Tellegen, A. (1987). An investigation of synesthesia. *Journal of Personality and Social Psychology, 52*, 981–987.

Ramachandran, V. S., & Hubbard, E. M. (2001). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings of the Royal Society Biological Sciences Series B, 268*(1470), 979–983. doi:10.1098/rspb.2000.1576

Rao, A., Nobre A. C., Alexander, I., & Cowey, A. (2007). Auditory evoked visual awareness following sudden ocular blindness: an EEG and TMS investigation. *Experimental Brain Research*, *176*(2)**,** 288-298.

Rich, A. N., & Karstoft, K-I. (2013). Exploring the benefit of synaesthetic colours: Testing for “pop-out” in individuals with grapheme–colour synaesthesia. *Cognitive Neuropsychology, 30*(1-2), 110-125.

Rich, A. N., Bradshaw, J. L., & Mattingley, J. B. (2005). A systematic, large-scale study of synaesthesia: implications for the role of early experience in lexical-colour associations. *Cognition, 98*, 53-84.

Rich, A. N., & Mattingley, J. B. (2002). Anomalous perception in synaesthesia: a cognitive neuroscience perspective. *Nature Reviews Neuroscience, 3*(1), 43-52.

Ridding, M. C., Brouwer, B., & Nordstrom, M. A. (2000). Reduced interhemispheric inhibition in musicians. *Experimental Brain Research, 133*, 249–53.

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology, 88*, 455-463.

Rogowska, A. (2011). Categorization of Synaesthesia. *Review of General Psychology, 15*(3), 213-227.

Rothen N., & Meier, B. (2014). Acquiring synaesthesia: insights from training studies. *Frontiers in Human Neuroscience, 8*:109. doi: 10.3389/fnhum.2014.00109. eCollection 2014.

Rothen, N., Wantz, A.-L., and Meier, B. (2011). Training synaesthesia. *Perception, 40*, 1248–1250. doi: 10.1068/p6984.

Rouw, R. (2013). *Synesthesia, Hyper-Connectivity, and Diffusion Tensor Imaging*. In Oxford Handbook of Synesthesia. : Oxford University Press. Retrieved 17 Sep. 2017, from http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199603329.001.0001/oxfordhb-9780199603329-e-025.

Rouw, R., & Scholte, H. S. (2007). Increased structural connectivity in grapheme-color synesthesia. *Nature Neuroscience, 10*, 792–797. doi:10.1038/nn1906.

Rouw, R., & Scholte, H. S. (2010). Neural basis of individual differences in synesthetic experiences. Journal of Neuroscience, 30(18), 6205-13. doi: 10.1523/JNEUROSCI.3444-09.2010.

Sachs, G. T. L. (1812). Historiae naturalis duorum leucaetiopum: Auctoris ipsius et sororis eius. Erlangen.

Sadato, N., Okada, T., Honda, M., & Yonekura, Y. (2002). Critical period for cross-modal plasticity in blind humans: a functional MRI study. *NeuroImage, 16*, 389–400.

Saenz, M., & Koch, C. (2008). The sound of change: visually-induced auditory synesthesia. *Current Biology*, *18*(15), R650–R651.

Sagiv, N., & Ward, J. (2006). *Cross-modal interactions: lessons from synaesthesia*. In: Martinez-Conde, S., Macknik, S. L., Martinez, L. M., Alonso, J-M., & Tse, P.U. (Eds). *Progress in Brain Research,* Volume 155, Part B, (pp.259-271). Elsevier.

Sagiv, N., Heer, J., Robertson, L. (2006). Does binding of synesthetic color to the evoking grapheme require attention? *Cortex, 42*, 232-242.

Sauseng, P., Klimesch, W., Heise, K. F., Gruber, W. R., Holz, E., Karim, A. A., Glennon, M., Gerloff, C., Birbaumer, N., & Hummel, F. C. (2009). Brain oscillatory substrates of visual short-term memory capacity. Current Biology, 19(21), 1846-52. doi: 10.1016/j.cub.2009.08.062.

