

**MULTISENSORY PERCEPTUAL DECISION-MAKING
RESOLVED IN TIME**

by

STEFFEN BÜRGERS

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

School of Psychology
College of Life and Environmental Sciences
University of Birmingham
March 2019

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Perceptual decision-making, inferring the “right” course of action from sensory information, is a vital process for any living organism. To get reliable perceptual estimates it is useful to integrate information from multiple modalities (e.g. vision and audition), provided they originate from the same source. In this thesis we scrutinized multisensory perceptual decision-making in time with a simple one versus two-flash segregation task, while presenting concurrent, irrelevant beeps. When stimuli are presented close in time, disparate flash-beep pairings lead to altered flash reports (e.g. 1-flash & 2-beep stimulation elicits an illusory second flash), a phenomenon known as the sound induced flash illusion. In chapter 1 we lay theoretical foundations of multisensory integration and perceptual decision making with a focus on the sound induced flash illusion. In chapter 2 we examine the temporal determinants of the sound induced flash illusion. In chapter 3 we characterize and compare the neural representations of perceptual choice outcomes throughout the perceptual decision-making process. In chapter 4 we investigate the long-standing hypothesis that the frequency of alpha oscillations indexes visual temporal resolution. Finally, in chapter 5 we summarise the findings obtained in the empirical chapters and discuss future directions.

*For my parents,
Sigrid & Norbert;
And for my family,
Tara & Judo*

Acknowledgements

First, I want to thank my supervisor, Uta Noppeney, for supporting me throughout this PhD. You are truly a scientific role model in your curiosity, way of critical thinking, methodological prowess and ability to place ideas and findings into the larger theoretical landscape.

Next, I want to thank my colleagues and friends in the Computational Cognitive Neuroimaging Group and psychology department whom I spent a large part of my life with inside and outside the office. Your contribution to this work is immeasurable in countless discussions about methods and theory, as well as emotional and motivational support. You are all great scientists and people. I hope our paths will cross again.

I also want to thank my parents Sigrid and Norbert, and my brother Jens for believing in me, regardless of what path in life I pursue.

Finally, I wish to express my gratitude and love to my wife Tara and stepson Judo. You have enriched my life in countless ways and I am very proud to be part of your chosen family.

Table of Contents

Acknowledgements.....	i
List of figures	vii
List of tables.....	ix
Chapter 1 - General introduction to multisensory perceptual decision-making	1
Principles of multisensory integration	2
Multisensory neural pathways	3
Investigating multisensory integration via perceptual illusions	4
Causal inference and optimality in multisensory integration	6
The sound induced flash illusion: Fission and fusion	7
Inferring perceptual and decisional processes from behavioural responses	10
Modelling psychometric functions.....	13
Response speed as a proxy for decision time	16
Neural correlates of perceptual choice formation	17
The role of alpha frequency in visual temporal resolution	19
Study outline and overview of experiments	20
Advantages and limitations of a large-scale within-subject design.....	23
Mechanisms of (multi-)sensory integration captured by signal detection theory, psychometric functions and response times	25
Synopsis of research questions	27
Chapter 2 - Temporal determinants of the sound induced fission and fusion illusion	28
Abstract.....	28
Introduction	29

Results.....	32
Psychometric function model and temporal resolution.....	32
Signal detection model: d'	38
Model free analysis	40
Summary of results	44
Discussion	45
Methods.....	49
Participants	49
Stimuli	49
Design and procedure.....	49
Experimental setup	51
Psychometric function analysis	51
Signal detection analysis.....	53
Accuracy and response time analysis	53
Statistical comparisons.....	54
Supplementary materials	55
Chapter 3 - Decoding real and illusory flash representations during perceptual decision-making	
.....	57
Abstract.....	57
Introduction	57
Results.....	62
Behavioural results.....	63
Univariate ERP results of perceptual outcome	65

Multivariate pattern decoding of perceptual outcome	69
Discussion	77
Methods	80
Participants	80
Stimuli	81
Design and procedure	81
Experimental setup	82
Staircase procedure to titrate perceptual threshold	82
Behavioural analyses	83
EEG preprocessing	85
EEG univariate analyses	86
EEG multivariate analyses	87
Supplementary materials	90
Chapter 4 - What is the role of posterior alpha frequency in visual temporal resolution?	93
Abstract	93
Introduction	94
Results	96
State alpha frequency does not influence visual temporal resolution	98
State alpha power does not generally index cortical excitability	101
Alpha phase might play a role in visual temporal resolution	101
Trait alpha frequency does not index visual temporal resolution	103
Control analyses	106

Discussion	106
Methods	111
Participants	112
Stimuli	112
Design and procedure	112
Experimental setup	114
EEG recording and preprocessing.....	115
Computation of pre-stimulus spectral features	115
Disambiguating between perceptual sensitivity and bias	116
Relating spontaneous oscillatory fluctuations to perceptual sensitivity and bias	117
Trait alpha peak frequency and power estimation	118
Fitting psychometric functions to quantify individual windows of perception	119
Contrasting trait alpha frequency and power with perceptual window length	120
Alpha source separation.....	120
Control analysis of event related potentials	122
Supplementary materials	123
Interpreting source level analyses in EEG data	123
Chapter 5 - General discussion.....	133
Summary of empirical findings	133
Chapter 2: Temporal determinants of the sound induced fission and fusion illusion ...	133
Chapter 3: Decoding real and illusory flash representations during perceptual decision making	134

Chapter 4: What is the role of posterior alpha frequency in visual temporal resolution?	
.....	135
Contributions, limitations and future directions	136
Conclusion	142
References	144

List of figures

Chapter 1: General introduction to multisensory perceptual decision-making

Figure 1.1 Neural and behavioural responses as a function of stimulus onset asynchrony.

Figure 1.2 Two equal-variance Gaussian signal detection model

Figure 1.3 Psychometric function model with logistic sigmoid

Figure 1.4 Study overview and design.

Figure 1.5 Exemplar testing schedule.

Chapter 2: Temporal determinants of the sound induced fission and fusion illusion

Figure 2.1 Study design and trial examples

Figure 2.2 Perceptual performance over time and model illustrations

Figure 2.3 Model based perceptual discriminability analysis of one versus two flashes

Figure 2.4 Model free perceptual discriminability analysis of one versus two flashes

Supplementary Figure 2.1 Typical study time table

Chapter 3: Decoding real and illusory flash representations during perceptual decision-making

Figure 3.1 Study design and example trial presentation

Figure 3.2 Behavioural results

Figure 3.3 ERPs of “see 2” versus “see 1” perceptual outcomes

Figure 3.4 Decoding perceptual outcomes

Figure 3.5 Classifier performance for confidence-matched behavioural outcomes.

Supplementary Figure 3.1 Decoding time courses locked to second stimulus onset

Supplementary Figure 3.2 Bayes factors for decoding accuracies locked to response

Chapter 4: What is the role of posterior alpha frequency in visual temporal resolution?

Figure 4.1 Hypothesis, study design and trial examples

Figure 4.2 Perceptual sensitivity by pre-stimulus alpha frequency

Figure 4.3 Psychometric functions and trait alpha frequency results

Supplementary Figure 4.1 Typical study time table

Supplementary Figure 4.2 Perceptual sensitivity by occipital and parietal alpha frequency

Supplementary Figure 4.3 Bias by pre-stimulus alpha frequency

Supplementary Figure 4.4 Perceptual sensitivity and bias by pre-stimulus alpha power

Supplementary Figure 4.5 Perceptual sensitivity and bias by pre-stimulus alpha phase

Supplementary Figure 4.6 Event related potentials differ between SOAs

Supplementary Figure 4.7 Trait alpha frequency control analysis

List of tables

Chapter 3: Decoding real and illusory flash representations during perceptual decision-making

Table 3.1 Time collapsed within condition decoding accuracies of perceptual outcome

Chapter 4: What is the role of posterior alpha frequency in visual temporal resolution?

Table 4.1 Trait alpha peak frequency versus perceptual window length

Supplementary Table 4.1 Source level between subject analyses of alpha peak frequency and perceptual discrimination threshold

Supplementary Table 4.2 Between subject analyses of alpha power and perceptual discrimination threshold

Chapter 1 - General introduction to multisensory perceptual decision-making

Perceptual decision-making is the process of using sensory signals to decide on the state of the external world. There is a large variety of different sensory organs throughout the animal kingdom, and a plethora of specializations within those modalities between species. Crucially, there is a trade-off between how well a given sense represents a specific aspect of the environment and how well it can generalize to others. This is true both on the larger scale of evolution, but also during neurodevelopment and perceptual learning throughout life (Gilbert, Sigman, & Crist, 2001). For example, human hearing is excellent at resolving incoming information in time, but has a coarse spatial resolution. Conversely, vision provides highly accurate estimates in space, but lacks temporal acuity. As both the environment and the sensory systems themselves are laden with noise, it is highly advantageous to combine information from multiple senses, provided there is a common underlying source (Alais, Newell, & Mamassian, 2010). How does the brain do this? How does it know when integrating information is appropriate? What are the underlying neural mechanisms? And how does perceptual information ultimately lead to a behavioural choice? In this chapter I want to briefly address these questions and outline the current landscape of (audio-visual) multisensory integration (MSI) as a whole and the sound induced flash illusion, in particular. This introduction will lay a foundation for the questions asked in the empirical chapters: What are the temporal determinants of 1 versus 2-flash segregation in the context of 0-2 beeps (chapter 2)? When are perceptual choices (1 or 2 flashes) represented in neural data throughout the decision process, and do they differ depending on auditory context (0-2 beeps; chapter 3)? Does the frequency of the prominent alpha rhythm of the brain represent a general marker of visual temporal resolution

(chapter 4)?

Principles of multisensory integration

Multisensory integration (MSI), as used in this thesis and throughout much of the literature, refers to a binding of unisensory inputs such that the multisensory neural response or perceptual outcome are different compared to if only a single modality had been presented (Stein et al., 2010). Early scientific endeavours into MSI date back to the 19th century (Urbantschitsch, 1880), but the arguably most important discoveries to spike interest in how the senses interact were made during the 1970s, demonstrating MSI in the response profiles of neurons in the superior colliculus (SC) of anaesthetized cats (Meredith, Nemitz, & Stein, 1987; Meredith & Stein, 1983, 1986; Morrell, 1972; Pöppel, 1973; Stein & Meredith, 1993; Stein, Stanford, & Rowland, 2014). This research revealed three cardinal rules of MSI: First, stimuli are more likely to be integrated if presented in close spatial proximity (spatial rule) (Meredith & Stein, 1986). Second, integration is more likely when stimuli occur close in time (temporal rule) (Meredith et al., 1987). And third, the net effect of multisensory integration is inversely proportional to the signal strengths (or reliability) of the unisensory inputs (inverse effectiveness rule) (Meredith & Stein, 1983). In other words, when unisensory input is weak, the benefit of multisensory integration is accordingly larger, leading to a stronger response than a simple linear combination of the unisensory signals (superadditivity) (Stein & Meredith, 1993). Furthermore, each of these stimulus factors affects the way any of the other factors influence MSI (Carriere, Royal, & Wallace, 2008; Ghose & Wallace, 2014; Krueger, Royal, Fister, & Wallace, 2009; Royal, Carriere, & Wallace, 2009). For example, the temporal profile by which stimuli are likely to be integrated changes as the spatial distance between the inputs is increased (Ghose & Wallace, 2014). Intriguingly, the same principles that govern MSI in SC neurons also seem to apply, for the most part, to MSI at the behavioural level (Figure 1.1) (Alais

et al., 2010; Diederich & Colonius, 2004; Hillock, Powers, & Wallace, 2011; Powers, Hillock, & Wallace, 2009; Spence, 2018; Wallace et al., 2004; Wallace & Stevenson, 2014).

Even though the three stimulus-driven rules explain much of both neural and behavioural data, MSI is also dependent on many other factors including task and stimulus context (van Atteveldt, Murray, Thut, & Schroeder, 2014), emotional and motivational state (Bruns, Maiworm, & Röder, 2014; Maiworm, Bellantoni, Spence, & Röder, 2012), prior expectation (Gau & Noppeney, 2016), attention (Macaluso et al., 2016) and awareness (Deroy, Spence, & Noppeney, 2016). Moreover, multisensory response profiles are malleable and reflect regularities in the environment that are likely to lead to useful internal representations (Stein et al., 2014). Indeed, during early neurodevelopment multisensory neural response properties emerge as experience with the environment and its temporal and spatial patterns of cross-modal events is accrued (Stein et al., 2014). This developmental process has a critical time period within which SC requires afferent input from unisensory neurons in association cortex, without which MSI cannot be learned later in life, even if the pathways are reactivated (Jiang, Jiang, & Stein, 2006; Jiang, Wallace, Jiang, Vaughan, & Stein, 2001; Stein et al., 2014; Stein, Wallace, Stanford, & Jiang, 2002). If multisensory perceptual learning during this critical period is successful, it continues throughout life and can happen even on very short time scales, allowing for (partial) adaptations to new environments (Arnold & Yarrow, 2011; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Stein et al., 2014).

Multisensory neural pathways

As alluded to earlier, the SC is the most thoroughly studied and well understood locus of multisensory neurons in the brain. It has a direct role in orienting behaviour both overtly by performing a saccade toward an external stimulus and covertly by allocating attention to it (Gandhi & Katnani, 2011; Kustov & Robinson, 1996). It receives inputs from unisensory

brainstem nuclei for vision, audition and somatosensation, as well as cortical inputs from multisensory areas such as the anterior ectosylvian and lateral suprasylvian sulci in the cat brain (Jiang et al., 2001). Interestingly, the neurons projecting to SC from these “association areas” are again unisensory (Stein et al., 2002). Moreover, when cortical projections to SC neurons are temporarily silenced, neurons will still respond to unisensory input from multiple senses, but no longer integrate them (Alvarado, Stanford, Vaughan, & Stein, 2007). Cortical input is thus essential for MSI in the SC.

In the cortex primary auditory and visual areas receive subcortical inputs from their respective unisensory pathways, as well as from the SC. Moreover, it has been discovered relatively recently that there are direct pathways between primary auditory and visual cortex, in particular to peripheral V1 (Clavagnier, Falchier, & Kennedy, 2004; Falchier, Clavagnier, Barone, & Kennedy, 2002; Rockland & Ojima, 2003). Multisensory integration was observed in both A1 and V1 (Driver & Noesselt, 2008; Ghazanfar, Maier, Hoffman, & Logothetis, 2005; Kayser & Logothetis, 2007; Lakatos, Chen, O’Connell, Mills, & Schroeder, 2007; Molholm et al., 2002). Such discoveries led to the theory that the entire neocortex is essentially multisensory (Driver & Noesselt, 2008; Ghazanfar & Schroeder, 2006). As mentioned before, however, there are also designated “association” cortical regions whose primary function is to integrate unisensory inputs. For audio-visual integration in humans those areas are predominantly located in the posterior parietal cortex and the superior temporal sulcus (STS) (Driver & Noesselt, 2008). Interestingly, multisensory neurons in the cortex are much less likely to show sub-additive or supra-additive responses like their sub-cortical counterparts in the SC, but instead tend to exhibit simple additive responses (Driver & Noesselt, 2008).

Investigating multisensory integration via perceptual illusions

Experimentally, MSI is often investigated with paradigms that evoke perceptual illusions in one modality due to another. An example in the spatial domain is the ventriloquist effect (VE). The classical VE occurs when participants are asked to locate an acoustic stimulus and an uninformative, but concurrent visual stimulus is presented. Depending on the spatial proximity of the two stimuli, as well as stimulus reliability, it is likely that participants merge the two unisensory inputs (Alais & Burr, 2004). Similarly, a sound can capture visual perception in time and even induce a percept of a second flash when only one flash is presented if a sound was synchronously presented with the first flash and a second sound occurs shortly after (Shams, Kamitani, & Shimojo, 2002). This phenomenon has become known as the double flash illusion or fission illusion (Shams, Kamitani, & Shimojo, 2000; Shams et al., 2002). The reason perceptual illusions are an excellent way of studying multisensory integration is that they inform us about the physical properties that cue the brain to integrate multimodal information. In the VE the manipulated dimension is spatial disparity, while in the fission illusion the manipulated dimension is stimulus onset asynchrony (SOA). The mapping between a physical attribute and perceptual responses is called a psychometric function (Figure 1.1). When the relative timing between stimuli is the manipulated variable, the time range within which they are integrated is called the temporal binding window (TBW).

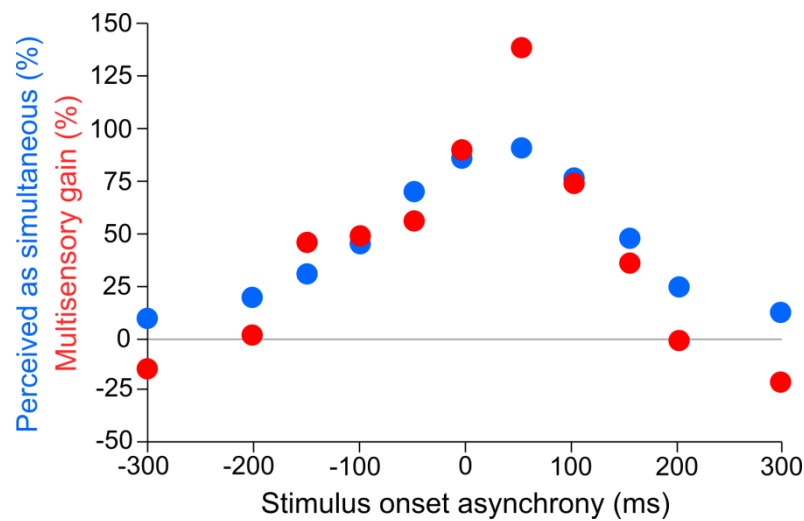


Figure 1.1. *Neural and behavioural responses as a function of stimulus onset asynchrony. Temporal tuning function of a representative neuron in cat superior colliculus reacting to audio-visual stimulation (red); and representative psychometric function of human simultaneity judgments in response to audio-visual stimulation (blue). Note the resemblance between neural and behavioural response profiles. Adapted from (Wallace & Stevenson, 2014).*

Causal inference and optimality in multisensory integration

In our everyday world the brain is bombarded with information from multiple senses. As a result, it has to constantly judge whether sensory input streams originate from the same source or separate outside events (Shams & Beierholm, 2010). In the former case it is ideal to combine the information into a unified percept that is less noisy than either of its parts. In the latter case it is desirable to segregate the input signals. When it can be assumed that stimuli from different modalities come from the same source the best possible estimate can be obtained by adding the unisensory estimates weighted by their respective reliabilities (forced fusion) (Alais & Burr, 2004; Ernst & Bühlhoff, 2004; Hillis, Ernst, Banks, & Landy, 2002). Conversely, when it is known that incoming information arises from separate sources the best estimate is obtained by processing the streams completely independently (full segregation) (Körding et al.,

2007). However, the number of sources in the environment is not usually known a priori, and the decision whether or not one or multiple sources exist has to be made based on the sensory input. Research has shown that this causal inference problem is solved by the brain in a near optimal fashion via Bayesian inference (Shams & Beierholm, 2010). For example, behavioural responses in a ventriloquist paradigm are best explained by a hierarchical Bayesian Causal Inference (BCI) model that incorporates uncertainty about the number of sources by adding the full segregation and full integration estimates weighted by their respective probabilities (Körding et al., 2007; Rohe & Noppeney, 2015a, 2015b). These weights in turn are based on parameters such as spatial and temporal proximity, as well as prior expectations and experiences with similar sensory input (van Atteveldt et al., 2014). Interestingly, this model averaging approach, even though optimal from a Bayesian perspective, is not applied by all study participants and choice of strategy seems to depend on the task specifics, suggesting that human performance is optimal in some cases, but not others (Körding et al., 2007; Wozny, Beierholm, & Shams, 2010).

The sound induced flash illusion: Fission and fusion

Following the Bayesian models of MSI, the VE illustrates the superiority of vision over audition in the spatial dimension. On the other hand, temporal acuity is much higher in the auditory domain compared to vision and this temporal superiority is illustrated by the sound induced flash illusion (SIFI) (Shams et al., 2000, 2002). In a SIFI paradigm one or more flashes are presented in conjunction with a variable number of sounds. The time interval between stimuli is on the order of 20-240 ms (Apthorp, Alais, & Boenke, 2013; Cecere, Rees, & Romei, 2015), and visual and auditory stimulus streams are synchronized, either starting simultaneously (Abadi & Murphy, 2014; Keil, Müller, Hartmann, & Weisz, 2014; Meylan & Murray, 2007) or being slightly shifted in time (Innes-Brown & Crewther, 2009; Kumpik,

Roberts, King, & Bizley, 2014; Mishra, Martinez, & Hillyard, 2008; Mishra, Martinez, Sejnowski, & Hillyard, 2007; Shams et al., 2002). If the number of flashes is smaller than the number of sounds participants are prone to perceive more flashes than are actually presented. This is the previously mentioned fission illusion. Conversely, when more flashes than sounds are presented participants are more likely to report seeing fewer flashes compared to unisensory visual trials (Andersen, Tiippana, & Sams, 2004; Shams et al., 2002). This is known as fusion illusion. Computational models of the SIFI have demonstrated that the number of perceived flashes is best explained by Bayesian inference when a varying number of visual and auditory stimuli are presented (Rohe, Ehrlis, & Noppeney, 2018; Shams, Ma, & Beierholm, 2005; Wozny, Beierholm, & Shams, 2008).