Saygin, A. P., Driver, J., & de Sa, V. R. (2008). In the Footsteps of Biological Motion and Multisensory Perception: Judgments of Audiovisual Temporal Relations Are Enhanced for Upright Walkers. *Psychological Science, 19*(5), 469-475. doi: 10.1111/j.1467-9280.2008.02111.x

Scheef, L., Boecker, H., Daamen, M., Fehse, U., Landsberg, M. W., Granath, D. O., Mechling, H., & Effenberg. A. O. (2009). Multimodal motion processing in area V5/MT: evidence from an artificial class of audio–visual events. *Brain Research, 1252*, 94-104.

Schlaug, G., Jancke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia, 33*, 1047–55.

Schroeder, C. A., & Foxe, J. (2005). Multisensoryt contributions to low-level, 'unisensory' processing. *Current Opinion in Neurobiology, 15*, 454-458.

Shanon, B. (1982). Colour associates to semantic linear orders. *Psychological Research, 44*, 75–83.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature Brief Communications*, 408, 788, doi:10.1038/35048669.

Simner, J., Carmichael, D. A. (2015). Is synaesthesia a dominantly female trait? *Cognitive Neuroscience, 6*(2-3), 68–76.

Simner J., Bain A. E. (2013). A longitudinal study of grapheme–color synesthesia in childhood: 6/7 years to 10/11 years. *Frontiers in Human Neuroscience, 7*, 603 10.3389/fnhum.2013.00603.

Simner J., Harrold J., Creed H., Monro L., Foulkes L. (2009). Early detection of markers for synaesthesia in childhood populations. *Brain, 132* 57-6410.1093/brain/awn292.

Simner, J., & Holenstein, E. (2007). Ordinal linguistic personification as a variant of synesthesia. *Journal of Cognitive Neuroscience, 19*(4), 694-703.

Simner, J. (2012). Defining synaesthesia. *British Journal of Psychology*, *103*, 1–15. doi:10.1348/000712610X528305.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S.A., Fraser, C., Scott, K., & Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception, 35*, 1024-1033.

Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., & Oakley, D. A. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cognitive Neuropsychology, 22*(8), 1069-1085. DOI:10.1080/02643290500200122.

Sinke, C., Halpern, J. J., Zedler, M., Neufeld, J., .Emrich, H. K., & Passie, T. (2012).Genuine and drug-induced synesthesia: a comparison. *Consciousness and Cognition, 21*, 1419-34.

Sinnett, S., Spence, C. & Soto-Faraco, S. (2007).Visual Dominance and attention: The Colavita effect revisited. *Perception and Psychophysics, 69*(5), 673-686.

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *Neuroimage, 17*, 1613–22.

Smilek, D., Dixon, M. J., & Merikle, P. M. (2005). Synaesthesia: discordant male monozygotic twins. *Neurocase, 11*(5), 363-70.

Smilek, D., Moffatt, B. A., Pasternak, J., White, B. N., Dixon, M. J., & Merikle PM. (2002). Synaesthesia: a case study of discordant monozygotic twins. *Neurocase, 8*(4), 338-42.

Snyder A. C., Foxe J. J. (2010). Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. *Journal of Neuroscience, 30*, 4024–4032

Spence, C. (2009). Explaining the Colavita visual dominance effect. *Progress in Brain Research, 176*, 245-258.

Spence, C. (2011). Cross-modal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73,* 971-995.

Spence, C., & Parise, C. V. (2012). The cognitive neuroscience of crossmodal correspondences. *I-Perception, 3*(7), 410–412. http://doi.org/10.1068/i0540ic.

Stevenson, R. A., Fister, J. K., Barnett, Z. P., Nidiffer, A. R., & Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Experimental Brain Research, 219*(1), 121–137. Published online 2012 Mar 24. doi: 10.1007/s00221-012-3072-1, PMCID: PMC3526341, NIHMSID: NIHMS380838.

Strauß, A., Wöstmann, M., & Obleser,J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience, 8*, 350. Published online 2014 May 28. doi: 10.3389/fnhum.2014.00350.