The fission illusion has been more widely studied, and is in some way more interesting, because an illusory flash is elicited without any visual input. Instead, in the fusion illusion two visual events may simply be shifted in time and integrated (Cuppini, Magosso, Bolognini, Vallar, & Ursino, 2014). Somewhat surprisingly then, the fission illusion is almost impervious to feedback learning and generally reported as indistinguishable from a real second flash (Abadi & Murphy, 2014; Rosenthal, Shimojo, & Shams, 2009). However, when monetarily rewarded for performance accuracy, participants were able to learn to segregate illusory from real flashes (Rosenthal et al., 2009). The fission illusion has been established as a more robust phenomenon than the fusion illusion, with lower between subject variability (Mishra et al., 2008, 2007), but this trend reverses when larger numbers of flashes and beeps are presented (Abadi & Murphy, 2014). Fission occurs both in the peripheral and central visual field, but is stronger in the periphery (Chen, Maurer, Lewis, Spence, & Shore, 2017; Kaposvári, Bognár, Csibri, Utassy, & Sári, 2014). Fusion is stronger in the centre and oftentimes not observed in the periphery (Chen et al., 2017; Kaposvári et al., 2014). A possible explanation for this could be that the

fusion illusion benefits from slower parvocellular processing (ventral pathway) compared to the faster magnocellular stream (dorsal pathway) (Kaposvári et al., 2014). Interestingly, the fission illusion was found to be a reliable phenomenon in both young and elderly adults, whereas susceptibility to the fusion illusion declines with age (Bolognini et al., 2016; DeLoss & Andersen, 2015; McGovern, Roudaia, Stapleton, McGinnity, & Newell, 2014).

Functional magnetic resonance imaging (fMRI) in humans revealed that both fission and fusion illusion lead to activity in retinotopically defined primary visual cortex V1 and V2, with an illusory second flash (fission illusion) leading to enhanced activity similar to a real second flash, and a fused single flash (fusion illusion) leading to attenuated activity similar to a real single flash (Watkins, Shams, Josephs, & Rees, 2007; Watkins, Shams, Tanaka, Haynes, & Rees, 2006). At the same time superior temporal sulcus (STS) displayed enhanced activity for both illusions (Watkins et al., 2007, 2006). In contrast, event related potentials (ERPs) showed a modulation by an illusory double flash measured relative to the onset of the second stimulus, as early as ~40-75 ms in visual cortex and ~50-80 ms in superior temporal cortex (Mishra et al., 2007). Instead, the fusion illusion modulated ERPs only after ~100-130 ms in superior temporal cortex and ~160-180 ms in visual cortex (Mishra et al., 2008). Furthermore, transcranial direct current stimulation (tDCS) applied over V1 or STS modulates fission, but not fusion (Bolognini, Rossetti, Casati, Mancini, & Vallar, 2011). Clinical research showed that the fission illusion is reduced in patients with visual field deficits proportional to lesion size in early visual cortex (low level sensory processing deficits), but is unaffected in patients with unilateral spatial neglect (higher order attentional processing deficits) (Bolognini et al., 2016). On the other hand, the fusion illusion was only observed for patients with unilateral spatial neglect, but neither for patients with visual field deficits nor healthy controls (Bolognini et al., 2016).

All in all, despite an impressive amount of studies employing SIFI paradigms, the underlying neural mechanisms are only partially understood. Both fission and fusion are related to activity in visual and multisensory cortex, but the temporal dynamics seem to be different. Fission seems to be more related to feedforward processing from auditory cortex to visual cortex, while the fusion illusion might be more related to polysensory feedback projections to V1. However, an important caveat with interpreting previous research is that results are often based on simple perceptual outcome measures, which can be confounded by biases that are independent of perceptual processing (Macmillan & Creelman, 2005). In the upcoming section we will address possible solutions to this issue.

Inferring perceptual and decisional processes from behavioural responses

When researching human perceptual decision-making we infer perceptual and cognitive processes from behavioural responses. As only the final outcome is observed this is not a trivial task. After all, perception is influenced not only by incoming sensory input, but also an array of cognitive factors such as attention and prior expectations (Gau & Noppeney, 2016; Macaluso et al., 2016). Moreover, participants will treat the same perceptual information differently depending on their decision rule (Macmillan & Creelman, 2005). Lastly, the motor response itself can be subject to errors, leading to actions that were not even intended (so called lapses) (Wichmann & Hill, 2001). How can we deal with this inferential dilemma? A classical and powerful framework for doing so is signal detection theory (SDT) (Macmillan & Creelman, 2005).

Consider a simple example from a detection task in which on any given trial a signal is either presented or not. In the context of this thesis we will encounter this situation whenever we compare 1-flash (noise) with 2-flash (signal) trials. If the signal is sufficiently weak (or the

noise sufficiently strong), participants will inevitably make mistakes that are purely due to perceptual noise, i.e. they will report “see 2”, even though 1-flash was presented. This is called a false alarm (FA). Alternatively, participants can correctly reject the presence of a signal on noise trials (correct rejection, CR). On signal trials participants can report the signal (hit) or fail to report it (miss). Oftentimes, researchers only consider hits, or correct responses (hits and correct rejections). However, each of these measures is not only affected by perceptual sensitivity (how well signal and noise can be perceptually distinguished), but also various types of biases such as response bias (e.g. prefer answering “see 2” on all trials) (Witt, Taylor, Sugovic, & Wixted, 2015).

A clever way to segregate perceptual sensitivity from bias is to model signal and noise distributions as Gaussians with unit standard deviation lying on an internal decision axis (Macmillan & Creelman, 2005). The noise distribution can be seen as residing on the 0-point of this axis, but due to random fluctuations in sensory processing any given sample taken from it (a given 1-flash trial presentation) can have a positive value, possibly exceeding the internal criterion for accepting it as a signal (Figure 1.2). Given empirical values of the proportions of false alarms and hits, it is possible to quantify perceptual sensitivity as the distance between the means of the two Gaussians, denoted as d' :

$$(1) \quad d' = z(HR) - z(FAR),$$

where z denotes the inverse cumulative distribution function of the standard normal distribution, thus converting probabilities (hit rate or false alarm rate) to z-scores. In addition, the decision criterion can be calculated and if it is measured relative to the crossing-point of the two Gaussians it is a measure of bias, denoted by C :

$$(2) \quad C = -0.5(z(HR) + z(FAR)).$$

Crucially, d' and C are independent of each other, provided that the assumptions of the model are correct (sensory noise is Gaussian and equal for signal and noise). This method does not solve the problem, however, of dissociating different types of biases from each other. For example, a simple response bias can lead to the same value of C as a perceptual shift of both signal and noise along the decision axis (Witt et al., 2015). Such a shift is possible when overall activity in visual cortex is boosted by attention or cross-modal input from non-visual sensory cortices (Cecere et al., 2015; Iemi, Chaumon, Crouzet, & Busch, 2017; J. Lange, Oostenveld, & Fries, 2013).

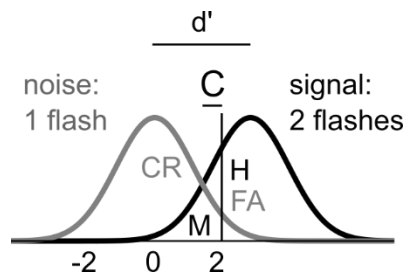


Figure 1.2. Two equal-variance Gaussian signal detection model. Both signal (black) and noise (grey) vary from trial to trial on an internal decision axis measured in standardized units. The ordinate denotes probability density. When a sample is taken from either distribution exceeds the internal decision criterion (black line) the observer reports a signal. The distance d' between the Gaussian expected values denotes perceptual sensitivity. The distance from the decision threshold to the ideal threshold (crossing point of the Gaussians) is denoted by C and a straightforward measure of bias. If the model assumptions are correct (Gaussian noise with equal-variance for both signal and noise), d' and C are independent. CR , correct rejection; M , miss; H , hit; FA , false alarm.

A different approach to measure perceptual sensitivity is to design the experiment in such a way that bias is minimized. This can be done with two-interval forced choice (2IFC) procedures. Specifically, on a given trial not only one event of interest is presented (either signal or noise), but both are displayed to the participant in random order (e.g. 1-flash, short delay, 2-

flash). As the order of signal and noise is random, a bias toward the second interval is overall not going to affect performance accuracy after pooling over signal leading and lagging presentations. Moreover, it is less likely that participants will adopt an interval bias compared to a bias for a perceptual outcome (Klein, 2001; Macmillan & Creelman, 2005). It is nevertheless desirable to minimize interval bias as much as possible, as it is invalid to simply average over interval orders if such a bias exists (García-Pérez & Alcalá-Quintana, 2011; Klein, 2001). This can be achieved (partially) by changing the response-key mapping to perceptual outcomes over the course of the experiment to avoid preferences in the response hand.

Modelling psychometric functions

The simple signal detection model described above only considers the absence of a signal (i.e. noise), and one signal level. However, when measuring the impact of SOA on MSI, we consider multiple signal levels, one for each SOA. In Figure 1.1 we illustrated simultaneity judgments of audio-visual stimuli as a function of SOA. Note that the progression of this psychometric function (mapping of physical variable to perceptual response probabilities) is sigmoid shaped when considering either only A-lagging and V-leading or A-leading and V-lagging conditions. In fact, this shape is readily observed in almost any psychophysical context, at least when a single physical variable is manipulated (Kingdom & Prins, 2016; Macmillan & Creelman, 2005). Consequently, researchers have developed ways to model psychometric curves with cumulative Gaussian, logistic and other sigmoid shaped functions, thereby capturing properties of the data with only a few model parameter estimates (Figure 1.3) (Kingdom & Prins, 2016; Klein, 2001; Macmillan & Creelman, 2005; Wichmann & Hill, 2001). The most common framework within which to do so is again signal detection theory. Hence, in a yes-no design we can imagine one Gaussian for each signal strength and one dividing criterion

that determines the decision boundary. The psychometric function is formally described by four parameters:

$$(3) \quad \psi(x; \alpha, \beta, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda) F_L(x; \alpha, \beta)$$

with

$$F_L(x; \alpha, \beta) = \frac{1}{1 + \exp(-\beta(x - \alpha))},$$

where x denotes SOA, F_L is the logistic function, and parameters α, β, γ and λ describe threshold (inflection-point), slope, guess-rate and lapse-rate, respectively. A popular choice is to fit the model to proportion of hits (in our yes-no task “see 2”), or proportion correct (for both signal and noise trials) (Cecere et al., 2015; Macmillan & Creelman, 2005; Samaha & Postle, 2015). In the latter case (and in 2IFC task designs) guess-rate should be 50 %, dictated by the amount of possible alternatives (although it should be noted that “guesses” are nevertheless based on the signal detection model and an evaluation of the internal decision variable (Kingdom & Prins, 2016; Klein, 2001)). Instead, lapse-rate captures responses that were independent of the perceptual evidence (e.g. random finger twitches). Finally, threshold and slope describe the location (in terms of SOA) and steepness of the function. Both threshold and slope estimates can be influenced by decision/response bias (García-Pérez & Alcalá-Quintana, 2011; Gold & Ding, 2013; Klein, 2001). In the 2IFC task, as interval bias should theoretically be minimal, this allows for a direct interpretation of psychometric function threshold as the window of perceptual resolution within which two flashes can be resolved 50 % of the time, discounting guesses and lapses. Conversely, slope can be interpreted as the variability of this window. If interval bias does play a significant role, parameter estimates will be inaccurate (García-Pérez & Alcalá-Quintana, 2011; Klein, 2001). Another issue arises from perceptual learning effects over consecutive sessions, which can also be attributed to criterion shifts in some cases (Law & Gold, 2010). When psychometric curves are shifted over different sessions

it is not correct to pool data over them before fitting the psychometric function, as this will lead to artificially shallower slopes (García-Pérez & Alcalá-Quintana, 2011). However, in this thesis we neglected this pitfall simply because single session psychometric function fits did not pass goodness of fit tests in many cases due to insufficient observations within a session (i.e. a testing day).

Note that while yes-no task psychometric function threshold estimates have been previously used to quantify temporal binding windows (Cecere et al., 2015; Samaha & Postle, 2015), these estimates are not purely based on perceptual sensitivity, but also influenced by bias (Klein, 2001). This can be seen when considering the previously mentioned view of one Gaussian situated at each SOA and one dividing criterion for perceptual choice formation. By shifting the criterion, the psychometric function, too, will shift. Moreover, there is no optimal criterion placement, because the criterion cannot be placed in-between all of the Gaussians at once. This also means that depending on trial history participants are likely to adjust the criterion in the appropriate direction (this is generally true, but more pronounced when multiple SOAs are presented in random order).

In chapter 2 we will apply psychometric function and the one-dimensional signal detection model to investigate the temporal determinants of two-flash segregation in variable audio-visual contexts (0-2 beeps).

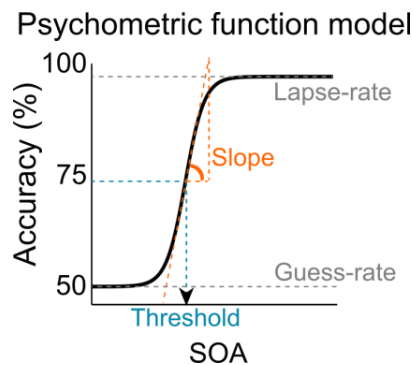


Figure 1.3. Psychometric function model with a logistic sigmoid. Toy model of a psychometric function fit to proportion correct over both signal and noise trials. Guess-rate and lapse-rate are nuisance parameters that capture the lower and upper bound of the function. Threshold and slope determine the position and steepness of the function, respectively. In this study threshold is used as a measure of the temporal window of perceptual resolution, i.e. how well participants are able to segregate one from two flashes. Slope reflects the variability of this window.

Response speed as a proxy for decision time

Another factor that is readily measured in psychophysics tasks, but not considered by the models discussed above, is response speed. A common and reasonable assumption is that the response speed roughly corresponds and is proportional to the speed of the decision, as long as participants' response window is unrestrained. There is an intuitive and long-established trade-off between response accuracy and response speed: When accuracy is valued, response time increases and vice versa (Heitz, 2014; Liu & Watanabe, 2012). Indeed, when there is a difference in response speed between experimental conditions, it is not necessarily correct to compare them in terms of signal detection model parameter estimates. For instance, if condition A is associated with higher sensitivity than condition B, but at the same time has slower response times, how are we to know whether sensitivity would be still larger in A if participants reacted as quickly as in condition B?

A potential solution to this issue is provided by sequential sampling models such as the influential drift diffusion model, which takes into account both accuracy and response speed (Ratcliff, Smith, Brown, & McKoon, 2016). In fact, it can be seen as an extension of the simple SDT model discussed above (Ratcliff, 1978; Ratcliff & McKoon, 2008). However, thus far none of these sequential sampling models has been adapted to take into account the causal inference problem, making them suitable for forced fusion settings only (Drugowitsch, DeAngelis, Klier, Angelaki, & Pouget, 2014). In this thesis we therefore opted for using the simple two equal-variance Gaussian signal detection model in line with previous investigations of the SIFI (McCormick & Mamassian, 2008).

Neural correlates of perceptual choice formation

Until now we have only examined neural responses in the context of MSI and early, largely stimulus evoked ERPs. Now, we turn to the next stage in which sensory evidence is used to decide on a particular behavioural choice. This is not meant to imply that early perceptual integration and decision formulation are entirely sequential. We very much take the view that higher order cognitive and lower level perceptual processes interact before, during and after stimulus presentation (Gau & Noppeney, 2016; Gilbert & Li, 2013; Iemi et al., 2017; Yang, 2017).

Since the discovery of neural response profiles resembling evidence accumulation akin to a decision variable in sequential evidence accumulation models in lateral intraparietal area (LIP) and frontal eye fields (FEF) of non-human primates, the field has expanded substantially (Hanks & Summerfield, 2017). According to the canonical view, a domain general decision variable is represented by neuronal firing rates, ramping up toward a fixed threshold at which point a response is initiated (Gold & Shadlen, 2002; Roitman & Shadlen, 2002). This is exactly akin to the decision variable in the standard drift diffusion model (Ratcliff, 1978; Ratcliff &

McKoon, 2008; Ratcliff et al., 2016). More recently, however, it was discovered that the neural correlates of perceptual decision making are considerably more complex (Hanks & Summerfield, 2017). For example, the classical ramping of activity toward a decision bound seems to hold only when looking at populations of neurons rather than single units (Bennur & Gold, 2011; Hanks & Summerfield, 2017; Park, Meister, Huk, & Pillow, 2014); and on a single trial neuronal responses might be more step-like rather than gradual, with slow build-up of activity only emerging when considering multiple trial iterations (Churchland et al., 2011; Ding, 2015). Another interesting discovery was that LIP neurons are not essential for decision making, as pharmacological inactivation leads to negligible performance changes (Katz, Yates, Pillow, & Huk, 2016). Finally, Rodent studies revealed recently that prefrontal neurons' firing rates demonstrate a binary link toward perceptual outcome, whereas parietal neurons code linear evidence accumulation (Hanks et al., 2015).

Remarkably, even though recordings of electromagnetic brain activity over a participant's scalp capture the synchronous activity of vast populations of neurons, accumulation-to-bound signals have also been reported in ERPs (Kelly & O'Connell, 2013; O'Connell, Dockree, & Kelly, 2012; Twomey, Murphy, Kelly, & O'Connell, 2015). The centro-parietal positivity (CPP) is a slow potential that ramps up toward a fixed threshold, with slope steepness being proportional to the sensory evidence and behavioural responses being initiated after the threshold is reached (Kelly & O'Connell, 2013; O'Connell et al., 2012). A crucial aspect of sequential sampling is that evidence is accumulated in an ongoing manner, which is clearly sensible if the stimulus to be evaluated is continuously presented. It is less clear what happens when a transient stimulus is evaluated, as is the case in the SIFI. Recently it has been demonstrated that the late component of the P300 resembles an accumulation to bound decision signal akin to the CPP (Twomey et al., 2015). The P300 is an ERP that occurs between

~300-600 ms after stimulus onset when participants are presented with infrequent, behaviourally relevant targets. The authors argue that the amplitude of the P300 scales inversely with the probability that a target signal is presented, because the start-to-bound distance is adjusted according to prior information (Smith & Ratcliff, 2004; Twomey et al., 2015). If this is true, then ERPs elicited by transient, frequent targets, should also carry information about this internal decision variable, although it might be obscured by early sensory evoked potentials. In that case multivariate pattern analysis, leveraging the information from all electrodes at once, might be able to reveal when perceptual outcomes are represented in the brain (Mostert, Kok, & de Lange, 2015).

We will address this question in chapter 3, applying multivariate pattern decoding to track and compare the neural representations of one versus two flash perceptual outcomes in the different auditory contexts (0-2 beeps), both relative to the stimulus and to response commitment.

The role of alpha frequency in visual temporal resolution

The temporal resolution of the visual system is vastly inferior compared to audition. The shortest discrete perceptual frame within which two identical pulses of light can be distinguished is dictated by the physical properties of the photon-receptors of the retina and should be approximately 34 ms (Reeves, 1996). This does not yet take into account any additional factors further down the processing hierarchy. Empirically, two flashes of light can be segregated in just under 50 ms (Hirsh & Sherrick Jr., 1961a; Kristofferson, 1967). In comparison, auditory perception resolves the temporal order of two otherwise identical events after just 2 ms delay (Hirsh & Sherrick Jr., 1961a).

Intriguingly, the most prominent brain rhythm over occipital regions – the alpha band – fluctuates at approximately 10 Hz, exactly twice the empirically measured window of temporal

integration (Berger, 1929). Unsurprisingly therefore, it was conjectured early on that perceptual cycles might be reflected in the waxing and waning of oscillatory brain activity (Kristofferson, 1967; VanRullen, 2018; White, 1963). Indeed, several studies found support in favour of a putative relationship between alpha frequency and visual temporal resolution (Cecere et al., 2015; Samaha & Postle, 2015; Varela, Toro, Roy John, & Schwartz, 1981), but contradictory and null findings were also reported (Gho & Varela, 1988; Walsh, 1952) (for a full review see (VanRullen, 2018)). As of yet, it is unclear whether these discrepancies are due to differences in experimental paradigms, stimulus parameters, task instructions, analysis methods, recording techniques or simply chance. It is also noteworthy that despite being the first brain rhythm to be discovered (and named (Berger, 1929)), the subcortical and putative cortical sources of alpha oscillations are still a matter of debate, and there are possibly multiple generators with functionally distinct roles (Freyer et al., 2011; Lopes da Silva, van Lierop, Schrijer, & Storm van Leeuwen, 1973; Naruse, Matani, Miyawaki, & Okada, 2010; Vijayan & Kopell, 2012).

If alpha frequency does index temporal cycles in vision, we should be able to replicate previous work on two-flash perceptual segregation showing a link between alpha frequency and temporal resolution (Cecere et al., 2015; Samaha & Postle, 2015). This question is addressed in chapter 4. Moreover, we will extend the perceptual context to the multisensory domain of the fusion illusion (2 flashes & 1 beep), and attempt to segregate perceptual sensitivity and response bias, which has not been done previously.

Study outline and overview of experiments

To comprehensively investigate the role of audio-visual perceptual decision making in the context of one- versus two-flash integration or segregation we conducted a single large-scale psychophysics study including three experimental designs (two interval forced choice, yes-no and yes-no at perceptual threshold) with three auditory contexts each (no beep, one beep,

two beeps; Figure 1.4). These were presented to 20 dedicated participants who completed the study over 5-9 testing days (Figure 1.5). Yes-no tasks additionally included acquisition of electroencephalographic (EEG) activity. Two-interval forced choice (2IFC) and the first yes-no task included 8 stimulus onset asynchronies for each two-event condition (e.g. ‘2 flash, 1 sound’) to allow for an estimation of individual temporal binding windows. In the 2IFC task psychometric function threshold is a good estimate of perceptual resolution, whereas in the yes-no task it is also influenced by bias (Klein, 2001). The 2IFC task is therefore crucial to get an unbiased estimate of participants’ temporal binding windows. The yes-no task is nevertheless important to measure concurrent brain activity, as in the 2IFC task EEG was not recorded due to ambiguity as to whether the first or second interval influenced the perceptual outcome. Moreover, perceptual bias is in itself an interesting factor, which can be extracted with the aforementioned two equal variance Gaussian signal detection model. At the behavioural level temporal binding windows are investigated in detail in the first empirical chapter (chapter 2), thereby considering tasks one (2IFC) and two (yes-no).

The third task (yes-no threshold) was devised to i. allow for a replication of results from the yes-no task in a very similar context (only for chapter 4) and ii. obtain a much larger number of trials at the critical SOAs at which participants responded ‘see 1’ and ‘see 2’ approximately equally often for the stimulus combinations ‘2 flash, 0 beep’ (unisensory flash fusion condition), ‘2 flash, 1 beep’ (fusion condition) and ‘1 flash, 2 beep’ (fission condition). Obtaining such a large number of trials allowed us to investigate the neural correlates of perceptual decision making prior to responding ‘see 1’ or ‘see 2’ depending on auditory context using multivariate pattern analysis in chapter 3 (thereby only considering data from the yes-no threshold task).

The role of alpha oscillation frequency as a subjective trait or variable state within individuals is investigated using data from all three tasks in chapter 4.

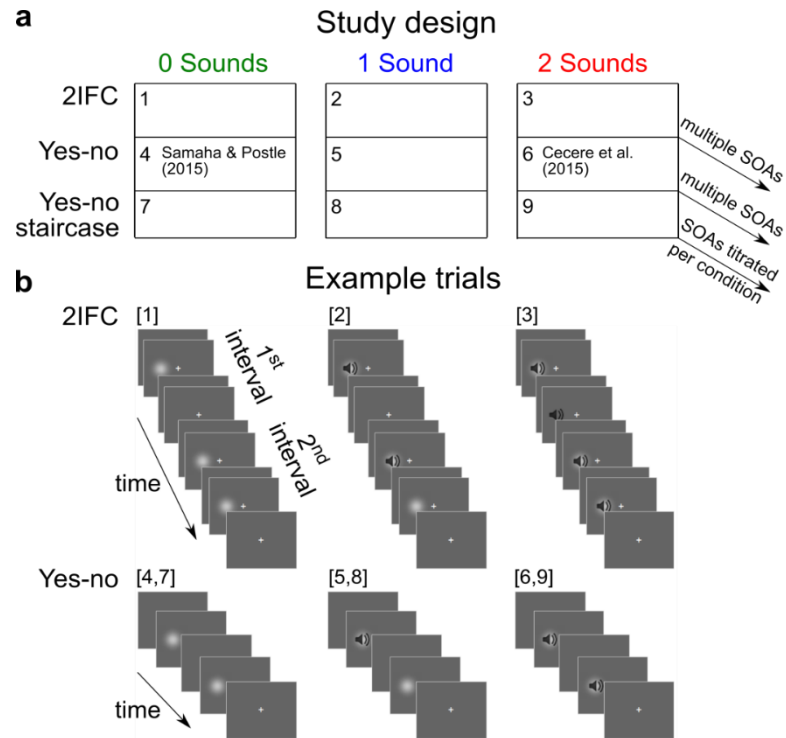


Figure 1.4. Study overview and design. *a*, Illustration of the 9 experiments with task coded by rows and number of sounds coded by columns. EEG was acquired only during yes-no tasks (experiments 4-9). Two yes-no tasks were performed, one with multiple SOAs to allow for the estimation of psychometric functions and one with a single SOA titrated to yield approximately 50% ‘see 1 flash’ and 50% ‘see 2 flash’ responses per participant for the following stimulus combinations: ‘2 flash + 0 sound’; ‘2 flash + 1 sound’; ‘1 flash + 2 sound’. *b*, Example trials for two-interval forced choice (2IFC; 1-3) and yes-no paradigms (4-9). In the 2IFC task the first or second interval contains two flashes and the other interval contains one flash. Participants reported which interval contained two flashes. In the yes-no task either one or two flashes are presented and participants reported how many flashes they perceived.

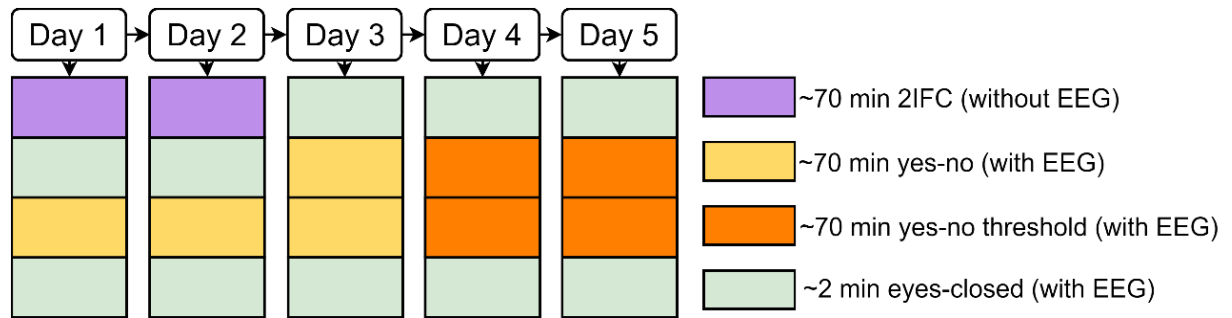


Figure 1.5. Exemplar testing schedule. A planned testing schedule consisted of 5 days with approximately 140 minutes pure task time and 4 minutes of relaxed eyes-closed EEG recordings per day. An experimental run included roughly 70 minutes of pure task time and breaks were flexibly offered in-between task blocks (~6 min task time per block). In the two-interval forced choice (2IFC) task no EEG was recorded.

Advantages and limitations of a large-scale within-subject design

When deciding on the particular study design before or during the piloting phase of a new research project scientists inevitably face the challenge of choosing between several trade-offs. For instance, when a study is short participants will be more alert, which might improve data quality, and more participants can be tested within a shorter time frame. Yet, signal to noise ratio will be inferior to a long study with more trials and participants' performance will be more variable due to learning effects. In this study we decided to test the same 20 participants for a long period of time over multiple tasks and flash-beep conditions to maximize signal to noise ratio (i.e. acquire a large number of trials per condition) and allow for comparisons between tasks within participants, thereby testing the reliability of effects over time and between slightly different task contexts (i.e. do effects replicate / generalize?). This is a particularly strong approach when testing for within subject effects, but somewhat limited for between subject population inference, as the same participant sample is tested throughout all tasks.

We chose to intermix different flash-beep conditions within blocks to minimize response bias and strategy effects that participants might develop in less balanced or diverse task designs (e.g. only presenting 1 flash, 2 beep trials would likely lead to large variability in response bias between participants). This choice does, however, make our results somewhat less comparable to previous work where only a subset of our flash-beep combinations was presented (Cecere et al., 2015; Samaha & Postle, 2015), a caveat that is particularly relevant for the interpretation of results in chapter 4 (what is the role of alpha frequency in visual temporal resolution). For example, it is possible that an intermixing of unisensory visual and audio-visual stimuli affects participants' strategy differently than a purely unisensory task, which in turn might affect perceptual or response bias. Overall, however, we believe that a balanced design with intermixed, pseud-randomized trial sequences of all flash-beep combinations is least likely to bias participants' strategy, and it should afford the best possible comparability between conditions as any change in strategy or bias is equally likely to affect any of the flash-beep combinations that are presented on the subsequent trial. Hence, we are sacrificing between study comparability to previous work for enhanced comparability between tasks and conditions in our own study.

While strategic components of perceptual decision making should in principle be captured by bias, it is possible that context effects and overall structure of an experiment influence participants' perceptual sensitivity due to fatigue and fluctuations in attention or vigilance (Luo & Maunsell, 2015; Veksler & Gunzelmann, 2018). It is therefore advisable to take into account temporal effects such as block number or day of testing, which are known to influence participants' vigilance and cognitive state (Veksler & Gunzelmann, 2018), and more recently alpha oscillation power and frequency (Benwell et al., 2019).

Mechanisms of (multi-)sensory integration captured by signal detection theory, psychometric functions and response times

We have outlined that full factorial psychophysics designs and a counterbalancing of response keys over recordings, in particular when using 2IFC tasks, should minimize response biases (e.g. always responding with the right hand). Moreover, such biases can be captured by the criterion parameter in the two equal variance Gaussian signal detection model, but the bias parameter also captures any other type of criterion shift, which can even be related to purely perceptual effects (Witt et al., 2015). Now, we want to more specifically relate different cognitive and perceptual processes to measures of perceptual sensitivity and bias, and speculate about potential differences depending on audio-visual context.

Previous research has shown that behavioural performance in multisensory contexts is frequently close to optimal such that stimuli that are likely to originate from the same source are “fused” according to their respective reliabilities, while stimuli that are unlikely to originate from the same external event are processed independently (Alais & Burr, 2004). When uncertainty about the world’s causal structure exists, Bayesian causal inference is performed and “forced fusion” and “full segregation” estimates are combined according to their respective probabilities and prior expectations (Körding et al., 2007). However, while optimality in multisensory integration is prevalent, it is not ubiquitous, and suboptimality exists in several forms. At the perceptual level, participants might put too much weight on the dominant sensory signal, which can be construed as a bias towards reporting the dominant percept, even when participants perceive stimuli to originate from separate sources ((Bertelson & Radeau, 1981; Wallace et al., 2004)). Similarly, at the causal inference level, participants might opt for suboptimal strategies of how to combine “forced fusion” and “full segregation” estimates such as model selection (choosing the estimate associated with a higher probability) or probability

matching (choosing an estimate at random proportional to estimated probabilities; (Wozny et al., 2010)). The crucial cue for causal inference manipulated in this thesis is temporal asynchrony. Hence, we would expect participants' perceptual estimates to transition from "forced fusion" at low temporal asynchronies to "full segregation" at large temporal asynchronies, while intermediate SOAs should be associated with Bayesian causal inference where both possibilities have a significant non-zero probability. How would this manifest in terms of psychometric function and signal detection model parameters and response times?

Perceptual sensitivity reflected by d' and psychometric function threshold in the 2IFC task are direct measures of how the perceptual system is able to segregate between one and two flash presentations. As such d' should increase as a function of SOA, and 2IFC threshold will serve as a measure of temporal window within which temporally separate events are likely to be perceptually integrated. Because multisensory conditions with congruent flash-beep presentations should lead to an increase in perceptual reliability, it might not necessarily be expected that multisensory TBWs will differ from unisensory visual TBWs. However, we might expect to see an inverted U-shaped curve of response times in the 2-flash, 2-flash & 1-beep and 1-flash & 2-beep conditions, indicating more computationally intensive model averaging strategies when the causal structure is uncertain (Wozny et al., 2010).

Instead, estimates of bias (C) will likely capture response biases toward the number of perceived auditory stimuli, which has been observed in both fission and fusion illusion contexts when SOA was fixed (Chen et al., 2017; McCormick & Mamassian, 2008). Such biases are unlikely to be affected by temporal asynchrony if trials with different SOAs are randomly intermixed. However, in the context of multiple SOAs decision criterion is very likely to change substantially depending on trial history. While these effects will probably average out over the course of the experiment, they will reduce signal to noise ratio. If participants adopt a

suboptimal strategy for combining forced fusion and full segregation estimates when the causal structure is uncertain, this should manifest as a shift in bias for intermediate SOAs (e.g. for model selection at early SOAs forced fusion is more likely leading to a bias toward the auditory stimulus, whereas at large SOAs full segregation is more likely leading to a bias toward the visual stimulus – which would manifest as no bias in the signal detection model). However, the only condition in which criterion shifts can be estimated as a function of SOA is in the yes-no 2-beep condition and fluctuations in bias might also simply be due to perceptual outcome.

Synopsis of research questions

The main question we aimed to address in this thesis is to what extent one versus two flash temporal discrimination is influenced by the concurrent presentation of auditory beeps, in particular when flash-beep pairings are mismatched, leading to the sound induced fusion illusion (reporting a single flash when two flashes and one beep are presented) and fission illusion (reporting two flashes when one flash and two beeps are presented). Specifically, we were interested to what extent perceptual sensitivity and bias manifest as a function of stimulus onset asynchrony depending on auditory context (chapter 2); similarities or differences in EEG patterns associated with perceptual decisions (chapter 3); and whether the long-standing alpha temporal resolution hypothesis of visual perception is i. related to perceptual sensitivity rather than bias and ii. extendable over unisensory and multisensory contexts.

Chapter 2 - Temporal determinants of the sound induced fission and fusion illusion

The following chapter is being prepared for scientific publication and is the result of collaboration between Uta Noppeney and Steffen Bürgers. U.N. conceived the study. S.B. and U.N. designed the experiments. S.B. collected the data. S.B. analysed the data. S.B. and U.N. interpreted the data. S.B. wrote the manuscript. U.N. provided direct supervision at all stages of the process.

Abstract

In the sound induced flash illusion, the number of presented flashes is perceived differently depending on the number of concurrently presented beeps. This multisensory integration depends on the temporal asynchrony between the input stimuli. However, little is known about how the fission illusion (reporting an illusory second flash) and fusion illusion (reporting one flash when two are presented) are affected by temporal asynchrony irrespective of (response) bias. Therefore, we examined one versus two flash perceptual outcomes as a function of stimulus onset asynchrony accompanied by 0-2 beeps in both two-interval forced choice and yes-no tasks. Psychometric function slope and d' consistently showed a larger temporal integration window for fission compared to 2-flash fusion and the fusion illusion. At the same time, fusion was associated with higher sensitivity and faster responses relative to the 0-beep condition. This effect was driven by performance benefits on 1-flash, 1-beep trials compared to 1-flash, 0-beep trials. In addition, both fission and fusion illusion were associated with bias toward reporting the presented number of beeps. These findings suggest a larger temporal binding window for fission than unisensory flash fusion or the fusion illusion, and highlight the importance of dissociating perceptual sensitivity and bias.

Introduction

We perceive our environment in an inherently multisensory fashion with each modality providing information about the external world. When it can be assumed that the same external event gives rise to input in multiple sensory modalities (e.g. vision and audition), the best internal representation is formed by integrating the individual sensory signals weighted by their respective reliabilities. Conversely, when it can be assumed that individual sensory signals come from separate outside events they should be processed independently (Körding et al., 2007). The brain uses prior experience as well as the sensory inputs themselves to infer whether a common or separate outside sources stimulate the senses (Körding et al., 2007; Rohe & Noppeney, 2015b; Shams, 2012).

A strong cue for a common source is temporal coincidence, as evidenced by temporal tuning functions of multisensory neurons in the cat superior colliculus (Meredith et al., 1987; Meredith & Stein, 1986) and temporal binding windows (TBW) relating physical asynchrony to perceptual response probabilities (Stevenson, Fister, Barnett, Nidiffer, & Wallace, 2012). A mistuning of unisensory or multisensory TBWs is likely to lead to wrong inferences about the world and has been implicated in developmental disorders such as dyslexia, autism and schizophrenia (Wallace & Stevenson, 2014).

A relatively simple, but powerful way to study audio-visual multisensory perception is via the sound induced flash illusions (SIFIs) (Shams et al., 2000, 2002). In the SIFIs the number of perceived flashes is altered by concurrently presented mismatched numbers of acoustic stimuli illustrating the relatively higher temporal resolution (i.e. higher temporal reliability) of audition compared to vision (Shams et al., 2000, 2002, 2005). Specifically, in the sound induced fission illusion (or double flash illusion) an illusory second flash is reported when a single flash and two beeps are presented. Conversely, in the sound induced fusion illusion a single flash is

reported when two flashes are presented together with a single beep (both illusions also occur for larger disparate numbers of flashes and beeps (Shams et al., 2002)).

Following the principle of temporal coincidence, the likelihood of either the fission or fusion illusion diminishes with increasing asynchrony between consecutive stimuli. The fission illusion TBW has been variably reported to be within 70 - 300 ms, but varies considerably between participants and studies (Apthorp et al., 2013; Bidelman, 2016; Bidelman & Heath, 2018; Cecere et al., 2015; Foss-Feig et al., 2010; Hamilton, Wiener, Drebing, & Coslett, 2013; Narinesingh, Goltz, & Wong, 2017; Shams et al., 2000, 2002) and some proportion of 2-flash reports remains even after 300 ms (Bidelman & Heath, 2018; Foss-Feig et al., 2010). The fusion illusion has been reported to abate at approximately 70 -100 ms (Apthorp et al., 2013; Narinesingh et al., 2017). Interestingly, only two studies directly compared fission and fusion illusion TBWs and found no significant difference between them for young adult participants (Apthorp et al., 2013; McGovern et al., 2014), but a larger fission window for older adults (McGovern et al., 2014).

An important caveat in using perceptual reports to quantify perceptual binding is that they are impacted not only by perceptual sensitivity, but also perceptual and decisional bias (Witt et al., 2015). Signal detection theory segregates these factors in the form of d' (sensitivity) and criterion C (bias) (Macmillan & Creelman, 2005). However, studies using signal detection measures as a function of SOA have not directly compared fission and fusion conditions (Bidelman & Heath, 2018; McGovern et al., 2014). When deploying a fixed SOA, the fission illusion (1 flash, 2 beeps versus 2 flashes, 2 beeps) has been associated with both decreased perceptual sensitivity compared to a unisensory context (Chen et al., 2017; Kaposvári et al., 2014; McCormick & Mamassian, 2008; Vanes et al., 2016; Watkins et al., 2006) and bias toward 2-flash reports (Chen et al., 2017; Kaposvári et al., 2014; McCormick & Mamassian,

2008; Vanes et al., 2016). Conversely, the fusion illusion (2 flashes, 1 beep versus 1 flash, 1 beep) has been associated with decreased perceptual sensitivity (Chen et al., 2017; Kaposvári et al., 2014; Mishra et al., 2008; Vanes et al., 2016) and a shift in bias toward reporting a single flash (Chen et al., 2017; Kaposvári et al., 2014; Vanes et al., 2016).

When measuring perceptual performance at multiple SOAs it is possible to fit psychometric functions that take into account all stimulus levels (SOAs) at once, summarizing the data with parameter estimates of threshold (inflection point) (Cecere et al., 2015; Samaha & Postle, 2015) and slope (steepness of the function). In a yes-no context, researchers have previously pooled over 1-flash and 2-flash presentations and fit psychometric curves to performance accuracy, using the threshold as a measure of temporal resolution (Samaha & Postle, 2015). However, threshold estimates are also affected by bias (Macmillan & Creelman, 2005). To circumvent bias influencing threshold estimates, two-interval forced choice (2IFC) paradigms can be used, which lead to theoretically unbiased responses.

Lastly, it is important to account for response time differences between conditions due to the speed-accuracy trade-off (Liu & Watanabe, 2012), which is neither considered by psychometric function nor simple two equal-variance Gaussian signal detection models.