Ten Oever, S., Romei, V., van Atteveldt, N., Soto-Faraco, S., Murray, M. M., & Matusz, P. J. (2015). The COGs (context, object, and goals) in multisensory processing. *Experimental Brain Research, 234*(5), 1307-1323. DOI: 10.1007/s00221-016-4590-z.

Terhune, D. B., Tai, S., Cowey, A., Popescu, T., & Cohen Kadosh, R. (2011). Enhanced cortical excitability in grapheme-color synesthesia and its modulation. Current Biology, 21(23):2006-9. doi: 10.1016/j.cub.2011.10.032.

Tomson, S. N., Avidan, N., Lee, K.,Sarma, A. K., Tushe, R., Milewicz, D. M., Bray, M., Leal, S. M., & Eagleman, D. M. (2011). The genetics of colored sequence synesthesia: Suggestive evidence of linkage to 16q and genetic heterogeneity for the condition. *Behavioural Brain Research, 223*(1), 48-52.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136.

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.

Tyler, C. (2005). *Varieties of synesthetic experience*. In: Robertson L, Sagiv N, editors. Synesthesia: Perspectives from Cognitive Neuroscience. Oxford University Press. pp. 34–44.

Van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory Integration: Flexible Use of General Operations. *Neuron, 81*(6), 1240-1253.

Van Leeuwen, T. M., Petersson, K. M., & Hagoort, P. (2010). Synaesthetic Colour in the Brain: Beyond Colour Areas. A Functional Magnetic Resonance Imaging Study of Synaesthetes and Matched Controls. *PLOS One*, https://doi.org/10.1371/journal.pone.0012074.

Vaudano, E., Legg, C. R., Glickstein, M. (1991). Afferent and efferent connections of temporal association cortex in the rat: a horseradish peroxidise study. *European Journal of Neuroscience, 3*, 317–330.

Vosskuhl, J., Strüber, D., & Herrmann, C. S. (2015). Transcranial alternating current stimulation. Entrainment and function control of neuronal networks. *Nervenarzt, 86*(12):1516-22. doi: 10.1007/s00115-015-4317-6.

Wade, A. R., Brewer, A. A., Rieger, J. W., & Wandell, B. A. (2002). Functional measurements of human ventral occipital cortex: retinotopy and colour. *Philosophical Transactions of the Royal Society London, Series B: Biological Science*, *357*, 963–973.

Wagner, K., & Dobkins, K. R. (2011). Synaesthetic associations decrease during infancy. *Psychological Science, 22*(8), 1067-72. doi: 10.1177/0956797611416250. Epub 2011 Jul 19.

Ward, J. & Banissy, M. J. (2015). Explaining mirror-touch synaesthesia. *Cognitive Neuroscience, 6*(2-3), 118-133.

Ward, J. (2013). Synesthesia. *Annual Review of Psychology*, *64*, 49–75. doi:10.1146/annurevpsych-113011-143840.

Ward, J., Li, R., Salih, S., & Sagiv, N. (2007). Varieties of grapheme-colour synaesthesia: a new theory of phenomenological and behavioural differences. Consciousness and Cognition, 16(4), 913-31. Epub 2006 Nov 27.

Ward, J., & Sagiv, N. (2007). Synaesthesia for Finger Counting and Dice Patterns: A Case of Higher Synaesthesia? *Neurocase, 13*, 86-93.

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound–colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex*, *42*, 264–280. doi:10.1016/S00109452(08)70352-6.

Ward, J., Li, R., Salih, S., & Sagiv, N. (2007). Varieties of grapheme-colour synaesthesia: A new theory of phenomenological and behavioural differences. *Conscious and Cognition, 16*, 913–931.

Ward, J., & Mattingley, J. B. (2006). Synaesthesia: An overview of contemporary findings and controversies. *Cortex, 42*, 129-136.

Ward, J., & Simner, J. (2003). Lexical-gustatory synaesthesia: linguistic and conceptual factors. *Cognition, 89*(3), 237-61.