In this study, we aimed to comprehensively characterize the temporal determinants of 1-flash versus 2-flash perceptual segregation accompanied by 0-2 beeps using model based (psychometric function, two equal-variance Gaussian signal detection) and model free (hit rate, false alarm rate, response time) analyses.

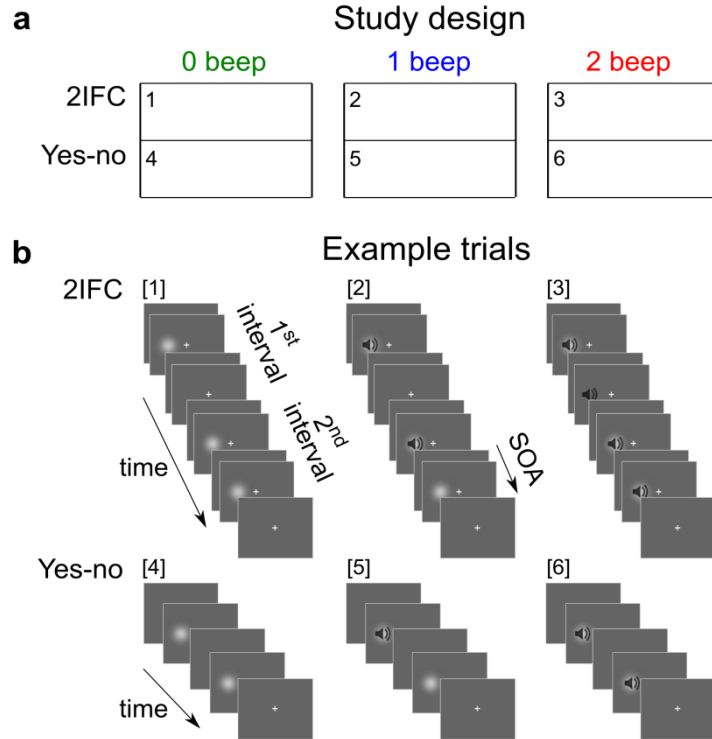


Figure 2.1. Study design and trial examples. *a*, Illustration of the 6 experiments included in this study with paradigm coded by rows and number of beeps coded by columns. *b*, Example trials for 2IFC (exp. 1-3) and yes-no tasks (exp. 4-6). In the 2IFC task the first or second interval contained 2 flashes and the other interval contained 1 flash (examples only show 2 flashes in second interval). Participants reported which interval contained 2 flashes. In the yes-no task either one or two flashes were presented and participants reported how many flashes they perceived (only 2-flash examples shown).

Results

To scrutinize the temporal determinants of 1-flash versus 2-flash perceptual segregation in a unisensory and audio-visual context, we conducted a psychophysics study consisting of 2 (task: 2IFC, yes-no) x 3 (number of beeps: 0-2) experiments (Figure 2.1). Twenty adult human participants completed the study.

Psychometric function model and temporal resolution

First, we characterized perceptual performance accuracy over both signal (2-flash) and noise (1-flash) trials by fitting psychometric functions (Figure 2.2a). Results are depicted in Figure 2.3a.

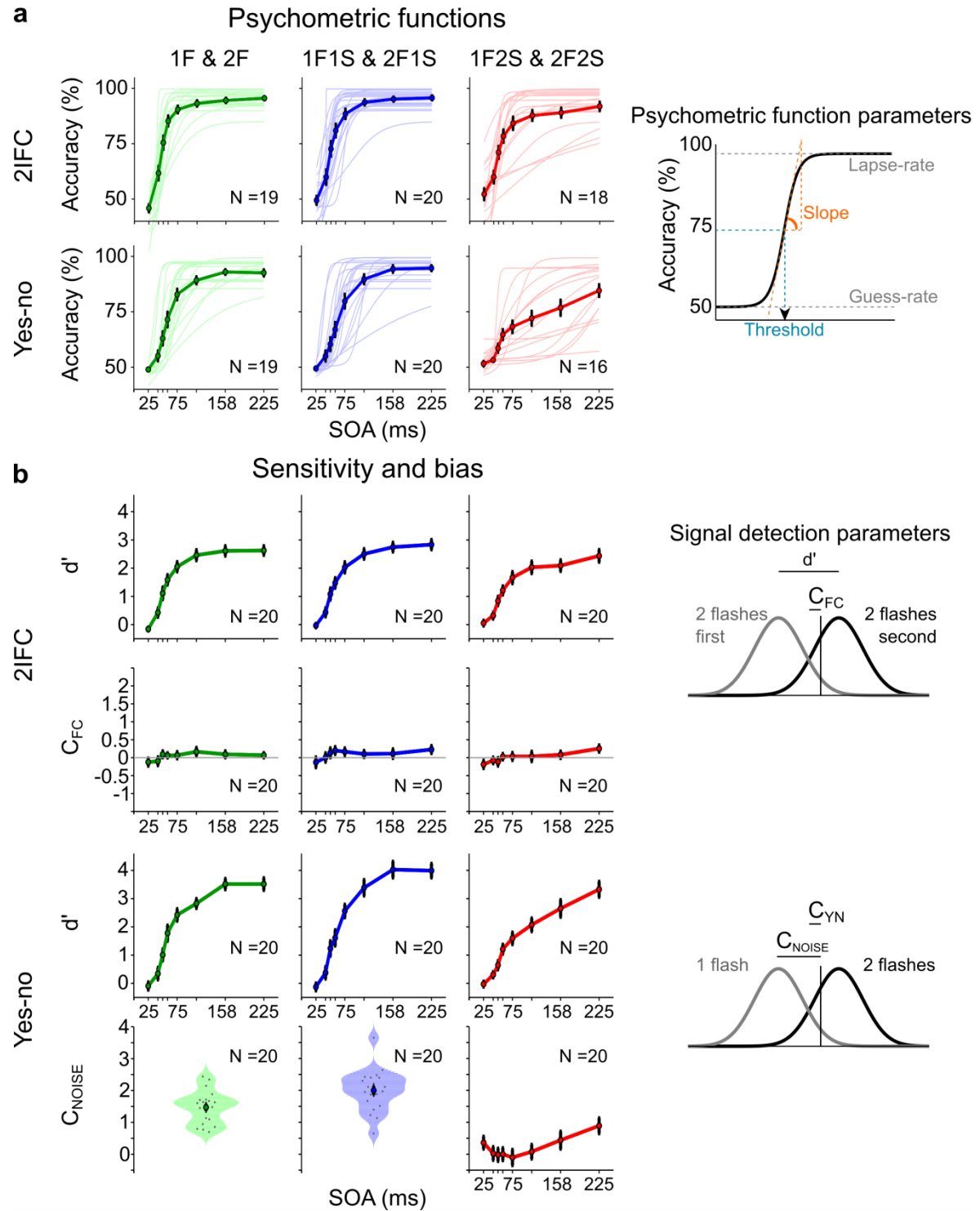


Figure 2.2. Perceptual performance over time and model illustrations. *a*, Grand average response accuracies (connected with bold lines) and single subject psychometric function fits (thin lines) for 2IFC (exp. 1-3, row 1) and yes-no (exp. 4-6, row 2) tasks. Each yes-no figure depicts data from 1-flash and 2-flash trials. Columns denote the number of concurrently presented beeps (0-2). In the 2IFC task

psychometric function inflection points (threshold parameters) are a measure of perceptual resolution, with slope denoting its variance, as interval bias is largely absent. In the yes-no task psychometric function parameters are influenced by both perceptual sensitivity and bias. Missing data in rows 1 and 2 are due to poor psychometric curve fits. **b**, Two equal-variance Gaussian signal detection model parameters (d' , C_{FC} and C_{NOISE} ; model illustrated in right panel) as a function of SOA for the 2IFC task (exp. 1-3, row 1-2) and yes-no task (exp. 4-6, row 3-4). Note that C_{FC} indicates close to unbiased responses in the 2IFC task over SOAs. In the yes-no task C_{NOISE} captures the criterion of reporting a signal (2 flashes) when there is none (1 flash). It does not take into account hits (report “see 2” when 2 flashes are presented). Note that a difference in C_{NOISE} between conditions indicates that psychometric function threshold differences between conditions are partly due to bias. Error bars denote ± 1 SEM.

2IFC task. If participants have no bias (C_{FC}) toward either interval, 2IFC threshold and slope estimates are driven solely by perceptual sensitivity (García-Pérez & Alcalá-Quintana, 2011; Klein, 2001). To test this assumption, a two-way repeated measures (RM) ANOVA with factors condition (3 levels: 0-2 beeps) x SOA (8 levels: 25, 42, 50, 58, 75, 108, 158, 225 ms) was performed (see Figure 2.2b, row 2; note that participant 1 was excluded from the analysis, because data at SOA 158 ms was not acquired). There was a significant main effect of SOA ($F_{(3.60, 64.73)} = 2.90$; $p = 0.03$; $\eta_p = 0.14$), indicating that psychometric function threshold is influenced by interval bias. Post-hoc tests revealed that participants tend to favour the second over the first interval for 25 ms compared to 50 - 75 ms SOAs (50 ms: $t_{(19)} = -4.72$; $p = 0.004$; $d = -0.54$; 58 ms: $t_{(19)} = -4.20$; $p = 0.013$; $d = -0.80$; 75 ms: $t_{(19)} = -3.60$; $p = 0.049$; $d = -0.69$). However, C_{FC} did not significantly differ from zero ($F_{(1,18)} = 1.31$; $p = 0.27$; $\eta_p = 0.07$) and there was no main effect of condition ($F_{(2, 36)} = 0.65$; $p = 0.53$; $\eta_p = 0.04$) or condition by SOA interaction ($F_{(5.28, 94.94)} = 0.68$; $p = 0.65$; $\eta_p = 0.04$). It is therefore unlikely that bias plays a significant role when comparing threshold and slope estimates between conditions.

For 2IFC threshold, a one-way RM ANOVA revealed a significant main effect of condition ($F_{(1.29, 21.95)} = 4.88$; $p = 0.03$; $\eta_p = 0.22$). Post hoc tests showed a marginally significant positive difference between 1-beep and 2-beep presentations ($t_{(17)} = 2.60$; $p = 0.056$; $d = 0.79$). As normality of residuals in the 2-beep condition was in doubt, a complementary Wilcoxon sign-rank test was performed similarly supporting a significantly larger threshold for 1-beep compared to 2-beep presentations ($z = 2.68$, $p = 0.021$; $r = 0.45$).

Regarding 2IFC slope (log10-transformed), a one-way RM ANOVA identified a significant condition effect ($F_{(1.52, 25.78)} = 7.70$; $p = 0.005$; $\eta_p = 0.31$), driven by steeper slopes in 0-beep versus 2-beep ($t_{(17)} = 3.50$; $p = 0.008$; $d = 0.96$), and 1-beep versus 2-beep contexts ($t_{(17)} = 2.94$; $p = 0.018$; $d = 0.92$; sign-rank test: $z = 3.07$, $p = 0.006$; $r = 0.52$).

These results demonstrate that the temporal window of perceptual segregation of 1 versus 2 flashes when accompanied by 2 beeps is significantly different from the 0 and 1-beep contexts. Unfortunately, it is not possible to interpret all threshold estimates as a direct measure of perceptual resolution, because several 2-beep threshold values are negative. This does not mean that perceptual resolution is negative, but instead suggests that the mapping of SOA to perceptual performance is not sigmoid shaped for those particular participants and that performance at 25 ms SOA already exceeds the expected chance level.

A shallower slope in the 2-beep compared to 0-beep and 1-beep conditions indicates that participants are susceptible to the fission illusion over a longer time period compared to two-flash fusion or the fusion illusion.

Yes-no task. In the yes-no task threshold did not significantly differ between conditions ($F_{(1.043, 14.60)} = 0.44$; $p = 0.52$; $\eta_p = 0.03$). However, threshold estimates are difficult to interpret as decision criteria (C_{NOISE}) are set differently between beep conditions ($F_{(1.56, 29.67)} = 94.34$; $p < 0.001$; $\eta_p = 0.83$).

For yes-no slope (log10-transformed), condition did show a significant effect ($F_{(1.27, 17.81)} = 22.19$; $p < 0.001$; $\eta_p = 0.61$). Post hoc tests ascribed it to steeper slopes in 0-beep versus 2-beep conditions ($t_{(14)} = 4.41$; $p = 0.002$; $d = 1.15$), and 1-beep versus 2-beep conditions ($t_{(15)} = 5.09$; $p < 0.001$; $d = 1.33$). These results replicate the relatively longer window of integration in the 2-beep condition compared to 0 and 1-beep conditions found in the 2IFC task.

Inter-individual variability. Considering individual psychometric function model estimates, it is evident that inter-individual variability is frequently quite large, in particular for threshold estimates in the 2-beep condition and slope estimates in the 1-beep and 2-beep conditions (Figure 2.2a and Figure 2.3a). As mentioned, yes-no task psychometric function threshold and slope variability can be due to fluctuations in sensitivity or bias and are therefore difficult to interpret, whereas in the 2IFC task they are almost exclusively driven by sensitivity. However, large lapse-rates (which are notoriously difficult to estimate and should in principle have been minimal as participants were instructed to correct lapses, in which case trials were excluded from analysis), threshold estimates outside the scope of available data (< 0.025 s) and very steep slopes, which essentially make the psychometric function model transition from the lower bound to the upper bound within the interval between two SOAs, are problematic and have to be interpreted with caution (Wichmann & Hill, 2001; Wichmann & Jäkel, 2018). Nevertheless, none of these concerns exists for the following two-equal variance Gaussian signal detection model analyses, which should identify any spurious PF results.

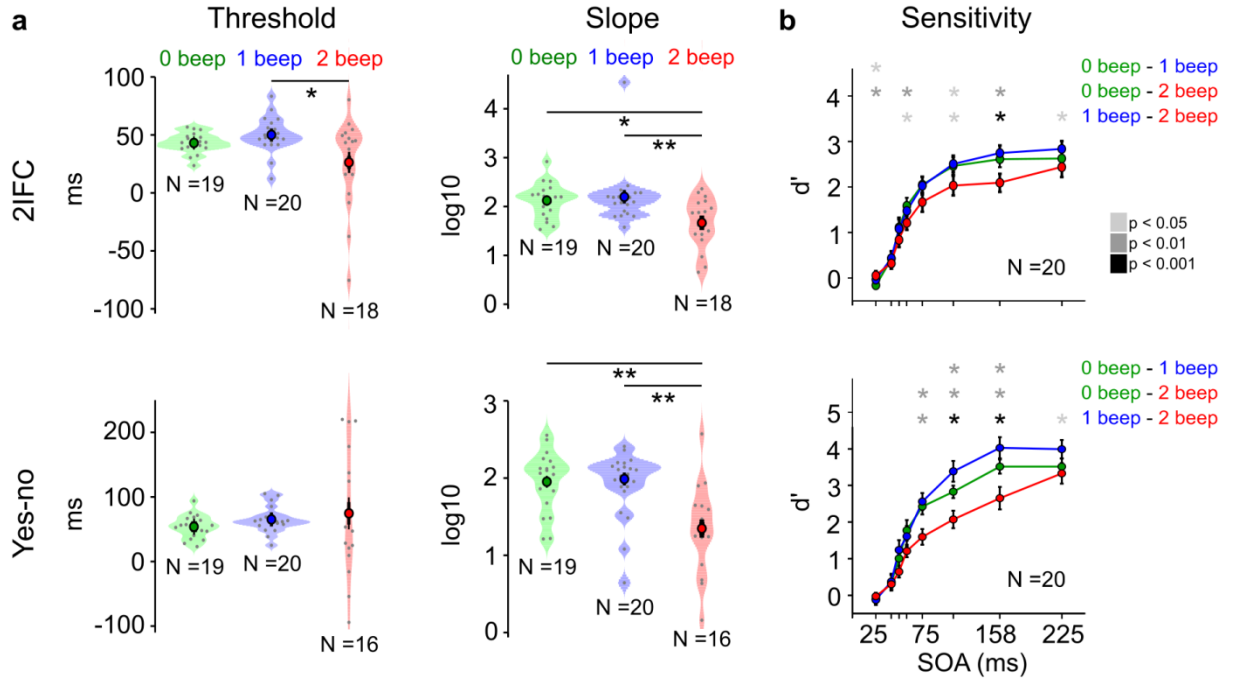


Figure 2.3. Model based perceptual discriminability analysis of one versus two flashes. *a*, Psychometric function threshold and slope estimate comparison between conditions (0-2 beeps) for the 2IFC task (exp. 1-3, row 1) and yes-no task (exp. 4-6, row 2). *b*, Sensitivity (d') for each condition as a function of SOA for the 2IFC task (row 1) and yes-no task (row 2). Error bars denote ± 1 SEM. Asterisks mark significant differences. In *a*: $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***).

Signal detection model: d'

Complementary to the psychometric function analysis, we conducted a simple two equal-variance Gaussian signal detection analysis separately for each SOA. Results are depicted in Figure 2.3b.

2IFC task. For 2IFC d' , a two-way RM ANOVA identified significant main effects of condition ($F_{(1.24, 22.36)} = 17.40$; $p < 0.001$; $\eta_p = 0.49$) and SOA ($F_{(2.40, 43.17)} = 105.47$; $p < 0.001$; $\eta_p = 0.85$). Additionally, there was a significant condition x SOA interaction ($F_{(6.27, 112.78)} = 3.79$; $p = 0.002$; $\eta_p = 0.17$). Post hoc contrasts revealed significantly lower sensitivity at 25 ms for 0-beep compared to 1-beep ($t_{(19)} = -3.45$; $p = 0.022$; $d = -0.41$) and 2-beep contexts ($t_{(19)} = -3.74$; $p = 0.008$; $d = -0.58$). Instead, for larger SOAs the 0-beep condition showed higher sensitivity

compared to 2-beep presentations (58 ms: $t_{(19)} = 4.37$; $p = 0.003$; $d = 0.50$; 108 ms: $t_{(19)} = 2.97$; $p = 0.039$; $d = 0.47$; 158 ms: $t_{(18)} = 3.95$; $p = 0.0065$; $d = 0.61$). Finally, the 1-beep condition showed consistently larger sensitivity than the 2-beep condition between 58 - 225 ms (58 ms: $t_{(19)} = 3.13$; $p = 0.028$; $d = 0.35$; 108 ms: $t_{(19)} = 3.61$; $p = 0.013$; $d = 0.53$; 158 ms: $t_{(18)} = 6.09$; $p < 0.001$; $d = 0.79$; 225 ms: $t_{(18)} = 3.50$; $p = 0.014$; $d = 0.45$).

Yes-no task. Similar to the 2IFC results, for yes-no d' , a two-way RM ANOVA revealed significant main effects of condition ($F_{(1.29, 23.26)} = 16.39$; $p < 0.001$; $\eta_p = 0.48$), SOA ($F_{(2.61, 47.01)} = 124.28$; $p < 0.001$; $\eta_p = 0.87$), and a significant interaction between them ($F_{(4.78, 86.07)} = 9.72$; $p < 0.001$; $\eta_p = 0.35$). Post hoc tests revealed a significantly lower sensitivity in 0-beep compared to 1-beep trials for 108 and 158 ms (108 ms: $t_{(19)} = -4.36$; $p = 0.003$; $d = -0.54$; 158 ms: $t_{(18)} = -4.10$; $p = 0.004$; $d = -0.48$). Moreover, d' was larger on 0-beep compared to 2-beep trials between 75 – 158 ms (75 ms: $t_{(19)} = 3.88$; $p = 0.007$; $d = 0.82$; 108 ms: $t_{(19)} = 4.07$; $p = 0.005$; $d = 0.83$; 158 ms: $t_{(19)} = 3.86$; $p = 0.006$; $d = 0.70$). In addition, d' was larger on 1-beep compared to 2-beep trials between 75 – 225 ms (75 ms: $t_{(19)} = 3.87$; $p = 0.006$; $d = 0.86$; 108 ms: $t_{(19)} = 7.35$; $p < 0.001$; $d = 1.17$; 158 ms: $t_{(19)} = 6.07$; $p < 0.001$; $d = 1.03$; 225 ms: $t_{(19)} = 2.99$; $p = 0.038$; $d = 0.51$).

In line with psychometric function results, both 2IFC and yes-no d' indicate higher sensitivity for 0-beep and 1-beep compared to 2-beep presentations, in particular for SOAs > 50 ms. The negative absolute and relatively lower sensitivity for 0-beep compared to 1-beep and 2-beep presentations in the 2IFC task at 25 ms can potentially be explained by visual summation effects. Specifically, Bloch's law states that a brief flash at perceptual luminance threshold is perceptually identical (i.e. a metamer) to a longer flash with half its luminance within a critical time period of up to 100 ms (Gorea, 2015). As stimulus luminance, contrast, or size are increased this critical duration shortens. Therefore, participants might be more likely

to respond “see 1” on 2-flash trials with very short SOAs compared to 1-flash trials, given that they have a relatively stronger flash percept (i.e. the two flash stimuli are temporally summed). In addition, a single beep can increase the salience of a flash (1-flash, 1-beep) or double flash (2-flash, 1-beep) (Noesselt et al., 2010). In doing so the perceived luminance increase due to temporal summation of visual signals alone might be attenuated. Alternatively, because both signal (2-flash) and noise (1-flash) are “boosted” by a concurrent beep, the difference in perceived luminance due to temporal summation of visual signals on 2-flash trials might be relatively less noticeable when concurrent sounds are presented. In the 2-flash, 2-beep condition the second beep should additionally facilitate segregation of the two flashes.

Lower sensitivity for 0-beep compared to 1-beep trials at large SOAs (108, 158 ms) seems surprising, considering that the fusion illusion by definition leads to lower than unisensory two-flash discriminability. However, this is only true for the 2-flash, 1-beep presentations, whereas 1-flash, 1-beep presentations should actually lead to improved perceptual reliability compared to 1-flash, 0-beep presentations (Noesselt et al., 2010). To disentangle multisensory perceptual effects on different flash-beep combinations we next turn to analyse performance accuracy (hits and false alarms) and response times.

Model free analysis

Here, we asked two questions: First, are differences in sensitivity between conditions due to differences in hit rate (2-flash trials) or false alarm rate (1-flash trials)? Second, how do response times relate to accuracy/sensitivity differences between conditions? The latter question is important to assess whether performance improvements based on sensitivity or accuracy might be confounded by response time increases. Results are illustrated in Figure 2.4.

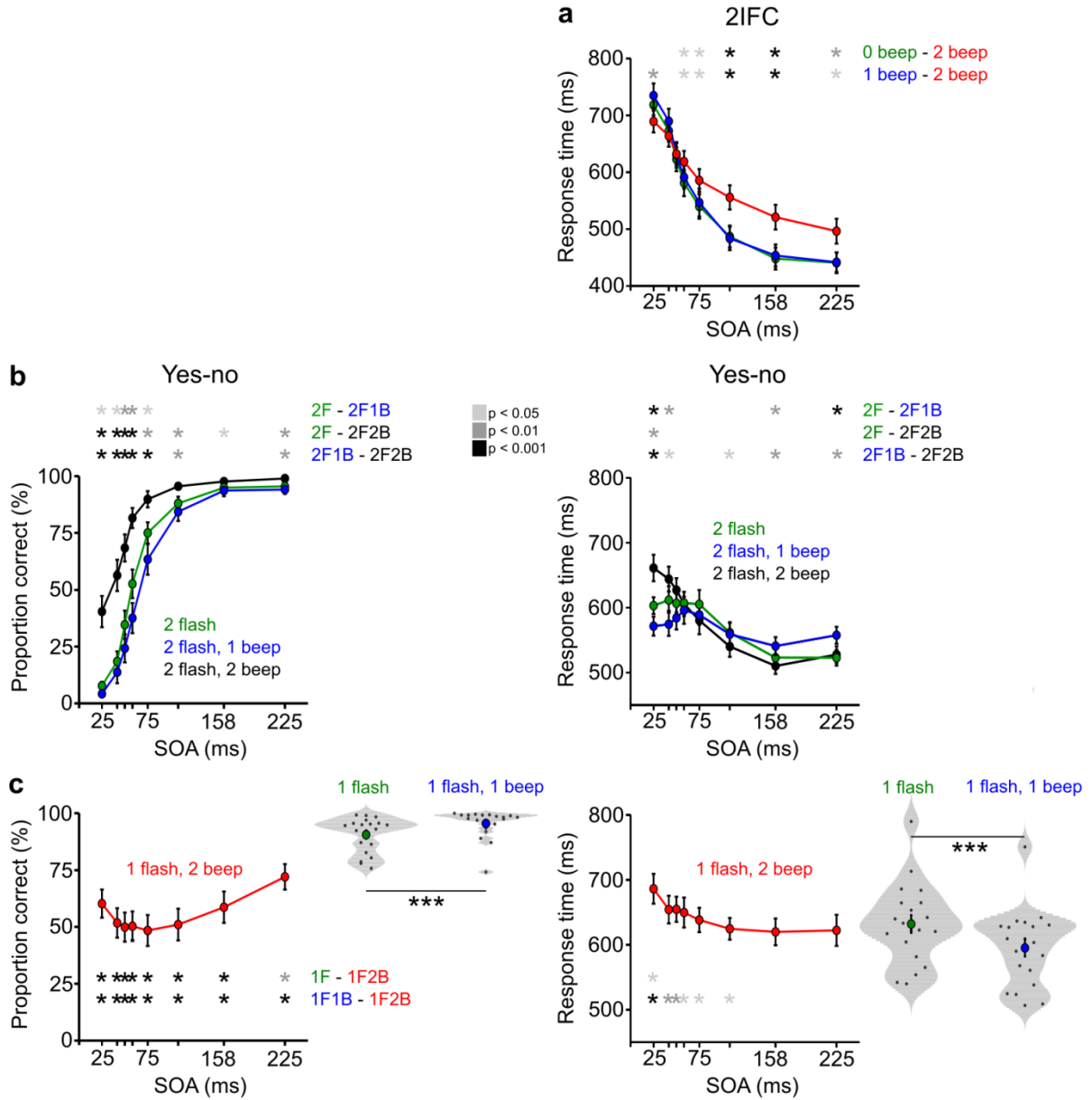


Figure 2.4. Model free perceptual discriminability analysis of one versus two flashes. *a*, Response times for 0, 1 and 2-beep conditions in the 2IFC task. *b*, *c*, Proportion correct (first column) and response times (second column) are shown for hit trials (*b*, 2 flashes presented) and false alarm trials (*c*, 1 flash presented) in the yes-no task. Error bars denote ± 1 SEM. Asterisks mark significant pairwise differences between conditions. Significance level is coded by shading or number of asterisks: $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***). 2F, 2-flash; 2F1B, 2-flash, 1-beep; 2F2B, 2-flash, 2-beep; 1F, 1-flash, 1F2B, 1-flash, 2-beep; 1F1B, 1-flash, 1-beep.

Hits and false alarms. For signal trials (2-flash) pair-wise Wilcoxon sign-rank comparisons were performed between conditions (0-2 beeps) at each SOA (25-225 ms; Figure 2.4b, left panel). Participants had more hits on 0-beep compared to 1-beep presentations (25 ms: $z = 2.58$, $p = 0.04$; $r = 0.41$; 42 ms: $z = 2.65$, $p = 0.04$; $r = 0.42$; 50 ms: $z = 3.32$, $p = 0.006$; $r = 0.53$; 58 ms: $z = 3.73$, $p = 0.002$; $r = 0.59$; 75 ms: $z = 2.95$, $p = 0.02$; $r = 0.47$). This effect is driven by bias rather than perceptual sensitivity (Figure 2.3b). Indeed, C_{YN} was consistently lower on 0-beep compared to 1-beep trials indicating a bias toward “see 2” reports ($F_{(1.47, 27.94)} = 94.80$; $p < 0.001$; $\eta_p = 0.83$). Conversely, performance on noise trials (1-flash) was worse for 0-beep compared to 1-beep presentations ($z = -3.58$, $p < 0.001$; $r = -0.80$). This demonstrates that increased sensitivity (Figure 2.3b) in the 1-beep compared to 0-beep condition is due to fewer false alarms rather than more hits.

Congruent 2-flash, 2-beep presentations showed improved performance compared to both 0-beep (25 ms: $z = 3.81$, $p < 0.001$; $r = 0.60$; 42 ms: $z = 3.92$, $p < 0.001$; $r = 0.62$; 50 ms: $z = 3.85$, $p < 0.001$; $r = 0.61$; 58 ms: $z = 3.82$, $p < 0.001$; $r = 0.60$; 75 ms: $z = 3.62$, $p = 0.001$; $r = 0.57$; 108 ms: $z = 3.44$, $p = 0.002$; $r = 0.54$; 158 ms: $z = 1.98$, $p = 0.049$; $r = 0.32$; 225 ms: $z = 2.95$, $p = 0.006$; $r = 0.47$) and 1-beep presentations (25 ms: $z = 3.78$, $p < 0.001$; $r = 0.60$; 42 ms: $z = 3.92$, $p < 0.001$; $r = 0.62$; 50 ms: $z = 3.92$, $p < 0.001$; $r = 0.62$; 58 ms: $z = 3.92$, $p < 0.001$; $r = 0.62$; 75 ms: $z = 3.80$, $p = 0.001$; $r = 0.60$; 108 ms: $z = 3.25$, $p = 0.004$; $r = 0.51$; 225 ms: $z = 3.10$, $p = 0.004$; $r = 0.49$). This is largely explained by a strong bias toward 2-flash reports as indicated by more false alarms (and lower criterion C_{YN}) on 1-flash, 2-beep trials (Figure 2.2b, Figure 2.4c). In addition, participants were more accurate (more correct rejections) in the 0-beep compared to the 2-beep condition (25 to 108 ms: $z = 3.92$, $p < 0.001$; $r = 0.88$; 158 ms: $z = 3.77$, $p < 0.001$; $r = 0.84$; 225 ms: $z = 3.10$, $p = 0.002$; $r = 0.69$). The same was true for 1-beep versus 2-beep presentations (25 to 108 ms: $z = 3.92$, $p < 0.001$; $r = 0.88$;

158 ms: $z = 3.78$, $p < 0.001$; $r = 0.85$; 225 ms: $z = 3.78$, $p < 0.001$; $r = 0.85$). Hence, enhanced perceptual discriminability for 0 and 1-beep versus 2-beep presentations is due to fewer false alarms rather than more hits.

Response times (2IFC). 2IFC response times are depicted in Figure 2.4c and results are largely in line with the sensitivity analysis (Figure 2.3b). A two-way RM ANOVA on response time with factors condition (0-2 beeps) and SOA (25-225 ms) identified a significant main effect of condition ($F_{(2,38)} = 18.84$; $p < 0.001$; $\eta_p = 0.50$), SOA ($F_{(2.26, 42.98)} = 62.33$; $p < 0.001$; $\eta_p = 0.77$), and a significant condition x SOA interaction ($F_{(4.59, 87.27)} = 9.91$; $p < 0.001$; $\eta_p = 0.34$). Post hoc tests revealed significantly faster 2-beep responses at short SOAs compared to 1-beep presentations (25 ms: $t_{(19)} = -3.70$; $p = 0.009$; $d = 0.50$). Instead, for intermediate and long SOAs participants responded slower on 2-beep compared to 0-beep (58 ms: $t_{(19)} = 2.79$; $p = 0.047$; $d = 0.40$; 75 ms: $t_{(19)} = 4.60$; $p = 0.001$; $d = 0.76$; 108 ms: $t_{(19)} = 7.78$; $p < 0.001$; $d = 0.76$; 158 ms: $t_{(18)} = 6.45$; $p < 0.001$; $d = 0.79$; 225 ms: $t_{(19)} = 3.97$; $p < 0.004$; $d = 0.62$) and 1-beep conditions (58 ms: $t_{(19)} = 2.72$; $p = 0.041$; $d = 0.32$; 75 ms: $t_{(19)} = 3.26$; $p = 0.017$; $d = 0.39$; 108 ms: $t_{(19)} = 8.52$; $p < 0.001$; $d = 0.77$; 158 ms: $t_{(18)} = 6.23$; $p < 0.001$; $d = 0.73$; 225 ms: $t_{(19)} = 3.39$; $p < 0.015$; $d = 0.62$).

Response times (yes-no). For signal trials (2-flash), a two-way RM ANOVA identified a significant main effect of SOA ($F_{(2.08, 39.52)} = 19.60$; $p < 0.001$; $\eta_p = 0.51$) and a significant condition by SOA interaction ($F_{(4.44, 84.37)} = 9.16$; $p < 0.001$; $\eta_p = 0.33$). Post hoc tests highlighted significantly slower RTs for 0-beep compared to 1-beep presentations for short SOAs (25 ms: $t_{(19)} = 4.79$; $p < 0.001$; $d = 0.51$; 42 ms: $t_{(19)} = 4.13$; $p = 0.003$; $d = 0.48$), and faster 0-beep RTs for large SOAs (158 ms: $t_{(19)} = -3.92$; $p = 0.005$; $d = -0.29$; 225 ms: $t_{(19)} = -6.27$; $p < 0.001$; $d = -0.62$). On congruent 2-flash, 2-beep trials participants responded slower on short SOAs compared to both 0-beep (25 ms: $t_{(19)} = 3.79$; $p = 0.01$; $d = 0.76$) and 1-beep

presentations (25 ms: $t_{(19)} = 5.21$; $p < 0.001$; $d = 1.13$; 42 ms: $t_{(19)} = 3.38$; $p = 0.016$; $d = 0.84$). For 1-beep trials this pattern reversed at long SOAs and participants reacted faster on congruent trials (108 ms: $t_{(19)} = 2.81$; $p = 0.045$; $d = 0.25$; 158 ms: $t_{(19)} = 4.67$; $p = 0.001$; $d = 0.53$; 225 ms: $t_{(19)} = 3.75$; $p = 0.008$; $d = 0.52$).

Slower response times on 2-flash, 2-beep trials at short SOAs could be due to additional processing time for the second beep, which is not present on the other signal trials. Slower response times on 0-beep compared to 1-beep trials at short SOAs might be due to temporal summation effects, which are more pronounced in the 0-beep condition. Alternatively, it could be explained by a simple salience or alerting effect of the sound (Noesselt et al., 2010). At larger SOAs response time effects converge with hit rate results.

On noise trials (1-flash, X-beep), 0-beep responses were significantly slower than 1-beep responses ($t_{(19)} = 7.67$; $p < 0.001$; $d = 0.62$), in line with the accuracy effect (Figure 2.4c). Moreover, participants responded significantly slower when 2 beeps were presented and SOA was short compared to 0-beep (25 ms: $t_{(19)} = 3.62$; $p = 0.015$; $d = 0.64$) and for short and intermediate SOAs compared to 1-beep presentations (25 ms: $t_{(19)} = 5.28$; $p < 0.001$; $d = 1.09$; 42 ms: $t_{(19)} = 4.12$; $p = 0.004$; $d = 0.74$; 50 ms: $t_{(19)} = 4.16$; $p = 0.004$; $d = 0.80$; 58 ms: $t_{(19)} = 3.34$; $p = 0.017$; $d = 0.65$; 75 ms: $t_{(19)} = 3.23$; $p = 0.018$; $d = 0.59$; 108 ms: $t_{(19)} = 2.67$; $p = 0.045$; $d = 0.44$).

Taken together, the pattern of response times for both 2IFC and yes-no task fits well with the signal detection, hit rate and false alarm performance measures. Therefore, it is unlikely that psychometric function and signal detection model results are confounded by response time differences.

Summary of results

Both, psychometric function and d' analyses showed that two flash perceptual resolution in the context of 0-beep or 1-beep is higher compared to 2-beep contexts. This effect is driven by fewer false alarms (respond “see 2” when 1 flash was presented) rather than more hits (respond “see 2” when 2 flashes were presented). Moreover, d' results demonstrate that perceptual sensitivity is higher in the 1-beep condition compared to when no beeps are presented due to fewer false alarms in the 1-flash, 1-beep setting. Response times are largely in line with sensitivity and accuracy findings.

Discussion

The temporal windows within which the sound induced fission and fusion illusions occur have previously been investigated using simple perceptual reports that are affected by both perceptual sensitivity and bias (Apthorp et al., 2013; Bidelman, 2016; Cecere et al., 2015; Foss-Feig et al., 2010; Hamilton et al., 2013; Narinesingh et al., 2017; Shams et al., 2000, 2002). Only one study investigated d' for both fission and fusion as a function of SOA, but did not statistically compare them (McGovern et al., 2014). Yet, both illusions have been associated with decreased perceptual sensitivity in comparison with unisensory presentations, as well as bias toward reporting the number of beeps when SOA was fixed (Chen et al., 2017; Kaposvári et al., 2014; McCormick & Mamassian, 2008; Mishra et al., 2008; Watkins et al., 2006). Therefore, we aimed to clarify the temporal determinants of one versus two flash perceptual discrimination accompanied by 0-2 beeps by dissociating perceptual sensitivity from bias using psychometric functions and signal detection theory. To account for the speed accuracy trade-off, we conducted complementary analyses on response times.

Regarding sensitivity, our results show consistently lower temporal resolution for the 2-beep condition compared to either 0 or 1-beep presentations. This effect was driven by more false alarms (report “see 2” when a single flash was presented) rather than more hits (report

“see 2” when two flashes were presented) indicating a larger TBW for the fission illusion. Interestingly, these sensitivity differences emerged only for longer SOAs (75 – 225 ms). This fits well with previous work showing a larger temporal range within which the fission illusion occurs compared to the fusion illusion (Apthorp et al., 2013) and demonstrates that this performance difference is due to perceptual sensitivity instead of response or perceptual bias (Figure 2.3b). In addition, it puts previous work into perspective that showed larger variability in fusion illusion susceptibility compared to fission when presenting stimuli at a fixed SOA: Due to the comparatively narrow window within which the fusion illusion occurs for a given individual, between-subject variability will lead to large fluctuations in fusion illusion rate when SOA is fixed (Andersen et al., 2004; Cuppini et al., 2014; Mishra et al., 2008; Shams et al., 2002).

On the other hand, the 1-beep condition showed similar or even larger sensitivity (Figure 2.3b) compared to unisensory flash presentations, because of fewer false alarms. This is consistent with previous research reporting lower sensitivity in the context of 1 compared to 0 beeps only for high but not low contrast peripherally presented flashes (Kaposvári et al., 2014). Indeed, when flash stimuli are weak, a single beep can boost the visual signal and lead to improved performance compared to unisensory stimulation (Noesselt et al., 2010). Once the fusion illusion diminishes sufficiently (i.e. for large SOAs), this can lead to higher two-flash discriminability in the 1-beep condition (as observed here, Figure 2.3b). In general, however, the fusion effect has been reported to be much smaller compared to the fission effect (Andersen et al., 2004; Shams et al., 2002).

Both 1-beep and 2-beep conditions were associated with a bias to report the number of presented beeps, similar to previous reports (Chen et al., 2017; Kaposvári et al., 2014; McCormick & Mamassian, 2008). Note that sensitivity (d') in this task plausibly reflects

perceptual processes, whereas C could reflect perceptual or response/decisional bias (Witt et al., 2015).

Fission and fusion seem to be two specific, but related instances of audio-visual integration, which can be explained by simple (neuro-)computational models (Cuppini et al., 2014; Shams et al., 2005). Nevertheless, evidence suggests that the underlying neural mechanisms are at least partially distinct. The fission illusion is stronger when flashes are displayed in the periphery (driven by d') rather than foveally, but occurs for both locations reliably (Chen et al., 2017; Kaposvári et al., 2014). In contrast, the fusion illusion has been frequently reported to be stronger for centrally presented flashes (Chen et al., 2017) (although this can be driven by bias (Chen et al., 2017)). A proposed explanation for this is that the temporal resolution of the parvocellular, ventral pathway (driven foveally) is lower compared to the magnocellular, dorsal pathway (driven peripherally) (Kaposvári et al., 2014). Hence, the relatively weaker fusion illusion is more likely to be observed when visual resolution is low. Interestingly, connectivity between A1, but also higher order polysensory areas and V1 is more developed in the periphery in monkeys (Clavagnier et al., 2004; Falchier et al., 2002; Rockland & Ojima, 2003), potentially explaining stronger fission illusions in the periphery (Shams, 2012).

There are also findings from neuroimaging, neuropsychology and neurostimulation showing discrepancies between fission and fusion. Functional magnetic resonance imaging (fMRI) demonstrated that both illusions lead to activity in retinotopically defined V1 and V2. Specifically, an illusory flash (fission illusion) led to augmented BOLD activity and a fused single flash (fusion illusion) led to attenuated activity (Watkins et al., 2007, 2006). Higher order polysensory areas in superior temporal sulcus (STS) exhibited enhanced activity for both illusions (Watkins et al., 2007, 2006). On the other hand, event related potential (ERP) studies

identified different time courses of modulations by an illusory double flash (fission modulation relative to stimulus 2 onset: ~40-75 ms in visual cortex; ~50-80 ms in STS (Mishra et al., 2007)) or single flash (fusion: ~100-130 ms in STS; ~160-180 ms in visual cortex (Mishra et al., 2008)). Furthermore, transcranial direct current stimulation (tDCS) applied over V1 or STS modulates fission, but not fusion (Bolognini et al., 2011), and illusion prevalence is differently expressed for patients with visual field deficits or unilateral spatial neglect (for details see (Bolognini et al., 2016)). Finally, the fission illusion was found to be a reliable phenomenon in both young and elderly adults, whereas susceptibility to the fusion illusion declines with age (Bolognini et al., 2016; DeLoss & Andersen, 2015; McGovern et al., 2014).