Wasserman, E. M., Greenberg, B. D., Nguyen, M. B., & Murphy, D. L. (2001). Motor cortex excitability correlates with an anxiety-related personality trait. *Biological Psychiatry, 50*(5), 377-82.

Watson, A.B., Ahumada, A.J. (1985) Model of human motion sensing. *Journal of the Optical Society of America, A2*, 322-342.

Watson, M. R., Akins, K. A., Spiker, C., Crawford, L., & Enns, J. T.. (2014). Synesthesia and learning: a critical review and novel theory. *Frontiers in Human Neuroscience, 8*, 98. Published online 2014 Feb 28. doi: 10.3389/fnhum.2014.00098 PMCID: PMC3938117.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception Psychophysics, 33*(2):113-20.

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.

Appendices

Appendix 1: Motion Energy extracted from each video, ranked from least to most motion.

Stimulus	Motion Energy
Hammer Hitting Nail	0.095
TV Smash	0.103
Multiple Clocks	0.104
Golfer	0.115
Silent Scream	0.121
Tennis Ball Bounce (Slow Motion)	0.152
Bumper Cars	0.157
Footsteps Beach	0.160
Ballerina	0.178
Bouncing Tennis Ball	0.196
Riviera	0.196
Rollercoaster	0.207
Newton's Cradle	0.224
Fireworks	0.248
Bouncing Black Balls	0.262
Punch bag Rapid Punches	0.273
Blue Twinkling Lights	0.275
LED Squares	0.327
Disco Lights	0.332
Police Car Lights	0.335
Orange Twinkling Lights	0.351
Police Car Lights & Passing Traffic	0.362
Random Moving Dots	0.450
Spinning Dot Globe	0.821

Appendix 2: Mean rating and standard deviation for each questionnaire item presented by previous awareness of experiencing vEAR

	Previously aware of hearing visual events?					
	No		Not Sure		Yes	
Stimulus	M	SD	M	SD	M	SD
Ballerina	0.337	0.844	0.771	1.245	1.309	1.516
Spinning Dot Globe	0.429	0.945	0.897	1.319	1.572	1.579
Random Moving Dots	0.433	0.922	0.907	1.279	1.556	1.583
Riviera	0.432	0.924	0.927	1.328	1.706	1.632
Bumper Cars	0.562	1.046	1.149	1.416	1.859	1.576
Disco Lights	0.601	1.086	1.170	1.420	2.045	1.644
Tennis Ball Bounce (Slow motion)	0.673	1.102	1.320	1.458	1.937	1.595
LED Squares	0.613	1.085	1.283	1.466	2.227	1.712
Orange Twinkling Lights	0.669	1.134	1.386	1.509	2.264	1.690
Police car Lights & Passing Traffic	0.690	1.143	1.421	1.523	2.178	1.644
Multiple Clocks	0.739	1.209	1.412	1.549	2.266	1.779
Blue Twinkling Lights	0.750	1.171	1.450	1.490	2.282	1.660
Footsteps Beach	0.859	1.222	1.590	1.517	2.246	1.658
Golfer	0.903	1.285	1.654	1.605	2.329	1.667
Police car Lights	0.858	1.256	1.614	1.568	2.517	1.618
Rollercoaster	0.939	1.332	1.742	1.631	2.412	1.693
Bouncing Tennis Ball	1.032	1.275	1.759	1.518	2.390	1.554
Bouncing Black Balls	1.043	1.340	1.866	1.606	2.720	1.607
TV Smash	1.136	1.401	1.955	1.646	2.743	1.664
Punch bag Rapid Punches	1.197	1.413	2.059	1.592	2.887	1.559
Silent Scream	1.285	1.483	2.193	1.686	2.879	1.722
Newton's Cradle	1.307	1.463	2.234	1.653	3.049	1.567
Fireworks	1.227	1.485	2.239	1.682	3.203	1.601
Hammer Hitting Nail	1.428	1.504	2.329	1.648	3.194	1.550
Grand Total	M	SE	M	SE	M	SE
	0.839	0.013	1.555	0.019	2.324	0.021