All in all, it seems clear that fission and fusion are at least in part distinguishable phenomena in terms of neural processing. However, many of these neuroimaging and neuropsychological studies did not segregate perceptual effects from bias. It is therefore possible that some of these findings can be explained by different biases for fission and fusion. Furthermore, findings are often based on the absence of an effect, which has to be interpreted with caution, especially given that the fusion illusion is weaker in general and exhibits more between subject variability when a single SOA is presented (Andersen et al., 2004; Cuppini et al., 2014; Mishra et al., 2008; Shams et al., 2002).

In conclusion, we demonstrated a consistently larger temporal window of integration in the fission illusion compared to both fusion illusion and unisensory two-flash fusion. This is a genuine perceptual effect and not confounded by the speed-accuracy trade-off. Both fusion and fission showed a bias toward reporting the number of presented beeps. These results fit well with previous work and stress the relative strength of the fission illusion compared to fusion, which is reflected in psychometric function slope, but not threshold. Moreover, they highlight

the importance of dissociating perceptual sensitivity and bias, and to consider response time differences between conditions.

Methods

Participants

After giving informed consent 20 right-handed healthy adults (11 female, mean age: 22.4; age range 19 - 30) completed the study. An additional six participants were excluded after the first testing session, because the eye-tracker could not be reliably calibrated (three participants) or because participants responded too slowly, multiple times or not at all on > 10% of trials in the first two-interval forced choice run (Supplementary Figure 2.1). All participants had normal or corrected to normal vision and reported unimpaired hearing. Participation was compensated with £7.50 per hour. Ethical approval was granted by the University of Birmingham Science, Technology, Engineering, and Mathematics Review Committee (approval number ERN_11-0429AP22).

Stimuli

The visual stimulus was a truncated Gaussian light-grey circular blob with a diameter of 4° and a standard deviation of 1.34° (maximum luminance 12.91 cd/m^2) presented on a dark-grey background (0.71 cd/m^2) for approximately 2 ms. The auditory stimulus was a 2 ms pure tone (3500 Hz) with a 0.5 ms linear ramp at on- and offset (maximum amplitude at the left earpiece was measured at 80 dB SPL). All stimuli were presented 15° to the left of a central light-grey fixation cross (12.91 cd/m^2). Onset of audio-visual stimulus pairs was synchronous. All stimuli were created in Matlab 2014a (Mathworks, Natick, MA, USA) and presented with Psychtoolbox 3 (<http://psychtoolbox.org>).

Design and procedure

The study consisted of 2 (task: Two-interval forced choice, yes-no) x 3 (two flashes presented with: 0-2 sounds) experiments (Figure 2.1). A normal testing schedule consisted of 5 days, each including two task runs (~70 min of task performance per run; see Supplementary Figure 2.1 for details). Before the first task run of a session, participants completed 48 practice trials.

Two-interval forced choice task. In the two-interval forced choice (2IFC) task (exp. 1-3) a target (2-flash, X-sound) and distractor (1-flash, X-sound) interval were presented with 800 ms delay in random order. Participants responded which interval contained two flashes. Stimulus presentations followed a factorial 2 x 3 x 8 design, denoting the interval order (1-2 or 2-1 flashes), number of sounds (0-2), and stimulus onset asynchronies (SOAs; 0.025, 0.042, 0.05, 0.058, 0.075, 0.108, 0.158, 0.225 ms). The response window was 1.5 s. It took another 1.5 s plus jitter (uniformly sampled between 0 and 250 ms) until the next trial started. To distinguish inter-trial intervals from 2IFC-interval delays the fixation cross was rotated by 45° after each trial. A run consisted of 12 blocks of 96 fully randomized trials of each type. Two 2IFC runs were acquired per participant for a total of 48 trials per condition. Importantly, the 2IFC task design leads to theoretically unbiased responses after pooling over trials with different interval orders.

Yes-no task. In the yes-no task (exp. 4-6) a single interval was presented per trial and participants indicated whether they perceived one or two flashes (Figure 2.1). Note that in this case response bias does play a role. An additional 0.6 s delay was included between trials to ensure that pre-stimulus EEG signals did not include response related activity (not relevant here). Each block consisted of 97 trials and included one catch trial where no stimulus was presented. Overall 96 trials were acquired per condition.

Behavioural responses and key mappings. In the 2IFC and yes-no task (exp. 1-6) participants responded with their index fingers using the 'F' and 'J' keys on a standard computer keyboard.

An equal number of trials were acquired for each of the two possible key mappings for each task. Key mapping was counterbalanced between sessions and participants. Participants were given at least 48 practice trials to get accustomed to a change in key mapping. They were instructed to correct mistakes if they noticed them.

Behavioural exclusion criteria. All trials with multiple responses on the current or previous trial were discarded from all analyses (2IFC: $M = 109.9$, range: 20, 260; Yes-no: $M = 196.6$, range: 23, 590). In addition, trials with response times faster than 100 ms after last stimulus onset were excluded (2IFC: $M = 14.89$, range: 1, 36; Yes-no: $M = 0.54$, range: 0, 7). This resulted in the following trial distributions: 2IFC (exp. 1-3; $M = 2170.3$, range: 1940, 2278); Yes-no (exp. 4-6; $M = 4410.6$, range: 3860, 4756).

Experimental setup

Testing took place in a darkened room with additional light-shielding the stimulus PC and participant. The presentation screen was a 19'' cathode ray tube (CRT) display with a refresh rate of 120 Hz and a resolution of 1024 x 768 pixels. Sounds were presented via EARtone 3A Insert Earphones (Aearo Company Auditory Systems, 1997). Participants' gaze position was tracked with an EyeLink 1000 Plus eye-tracker to be able to exclude trials where central fixation was not maintained (SR Research Ltd., 2014).

Psychometric function analysis

To obtain an estimate of perceptual window length for each participant and flash-sound pairing (i.e. 1FXS & 2FXS) in the 2IFC (exp. 1-3) and yes-no task (exp. 4-6) we described response accuracy as a psychometric function of SOA. Formally, this can be written as follows:

$$(1) \quad \psi(x; \alpha, \beta, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda) F_L(x; \alpha, \beta)$$

with

$$F_L(x; \alpha, \beta) = \frac{1}{1 + \exp(-\beta(x - \alpha))}$$

where x denotes SOA, F_L is the logistic function, and parameters α, β, γ and λ describe threshold (inflection-point), slope, guess-rate and lapse-rate, respectively. The threshold parameter α is the point at which participants can distinguish 1 from 2 flashes 50% of the time, discounting guesses and lapses. It therefore serves as a measure of perceptual resolution (Samaha & Postle, 2015). However, this measure is affected by participants' bias. Hence, the 2IFC task, which theoretically leads to unbiased responses (Macmillan & Creelman, 2005) provides an estimate of α (almost) purely driven by perceptual sensitivity, while the yes-no task α estimates are probably tainted by bias.

Psychometric function parameters were estimated using maximum likelihood estimation (MLE), as implemented in the Palamedes toolbox (Prins & Kingdom, 2018) for Matlab. First, a grid search was performed to find the most likely parameter values. Second, a Nelder-Mead simplex search (Nelder & Mead, 1965) found the maximum likelihood estimates (MLE) for threshold and slope parameters, keeping guess-rate and lapse-rate fixed. Finally, all parameters were allowed to vary in a second Nelder-Mead simplex search to reveal the final parameter estimates.

To improve estimation of lapse-rate, psychometric functions were jointly estimated over conditions (1F & 2F, 1F1S & 2F1S, 1F2S & 2F2S), and a single lapse-rate parameter was included. The guess-rate was allowed to vary (and not fixed at 0.5), because participants might react to stimulus salience differences from temporal summation effects (Gorea, 2015). Goodness of fit was assessed using a bootstrapping procedure (Wichmann & Hill, 2001). When goodness of fit of a model was insufficient psychometric functions were estimated separately for different conditions.

Signal detection analysis

For each SOA and condition (0-2 beeps), d' (sensitivity) and criterion (C) were estimated under the assumptions of a two equal-variance Gaussian signal detection model (Macmillan & Creelman, 2005). In the yes-no task participants' internal 1-flash, X-sound representations are drawn from the noise distributions and 2-flash, X-sound representations are drawn from the signal distributions. In the 2IFC task the noise distribution is defined as "2 flashes in first interval", whereas the signal becomes "2 flashes in second interval". Sensitivity d' is the distance between a given pair of signal and noise distribution means, which can be estimated as follows:

$$(2) \quad d' = z(HR) - z(FAR),$$

where HR denotes hit rate (e.g. proportion of "see 2" responses when 2 flashes were presented) and FAR denotes false alarm rate (e.g. proportion of "see 2" responses when 1 flash was presented). The estimate of criterion C_{NOISE} relative to the noise distribution only is computed as follows:

$$(3) \quad C_{NOISE} = -z(FAR),$$

and captures participants' tendency to report "see 2" when 1 flash was presented. Alternatively, the criterion is computed relative to the theoretically optimal criterion placement, creating a measure of bias:

$$(4) \quad C = -0.5(z(HR) + z(FAR)),$$

where $C > 0$ marks a bias toward "see 1" (yes-no) or "2 flashes in first interval" (2IFC). Note that in the yes-no task there is no optimal criterion placement relative to a single point, because multiple signal distributions (one for each SOA) exist. A small constant of 0.1 was added to both hits and false alarms to avoid infinities in the calculation of d' and criterion.

Accuracy and response time analysis

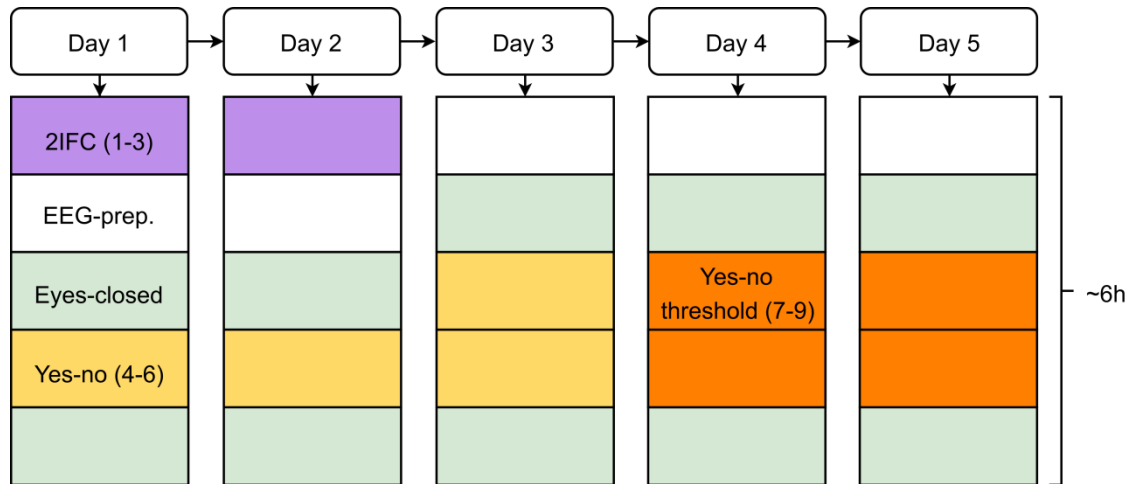
The simple signal detection model used here does not take into account response times. However, due to the speed-accuracy trade-off performance accuracy and response speed are inherently related (Liu & Watanabe, 2012). Moreover, as d' measures the distance between the means of the noise (1-flash) and signal (2-flash) distribution, an increase in d' can be due to fewer false alarms (noise trials) or more hits (signal trials). Therefore, we compared both response times and the proportion of correct trials separately between 2-flash conditions (hit rate), and 1-flash conditions (false alarm rates).

Statistical comparisons

Standard parametric statistics (repeated measures ANOVAs and paired t-tests) were used if assumptions were met, as tested by visual inspection of quantile-quantile plots, residual plots and Shapiro Wilk tests (normality rejected at the 0.001 level). The sphericity assumption was rejected at the 0.05 level with Mauchly's test of sphericity and the Greenhouse-Geisser correction was used when necessary. Otherwise standard non-parametric tests were performed (Wilcoxon sign-rank test). Post-hoc tests were corrected with the Bonferroni-Holm method to account for type 1 error inflation due to multiple comparisons (Holm, 1979).

In the 2IFC task only 7 instead of 8 SOAs were acquired for one participant. This participant was excluded from ANOVA, but not post-hoc comparisons for the SOAs where data was available.

Supplementary materials



Supplementary Figure 2.1. Typical study time table.

A planned testing schedule consisted of 5 days with approximately 6 hours per day. Square blocks indicate experimental runs, EEG preparation and eyes-closed resting state recordings. An experimental run included roughly 70 minutes of pure task time and breaks were flexibly offered in-between task blocks (~6 min task time per block). Eyes-closed runs lasted between 2 and 3 minutes.

Chapter 3 - Decoding real and illusory flash representations during perceptual decision-making

The following chapter is being prepared for scientific publication and is the result of collaboration between Uta Noppeney and Steffen Bürgers. U.N. conceived the study. S.B. and U.N. designed the experiments. S.B. collected the data. S.B. analysed the data. S.B. and U.N. interpreted the data. S.B. wrote the manuscript. U.N. provided direct supervision at all stages of the process.

Abstract

In the sound induced flash illusion, the number of perceived flashes is altered towards the number of concurrently presented beeps: In the fission illusion an illusory second flash is reported, whereas in the fusion illusion two flashes are merged into one. Event related potentials (ERPs) suggest that fission and fusion time courses are distinct early on (within 200 ms post-stimulus), but little is known about how perceptual decisions differ close to the behavioural response. Here, we leveraged the advantages of multivariate pattern classification to track and compare the perceptual representations of one versus two flashes in a unisensory, fission and fusion illusion setting (1-2 flashes accompanied by 0-2 beeps). Perceptual representations of one versus two flashes could be decoded approximately between 100 and 400 ms post-stimulus onset for 0 and 1-beep, but not 2-beep presentations. Prior to response, perceptual choice was decodable in a sustained fashion for all conditions. Remarkably, representations generalized well between 0 and 1-beep conditions, but neither 0, nor 1-beep representations generalized to the fission illusion (or vice versa). These findings suggest that real and illusory 2-flash neural representations differ during perceptual decision-making.

Introduction

The brain seamlessly combines inputs from multiple senses to form close to optimal representations of its environment following Bayesian inference (Körding et al., 2007; Rohe & Noppeney, 2015a; Shams, 2012; Shams et al., 2005). Inputs that are likely to originate from a common source are integrated weighted by their respective reliabilities to create a more reliable estimate than either of its unisensory parts (Ernst & Bühlhoff, 2004; Körding et al., 2007). Striking examples are the sound induced flash illusions (SIFIs), in which the number of perceived flashes is altered by concurrently presented beeps (Shams et al., 2000, 2002). Specifically, the fission illusion leads to “see 2” perceptual reports when 1 flash and 2 beeps are presented, whereas the fusion illusion leads to “see 1” reports when 2 flashes and 1-beep are presented.

Even though fission and fusion can be explained by the same (neuro-)computational models (Cuppini et al., 2014; Shams et al., 2005), evidence suggests that the generating neural mechanisms are at least partially different (Bolognini et al., 2016, 2011; DeLoss & Andersen, 2015; McGovern et al., 2014; Mishra et al., 2008, 2007; Watkins et al., 2007, 2006). A first indication for such a difference comes from developmental research showing that fusion, but not fission declines with age (DeLoss & Andersen, 2015; McGovern et al., 2014). Furthermore, the propensity to fission is sometimes, but not always correlated with fusion rate within participants (Bolognini et al., 2016; Vanes et al., 2016). Psychophysics studies demonstrated that the fission illusion is more pronounced when flashes are presented in the periphery, which has been attributed to the more dense connections between A1 and polysensory areas with peripheral compared to central V1 (Chen et al., 2017; Shams, 2012). On the other hand, the fusion illusion tends to be stronger in the centre, presumably because of the relatively lower temporal resolution of the ventral parvocellular pathway (Chen et al., 2017; Kaposvári et al., 2014).

Neuroimaging research further supports (partially) separate neuronal mechanisms, identifying similar brain networks, but different processing time courses. Functional magnetic resonance imaging (fMRI) revealed stronger blood oxygen level dependent (BOLD) activations in retinotopically defined V1 for fission and decreased activation for fusion, similar to activity evoked by unisensory 2-flash and 1-flash presentations (Watkins et al., 2007, 2006). At the same time superior temporal sulcus (STS) activity increased for both illusions (Watkins et al., 2007, 2006). Instead, event related potential (ERP) analyses demonstrated distinct progressions of difference waves for illusion versus non-illusion outcomes (Mishra et al., 2008, 2007). Specifically, fission was associated with differential evoked activity between ~40-75 ms after second stimulus offset over occipital sensors, and between ~50-80 ms over superior temporal cortex (Mishra et al., 2007). Fusion showed an ERP difference after ~100 ms over superior temporal cortex and only after ~160 ms over occipital cortex (Mishra et al., 2008).

Finally, neurostimulation and lesion studies show differential modulation of fusion and fission (Bolognini et al., 2016, 2011). In particular, transcranial direct current stimulation (tDCS) modulated fission, but not fusion occurrence such that V1 stimulation lead to fewer and STS stimulation to more “see 2” reports (Bolognini et al., 2011). Patients with visual field deficits due to lesions in striate cortex were less susceptible to the fission illusion compared to healthy controls and patients with unilateral spatial neglect (Bolognini et al., 2016). Patients with unilateral spatial neglect showed both fission illusion and fusion illusion, but fusion was not observed for either of the other groups (Bolognini et al., 2016).

Taken together, therefore, it appears that fission and fusion share similar underlying neural architectures, involving activity in primary visual cortex, as well as polysensory areas such as STS. Yet, they may be generated in different manners. The time differences observed in ERPs suggest that fusion is more likely to be driven by feedback projections from higher

order multisensory areas to V1, while fission might be a more direct process due to feedforward projections from A1 to visual cortex. An important caveat to consider is that both fission and fusion are associated not only with perceptual changes, but also decisional (possibly also perceptual (Witt et al., 2015)) bias toward the number of presented beeps (Chen et al., 2017; Kaposvári et al., 2014; McCormick & Mamassian, 2008). Hence, some of these discrepancies could potentially be explained by different biases between the two illusions. For example, later time courses for fusion compared to fission could be explained by a perceptual fission and decisional fusion effect (Mishra et al., 2008, 2007). This interpretational problem is exacerbated by fusion being generally weaker than fission (Andersen et al., 2004; Shams et al., 2002).

Similarly, discrepant results might in part be explained by differences in metacognitive judgments, although such effects would be expected to occur somewhat later than early stimulus evoked activity (Müller et al., 2016; Tsalas et al., 2018). Metacognition in perceptual decision-making is participants' ability to assess the quality of the perceptual evidence (Deroy et al., 2016). One might conjecture that perceptual illusions would be associated with lower perceptual confidence than veridical precepts. Yet, previous work has shown that participants are as confident in their illusory flash reports as they are in real flashes (Abadi & Murphy, 2014). However, these findings were not based on rigorous psychophysical methods (i.e. they confound bias and sensitivity).

Surprisingly, previous work has focused solely on stimulus evoked activity, and little scrutiny has been given to the perceptual decision-making process leading up to response commitment. Moreover, the power of univariate analyses of ERPs is limited due to the multiple comparison problem (Luck & Gaspelin, 2017), and multivariate approaches are more suitable in detecting subtle effects distributed over a large number of electrodes (Grootswagers, Wardle,

& Carlson, 2016). In addition, multivariate pattern classification allows for a direct comparison of the neural patterns between conditions.

Therefore, to examine the underlying neural representations of one versus two flashes in the context of 0-2 beeps we applied a multivariate pattern decoding approach (Grootswagers et al., 2016). Importantly, we performed stimulus and response locked analyses, tracking the perceptual decision-making process from prior expectations (pre-stimulus) until the motor response. Furthermore, we directly compared the similarity of dynamic neural representations of one or two flashes between conditions using cross-condition generalization. If illusory flashes are represented in the same way as real flashes, we would expect to be able to generalize between 0-2 beep conditions. To minimize response bias, we deployed factorial designs and titrated perceptual performance to 50% performance accuracy in the critical 2-flash, 2-flash & 1-beep (fusion) and 1-flash & 2-beep (fission) conditions. In addition, we acquired confidence ratings of perceptual choices to gauge participants' metacognitive awareness of the internal flash representations and how it might differ depending on auditory context.

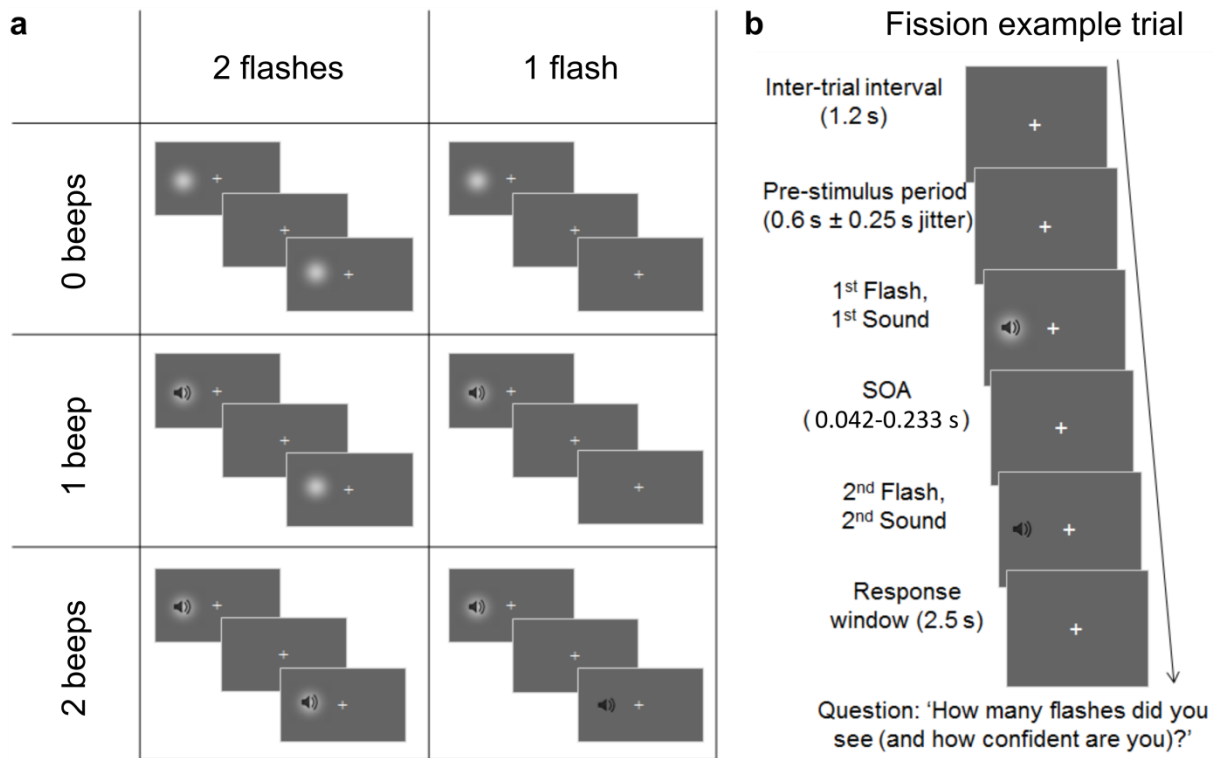


Figure 3.1. Study design and example trial presentation. *a*, Possible stimulus combinations on a given trial: 1-2 flashes paired with 0-2 beeps (catch trials are not depicted). *b*, Example fission trial (1 flash, 2 beeps) with details about temporal progression of stimuli. SOA was determined separately for 2-flash, 2-flash & 1-beep and 1-flash & 2-beep conditions for each participant to obtain an equal number of “see 1” and “see 2” responses. Note that screen dimensions, fixation cross and stimuli are not to scale or matched in contrast to the actual experiment.

Results

To scrutinize the one versus two flash perceptual decision-making processes in a unisensory and multisensory (0-2 beeps) context, we presented 20 young adult participants with stimuli close to their perceptual threshold (Figure 3.1). Specifically, a staircase procedure determined the stimulus onset asynchrony (SOA) at which participants reported “see 2” and “see 1” approximately equally often for the conditions 2-flash (unisensory flash fusion; $M = 56$ ms, range: 25, 75 ms), 2-flash & 1-beep (fusion illusion; $M = 63$ ms, range: 25, 92 ms) and 1-

flash & 2-beep (fission illusion; $M = 90$ ms, *range*: 42, 233 ms). Flash-beep combinations were presented in pseudo-random order from trial to trial.

Behavioural results

As expected from the staircase procedure, in the critical conditions (2-flash; 2-flash & 1-beep and 1-flash & 2-beep) participants responded “see 2” approximately 50% of the time (Figure 3.2b). Moreover, there was no significant d' condition effect ($F_{(1.37, 25.52)} = 2.80$, $p = 0.097$, $\eta_p = 0.128$), suggesting that titration of performance accuracy also lead to approximately equal sensitivity between beep contexts. Participants were significantly biased toward “see 2” reports in the 2-beep condition compared to either 0 or 1-beep contexts (Figure 3.2a; $F_{(1.41, 26.85)} = 52.50$, $p < 0.001$, $\eta_p = 0.734$; 0-beep – 2-beep: $t_{(19)} = 7.96$, $p < 0.001$, $d = 2.02$; 1-beep – 2-beep: $t_{(19)} = 7.62$, $p < 0.001$, $d = 2.23$; Bonferroni-Holm (bh) corrected (Holm, 1979)). This can be explained by 1-flash being considered as “noise” and 2-flash as “signal” (Figure 3.2d). Flipping the sign of C (or swapping signal and noise labels) in the 2-beep condition results in approximately equal criteria between beep contexts ($F_{(1.29, 24.47)} = 0.78$, $p = 0.42$, $\eta_p = 0.04$).

Considering only 2-flash trials, response time (RT) was significantly faster on “see 1” compared to “see 2” reports for both 0 and 1-beep conditions (upper panel in Figure 3.2c; $F_{(1,19)} = 5.79$; $p = 0.026$; $\eta_p = 0.233$; 0-beep: $t_{(19)} = -2.96$, $p = 0.024$, $d = -0.5284$; 1-beep: $t_{(19)} = -2.6376$, $p = 0.032$, $d = -0.49$; bh corrected). Furthermore, there was a significant main effect of condition ($F_{(2,38)} = 19.67$; $p < 0.001$), with significantly faster responses for both 0 and 1-beep compared to 2-beep presentations (0-beep – 2-beep: $t_{(18)} = -3.70$, $p = 0.005$, $d = -0.49$; 1-beep – 2-beep: $t_{(17)} = -2.68$, $p < 0.03$, $d = -0.34$; bh corrected). Similarly, on 1-flash trials RT was faster for “see 1” than for “see 2” reports in the 0 and 1-beep conditions (lower panel in Figure 3.2c; 0-beep: $t_{(19)} = -3.77$, $p = 0.004$, $d = -0.81$; 1-beep: $t_{(19)} = -3.37$, $p = 0.007$, $d = -0.79$; bh corrected).

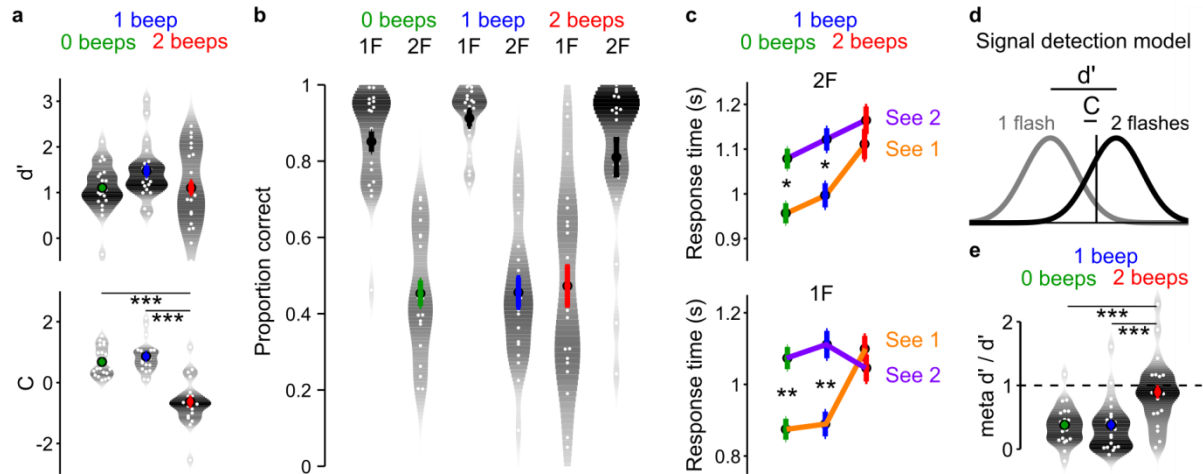


Figure 3.2. Behavioural results. *a*, Sensitivity (d') and bias (criterion C) for all conditions (0-2 beeps). White dots mark individual observations and grey shaded background marks the approximate sample distribution shape. Perceptual sensitivity was not significantly different between conditions. Participants were strongly biased toward “see 2” reports in the 2-beep compared to either of the other conditions. This bias is fully accounted for by 1-flash & 2-beep being considered as noise whereas 2-flash and 2-flash & 1-beep are considered as signals. *b*, Proportion correct for each flash-beep pairing. SOAs in the critical conditions 2-flash, 2-flash & 1-beep and 1-flash & 2-beep were titrated for 50% performance accuracy. In most cases performance in the control conditions (1-flash, 1-flash & 1-beep, 2-flash & 2-beep) was above chance. *c*, Response times by perceptual outcome (“see 1” or “see 2”) for 2-flash trials (upper panel) and 1-flash trials (lower panel). Participants tended to respond faster for “see 2” compared to “see 1” perceptual decisions. *d*, Signal detection model illustration. 1-flash presentations were treated as noise, 2-flash presentations as signal. *e*, Metacognitive efficiency (meta d' / d') quantifies how well participants are able to access perceptual evidence for making a confidence judgment about their perceptual choice. A value of 0 indicates no access, whereas a value of 1 indicates perfect access. Values smaller than 0 or larger than 1 are likely due to noise. Observers showed generally higher (and close to optimal) metacognitive efficiency in the 2-beep condition compared to both 0-beep and 1-beep. Error bars denote ± 1 within subject SEM. $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***), Bonferroni-Holm corrected. 1F, 1 flash; 2F, 2 flashes.

Choice confidence. In addition to assessing participants' perceptual decisions of seeing one or two flashes (level 1 choice) we gauged their metacognitive ability of assessing perceptual evidence in the form of a confidence judgment on a 4-point Likert scale (level 2 choice). To this end we computed meta d' , which is the perceptual sensitivity (d') value that would be expected if participants' had perfect access to their level 1 perceptual evidence when making a level 2 metacognitive judgment (Maniscalco & Lau, 2012). A ratio of meta- d'/d' can straightforwardly be interpreted as a measure of metacognitive efficiency, i.e. the degree to which observers are able to utilize the information available for the level 1 perceptual decision for their level 2 metacognitive judgment (Maniscalco & Lau, 2012). A metacognitive efficiency value of 1 signifies perfect access to level 1 perceptual evidence, whereas a value of 0 indicates no access.

There was a significant difference in metacognitive efficiency between auditory conditions ($F_{(1.47, 27.99)} = 23.09, p < 0.001, \eta_p = 0.55$). Specifically, both 0-beep and 1-beep metacognitive efficiency were significantly lower compared to the 2-beep condition, which on average was close to optimal (Figure 3.2e; 0-beep – 1-beep: $t_{(19)} = -0.00, p = 0.99, d = -0.00$; 0-beep – 2-beep: $t_{(19)} = -5.18, p < 0.001, d = -1.16$; 1-beep – 2-beep: $t_{(19)} = -5.15, p < 0.001, d = -1.06$). This finding fits potentially well with previous work demonstrating that when given a strong (monetary) incentive, participants were able to distinguish illusory from real '2-flash' percepts (Rosenthal et al., 2009), suggesting that participants might have used their metacognitive systems.

Univariate ERP results of perceptual outcome

To avoid confounds due to response related activity (or stimulus related activity when locking time courses to response) all trials with response times < 500 ms relative to second stimulus onset were discarded. In addition, to ensure differences between perceptual outcomes were not confounded by different trial numbers for recording sessions or response hands, we matched the

number of trials between response key mapping and perceptual outcome combinations. To be included in the analysis datasets needed to have at least 80 trials per condition.

Stimulus locked ERPs. First, we tried to replicate previous research reporting ERP differences between “see 1” and “see 2” perceptual outcomes for the fission and fusion illusion (Mishra et al., 2008, 2007). As the exact regions of interest were not reported by Mishra and colleagues, we instead used a functional localizer to determine electrodes and time points of interest (Luck & Gaspelin, 2017). Specifically, we contrasted physical one versus two flash presentations separately for 0-2 beep contexts (given that participants responded “see 1” for 0 and 1-beep trials and “see 2” for 2-beep trials). Paired t-tests were performed on all channels in the 0-400 ms time period relative to first stimulus onset. To correct for multiple comparisons we deployed cluster based permutation statistics (Maris & Oostenveld, 2007; Oostenveld, Fries, Maris, & Schoffelen, 2011). The functional localizer revealed distributed electrode clusters over the whole scalp with time ranges between ~150-400 ms after first stimulus onset. However, subsequent tests of perceptual outcome in the 2-flash, 2-flash & 1-beep and 1-flash & 2-beep conditions identified no significant clusters (Figure. 3.3a; all $p > 0.3$). Due to the variability in SOAs between participants, we furthermore investigated ERPs locked to the onset of the second stimulus. Again, no significant clusters were identified (all $p > 0.18$).

Differences in response speed between conditions (Figure 3.2c) are related to distinct preparatory activity (irrespective of response hand) (Hanks & Summerfield, 2017). For example, centro-parietal activity has been implicated in the accumulation of evidence toward a decision-threshold (Kelly & O’Connell, 2013; O’Connell et al., 2012; Twomey et al., 2015). Hence, it is possible that the same neural profile unfolds in a time lagged manner for the two perceptual decisions. To assess whether this is the case, we performed a control analysis where

response time was regressed out. As before, no significant differences emerged (Figure. 3.3c; locked to first stimulus: all p -values > 0.09 ; locked to second stimulus: all $p > 0.08$).

Note that by individually titrating SOAs for each participant and condition, this experiment is not optimized for high temporal precision, which is especially relevant for early ERP components (Mishra et al., 2008, 2007).

Response locked ERPs. ERPs leading up to response are much slower than stimulus evoked ERPs and should therefore be less affected by SOA variability (Hanks & Summerfield, 2017). Moreover, response locked activity is well suited to measure perceptual decision making and evidence accumulation for a particular choice or intention to act (O’Connell et al., 2012; Schurger, Sitt, & Dehaene, 2012). Therefore, an analysis was performed for the 800 ms time window leading up to the response including all electrodes. Consistent throughout conditions “see 2” responses were associated with a larger pre-response amplitude compared to “see 1” responses (Figure. 3.3b; 0-beep: $p = 0.012$; 1-beep: $p = 0.032$; 2-beep: $p = 0.084$; bh corrected). This pattern stayed largely unchanged after regressing out RT (Figure. 3.3d; 0-beep: $p = 0.015$; 1-beep: $p = 0.097$; 2-beep: $p = 0.012$; bh corrected).

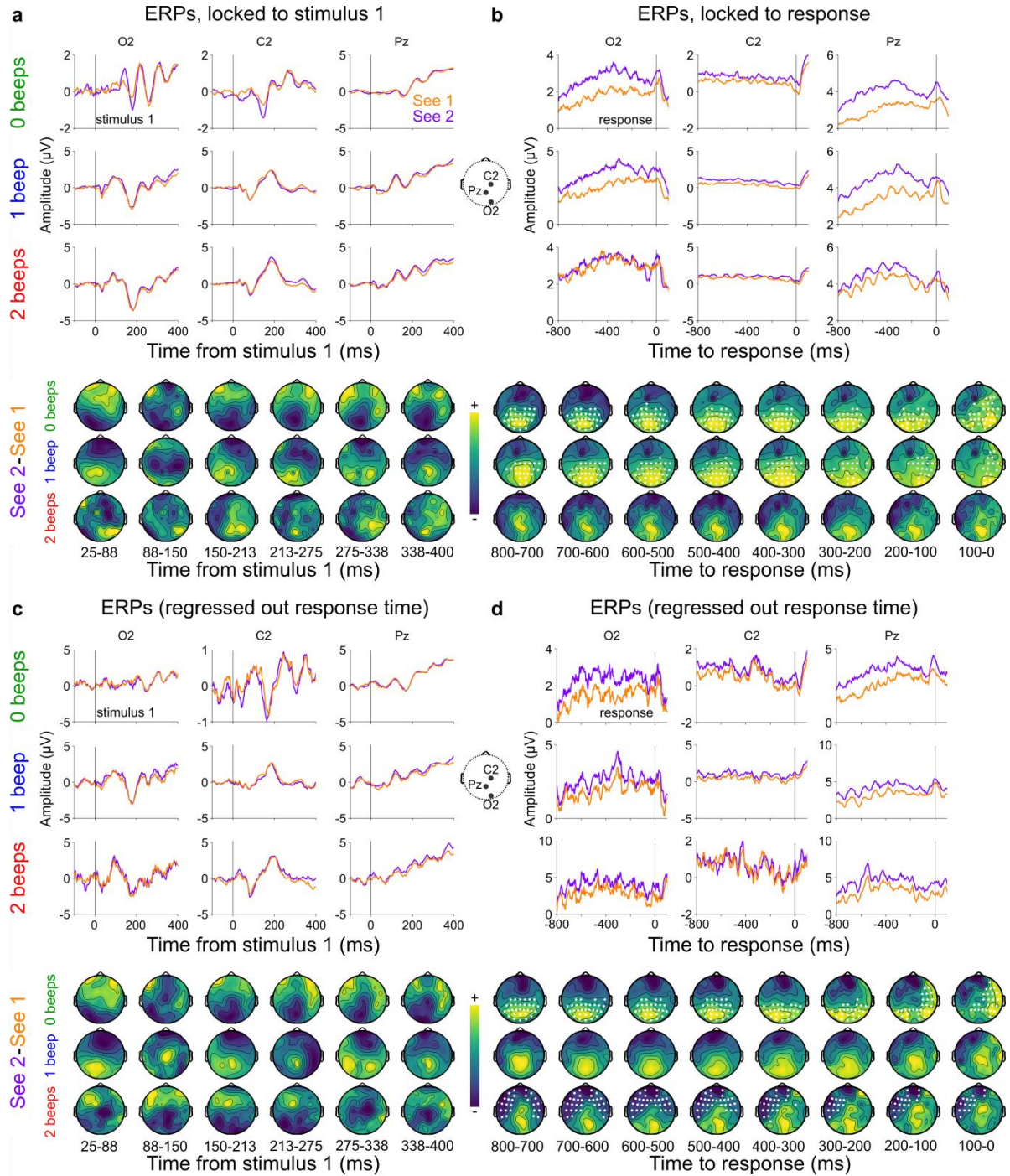


Figure. 3.3. ERPs of “see 2” versus “see 1” perceptual outcomes. a-d, Illustrative grand average ERPs for electrodes O2, C2 and Pz of perceptual responses and difference topographies (“see 2” – “see 1”) locked to stimulus 1 (a, c) or response (b, d). O2 and C2 were chosen based on previous work investigating the fission and fusion illusion (Mishra et al., 2008, 2007). Pz was chosen as a locus for measuring evidence accumulation in perceptual decision-making (Twomey et al., 2015). c, d, ERPs after

regressing out response time. White dots mark time points and electrodes contained in significant clusters (alpha level = 0.05).

Multivariate pattern decoding of perceptual outcome

A linear support vector machine classifier was trained to discriminate “see 2” from “see 1” perceptual outcomes for 2-flash, 2-flash & 1-beep and 1-beep & 2-flash conditions separately at each time point (Figure. 3.4a-d), and pooled over time (Figure. 3.4e-f). In the control analyses, we regressed out RT from single trial EEG time and channel samples for each condition in the training set, and applied the same correction to the test set to avoid creating dependencies between training and test data (Snoek, Miletić, & Scholte, 2019).

Stimulus locked decoding within conditions. Decoding accuracy (proportion correctly decoded labels) was significantly above chance when testing within conditions for 0-beep and 1-beep, but not 2-beep presentations (diagonals in Figure. 3.4a, c; violin plots in Figure. 3.4e). In the 0-beep condition three significant clusters emerged between 120 and 400 ms (119-150 ms: $p = 0.012$; 228-259 ms: $p = 0.019$; 369-400 ms: $p = 0.017$). Only the early cluster was retained after regressing out RT (119-166 ms: $p = 0.015$). In the 1-beep condition (fusion) there was a late cluster between 338 and 400 ms ($p = 0.003$). Moreover, there was an earlier trend, which was confirmed to be significant after regressing out RT (103-150 ms: $p = 0.008$; 306-400 ms: $p = 0.002$).

When training the classifier over the entire time window (0 - 400 ms post stimulus 1) decoding accuracy was significantly above chance for 0 and 1-beep, but not 2-beep presentations (Figure. 3.4e, Table 1). Results are largely unchanged when locking EEG time courses to second stimulus onset (Supplementary Figure 3.1).

Decoding accuracy fluctuating close to chance in the 2-beep condition (fission) could potentially be explained by the relatively larger SOA variability between participants (fission

range: 42, 233 ms; fusion range: 25, 92 ms; unisensory flash-fusion range: 25, 75 ms). In addition, fission analyses included fewer trials and participants, resulting in lower statistical power.

Stimulus locked decoding between conditions. Significant cross-condition generalization was observed when training on 2-flash trials and generalizing to either fusion (Figure. 3.4c, 322 - 353 ms: $p = 0.026$) or fission contexts (Figure. 3.4a; 353-369 ms: $p = 0.007$; Figure. 3.4c, 353 - 369 ms: $p = 0.018$). The latter was also significant when training on fission and generalizing to unisensory 2-flash presentations (353 - 369 ms: $p = 0.012$).

Response locked decoding within conditions. Each condition shows sustained above chance decoding accuracy leading up to response initiation, although only 0-beep and 1-beep conditions demonstrated statistically significant effects (diagonals in Figure. 3.4b; 0-beep: -569 – -428 ms: $p = 0.01$; -397 – -84 ms: $p = 0.002$; -6 – -88 ms: $p = 0.019$; 1-beep: -725 – -475 ms: $p = 0.003$; -366 – -209 ms: $p = 0.003$; -178 – -100 ms: $p = 0.016$). A very similar pattern with slightly attenuated decoding accuracies was obtained after regressing out RT (diagonals in Figure. 3.4d; 0-beep: -334 – -84 ms: $p = 0.002$; -6 – -41 ms: $p = 0.035$; 1-beep: -694 – -616 ms: $p = 0.014$; -319 – -209 ms: $p = 0.005$). This suggests that the classifier is driven by distinct neural activation patterns between perceptual outcomes rather than a similar activation pattern that is leading or lagging in time.

When classifying over the complete time window (-800 – 0 ms), all conditions demonstrated significantly above chance decoding accuracy. The same was true after regressing out RT, although the fission effect only tended toward significance (Figure. 3.4f; Table 1).

Response locked decoding between conditions. Between 0-beep and 1-beep presentations the classifier generalized consistently well (off-diagonals in Figure. 3.4b; 0-1 beep: ~-694 – -194 ms: all $p < 0.039$; 1-0 beep: ~-800 – -60 ms: all $p < 0.049$). This pattern was also observed after

regressing out RT (off-diagonals in Figure. 3.4d; 0-1 beep: -334 – -209 ms: $p = 0.005$; -53 – -6 ms: $p = 0.032$; 1-0 beep: ~-662 – -412 ms: all $p < 0.045$; -209 – -131 ms: $p = 0.008$). Instead, both 0-beep and 1-beep trained SVMs performed close to chance when tested on 2-beep presentations. This was also true when training on 2-beep and generalizing to 0 and 1-beep presentations. The same pattern was observed after regressing out RT (Figure. 3.4b, d).

To quantify evidence for the Null-hypothesis in the cross-condition generalization between 0 and 1-beep conditions with the 2-beep condition, we computed two-sided Bayes factors. Testing for both directions is sensible, because in principle it could be that a “see 2” fission neural representation leading up to response is the opposite of a fusion or unisensory neural pattern. For instance, the classifier might value neural patterns related to error-monitoring or cognitive control, which are related to the truthfulness of a neural representation (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Bayes factors almost exclusively support the Null-hypothesis (i.e. chance performance; Supplementary Figure 3.2).

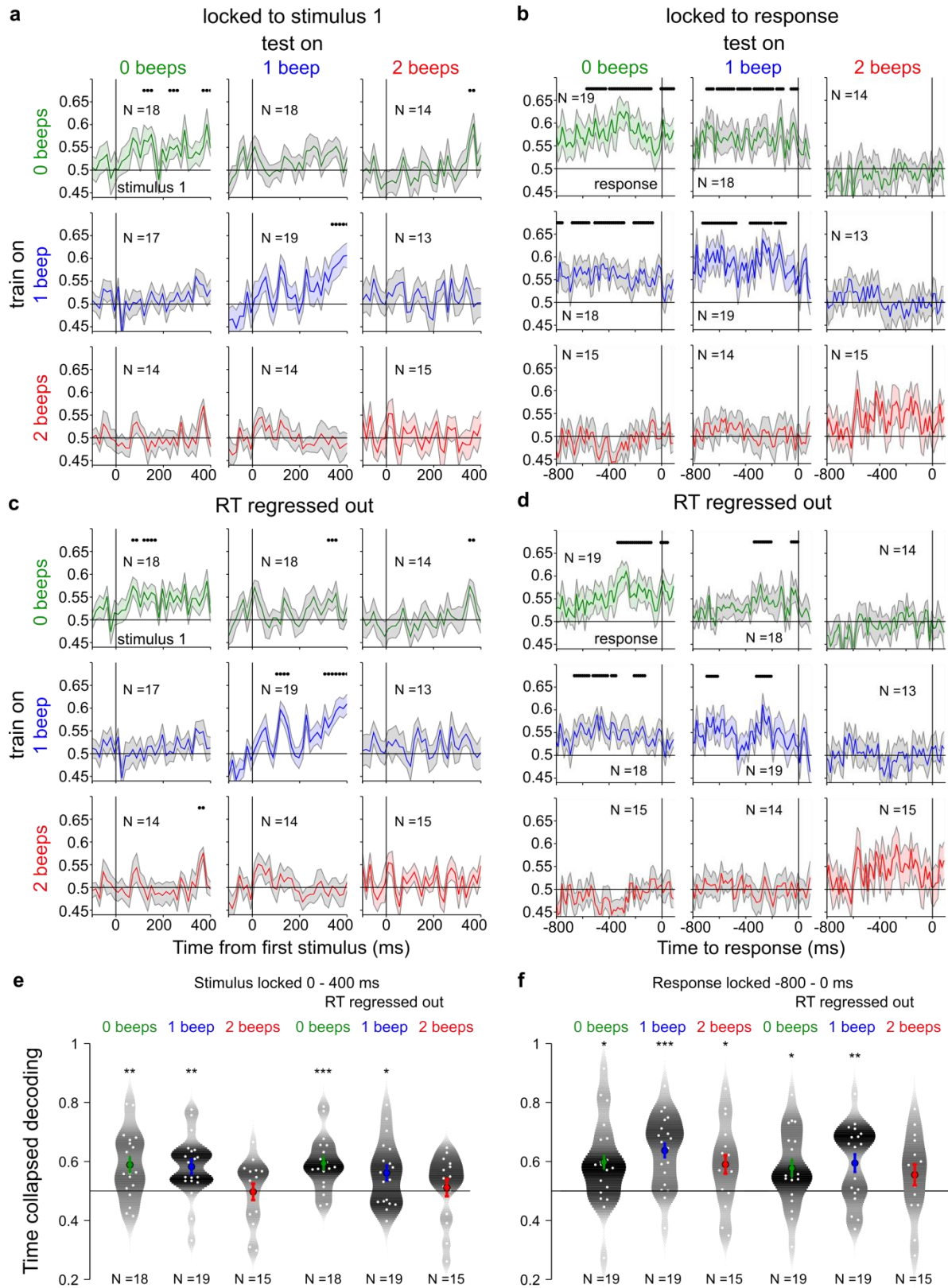


Figure. 3.4. Decoding perceptual outcomes. Ordinates show proportion correctly decoded labels (“see 1” or “see 2”). Chance level is 0.5. **a-d**, Within and across condition decoding accuracy time courses

locked to first stimulus onset (**a**, **c**) and response (**b**, **d**). Classifiers were trained on conditions described in row labels (foreground colour) and generalized to conditions denoted by column labels (background / error bar colour). Samples from clusters with decoding accuracies significantly above chance are marked with black dots. **e-f**, Time collapsed decoding accuracies within conditions (diagonals in **a-d**) locked to stimulus 1 (**e**) or response (**f**). Decoding accuracies significantly above chance are marked with asterisks, $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***). Error bars denote ± 1 SEM. RT, response time.

Table 3.1. Time collapsed within condition decoding accuracies of perceptual outcome

Condition	Stimulus locked				Response locked			
	N	t	p-value	d	N	t	p-value	d
0 beeps	18	3.28	0.004*	0.77	19	2.84	0.016*	0.65
1 beep	19	3.32	0.006*	0.76	19	4.81	0.0003*	1.10
2 beeps	15	-0.10	0.54	0.03	15	2.42	0.015*	0.63
RT regressed out								
0 beeps	18	4.26	0.0008*	1.00	19	2.56	0.020*	0.59
1 beep	19	2.42	0.026*	0.56	19	3.15	0.008*	0.72
2 beeps	15	0.38	0.35	0.10	15	1.50	0.078	0.39

d, Cohen's *d*. *P*-values are Bonferroni-Holm corrected per main effect (e.g. stimulus locked, RT not regressed out). Asterisks mark significance at the 0.05 level.

Assessing the role of confidence in classifying perceptual outcomes. Next, we assessed whether a difference in confidence for “see 1” versus “see 2” perceptual decisions between auditory conditions could potentially explain the different neural representations in the 2-beep condition compared to other auditory contexts. A RM ANOVA of condition (3 levels: 2-flash; 2-flash, 1-beep; 1-flash, 2-beep) x perceptual outcome (2 levels: “see 1” or “see 2” flashes) revealed a significant main effect of condition ($F_{(1.29, 24.45)} = 8.25$, $p = 0.005$, $\eta_p = 0.30$) and perceptual

outcome ($F_{(1, 19)} = 15.08$, $p < 0.001$, $\eta_p = 0.44$), but no significant interaction. Post hoc tests demonstrated significantly higher mean confidence on “see 1” compared to “see 2” responses, although there was only a trend toward significance in the 2-beep condition (0-beep: $t_{(19)} = 3.80$, $p = 0.001$, $d = 0.60$; 1-beep: $t_{(19)} = 4.38$, $p < 0.001$, $d = 0.78$; 2-beep: $t_{(19)} = 1.93$, $p = 0.068$, $d = 0.39$). Discrepancies in confidence scores between “see 1” and “see 2” flash responses suggest that the classifier might pick-up on these differences when trying to segregate the two classes (i.e. we might be decoding confidence, not perceptual outcomes).

We therefore repeated the decoding analyses for response-locked data after matching trial counts for all confidence levels between perceptual outcomes for each sound condition (2 flash; 2 flash & 1 sound; 1 flash & 2 sounds). Because we down-sampled the number of observations of the class with more trials this led to overall fewer samples. To mitigate this loss in statistical power we smoothed EEG time courses with a 28 Hz low-pass filter, resampled to 64 Hz and fed sliding time windows of three consecutive time samples into the classifier for the time resolved analyses (~50 ms). Moreover, we only investigated the 400 ms time window prior to response.

Within condition decoding accuracy resembled the previous results (diagonals in Figure 3.5a, left; 0-beep: ~-230 – -180 ms: $p = 0.047$; Figure 3.5a, right; 0-beep: $t_{(17)} = 2.34$, $p_{uncorr} = 0.016$, $d = 0.55$; 1-beep: $t_{(17)} = 1.73$, $p_{uncorr} = 0.050$, $d = 0.41$; 2-beep: $t_{(14)} = 1.94$, $p_{uncorr} = 0.037$, $d = 0.50$). This was also true when response time was regressed out (diagonals in Figure 3.5b, left; 0-beep: ~-230 – -180 ms: $p = 0.045$; 1-beep: ~-360 – -260 ms: $p = 0.019$; ~-230 – -180 ms: $p = 0.038$; Figure 3.5b, right; 0-beep: $t_{(17)} = 2.40$, $p_{uncorr} = 0.014$, $d = 0.57$; 1-beep: $t_{(17)} = 2.27$, $p_{uncorr} = 0.018$, $d = 0.53$; 2-beep: $t_{(14)} = 2.09$, $p_{uncorr} = 0.028$, $d = 0.54$).

Interestingly, cross-condition classification now revealed significant generalization between 1-beep and 2-beep conditions when trained on 2-beep and generalized to 1-beep, in

addition to the previously found generalization from 1-beep to 0-beep (Figure 3.5a). After regressing out response times these effects persisted, and there was also significant generalization from 1-beep to 2-beep (Figure 3.5b). It is therefore evident that the previous decoding analyses were influenced by differences in confidence scores between perceptual outcomes.

Can the influence of confidence ratings on classification of perceptual outcomes be explained by preparatory motor activity given that each confidence rating was associated with a different finger response (from pinkie = low confidence to index finger = high confidence)? Previous research suggests that preparatory motor activity of separate digits is difficult to distinguish with the relatively coarse spatial resolution of EEG, albeit not impossible with an optimized paradigm and decoding pipeline (Stankevich et al., 2016). However, in this design with only a single motor response per trial and no necessity for fully extending or flexing a finger to make a response, it is very unlikely that preparatory motor activity of distinct fingers influenced classification.

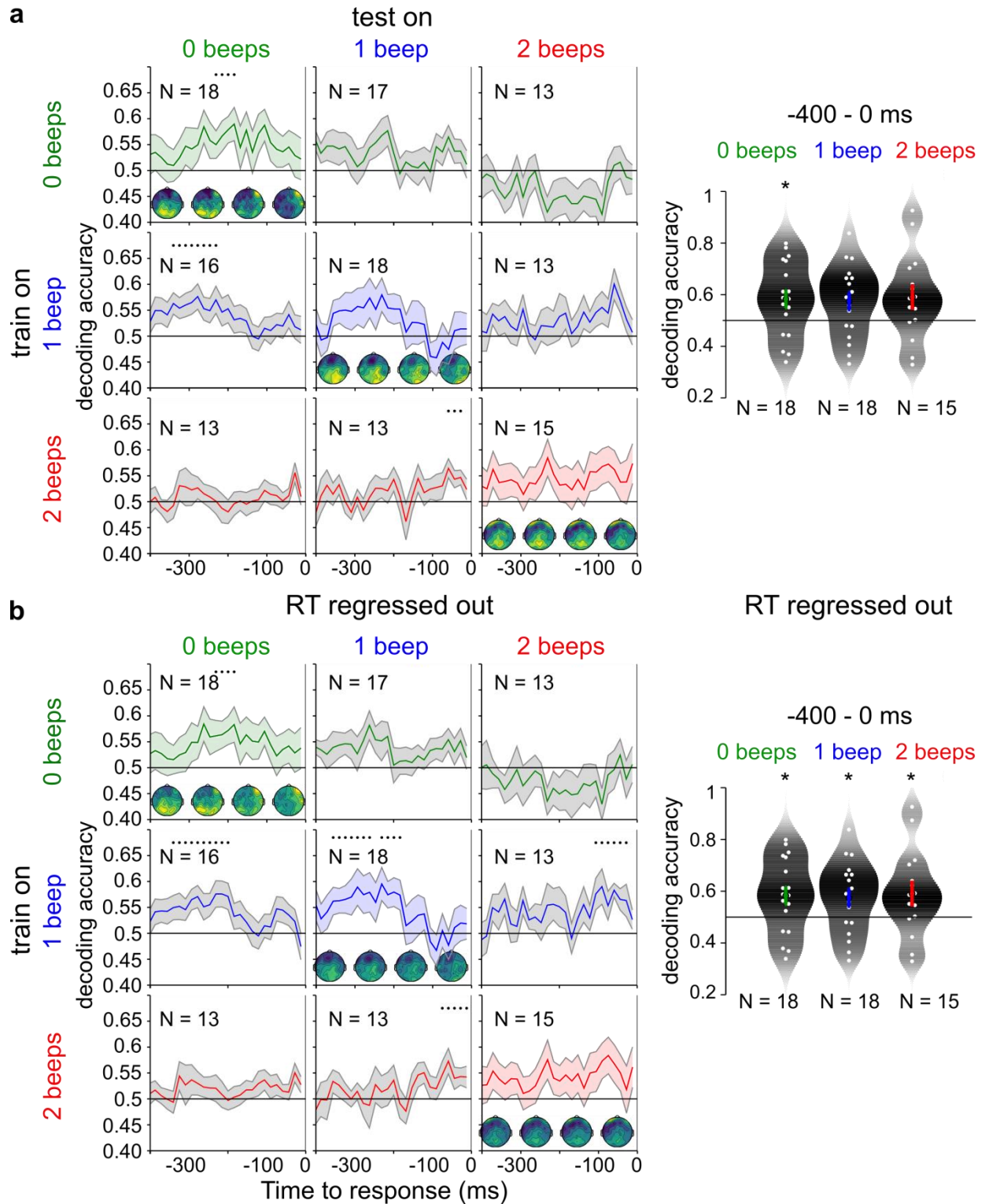


Figure 3.5. Classifier performance for confidence-matched behavioural outcomes. *a, b*, Ordinates show proportion correctly decoded labels (“see 1” or “see 2”). Chance level is 0.5. Left: Within and across condition decoding accuracy time courses locked to response (time = 0 ms). Classifiers were

trained on conditions described in row labels and generalized to conditions denoted by column labels. Topographies in diagonal denote the difference between perceptual outcome ERPs (“see 2” – “see 1”; colour map ranges from dark blue = negative to bright yellow = positive). Samples from clusters with decoding accuracies significantly above chance are marked with black dots. Right: Time collapsed decoding accuracies within conditions. Decoding accuracies significantly above chance are marked with asterisks, $p < 0.05$ (), $p < 0.01$ (**) and $p < 0.001$ (***). Error bars denote ± 1 SEM. RT, response time.*

Discussion

In the sound induced flash illusions the number of perceived flashes is altered by the number of concurrently presented beeps (Shams et al., 2000, 2002). Despite both fission and fusion being adequately explained by (neuro-)computational models of multisensory integration (Cuppini et al., 2014; Shams et al., 2005), ERP studies suggest that the time courses and underlying neural mechanisms are at least partially distinct (Mishra et al., 2008, 2007). To investigate these discrepancies further, in addition to univariate ERP analyses, we applied multivariate pattern decoding in a time resolved fashion. We thus tested whether and when “see 1” versus “see 2” perceptual representations emerge and generalize between different physical flash-beep contexts. Importantly, we explored the entire decision-making process beginning prior to stimulus presentation until the behavioural response.

EEG signatures prior to first stimulus onset (-100 – 0 ms) carried no decodable information about perceptual outcomes (Figure. 3.4a, c). Instead, stimulus-locked analyses revealed significant decoding accuracies between approximately 120 and 400 ms post-stimulus 1 onset for both 2-flash and 2-flash & 1-beep (fusion) presentations, but not for the 1-flash & 2-beep presentations (fission). Moreover, the classifier was able to generalize between 2-flash and 2-flash & 1-beep presentations for both early and late time epochs. In addition, 2-flash and

1-flash & 2-beep classifications generalized at late epochs (>300 ms). These results were relatively stable irrespective of whether response time was regressed out or not (Figure. 3.4a, c). However, exact time ranges have to be interpreted with caution due to the variability in SOAs between participants.

Response-locked analyses identified sustained perceptual choice representations for all conditions (0-2 beeps) in the ~ 600 ms time window leading up to the motor response, both with and without regressing out RT (diagonals in Figure. 3.4b, d; the 2-beep effect did not pass the 0.05 threshold for statistical significance). Remarkably, these neural patterns generalized well between 2-flash and 2-flash & 1-beep (fusion) conditions, but not at all between either of these conditions and 1-flash & 2-beep presentations (fission; off-diagonals in Figure. 3.4b, d; Supplementary Figure 3.2). Given that perceptual outcome could be decoded within the fission condition (at least to some extent; Figure. 3.4f; Table 1), this finding suggests that despite the perceptual decision always being the same (“see 1” or “see 2”), the neural representations thereof are different between unisensory 2-flash and fusion illusion contexts compared to the fission illusion.

However, a subsequent control analysis in which confidence ratings were matched between perceptual outcomes revealed that cross-condition generalization was indeed possible between fusion and fission contexts, but still not between unisensory and fission contexts. Hence, the fusion condition appears to contain neural features that generalize to both unisensory and fission settings. This could potentially be explained by sensory evoked activity due to 2 flashes being similar between 2-flash and 2-flash & 1-beep conditions in how it influences perceptual outcomes, whereas in the 1-flash & 2-beep condition it is not. Moreover, fission and fusion might share neural activity patterns influencing perceptual outcome that are related to auditory stimulation.

If perceptual representations between fission and the 0-beep condition are different, why are we able to generalize between 2-flash and 1-flash & 2-beep conditions post-stimulus? Considering that decoding accuracy within the fission condition is close to chance, it is unlikely that this short-lived effect (~16 ms) represents meaningful generalizations between conditions. Moreover, stimulus locked generalizations are particularly difficult to interpret (especially when they are transient), considering the variability in SOA not only over participants, but between conditions.

ERP topographies suggest posterior parietal cortex to be principally involved in perceptual outcome differences (Figure. 3.3b, d; although this is somewhat less clear after matching confidence scores; Figure 3.5). Indeed, previous work on perceptual decision making has established posterior parietal cortex as a seat of perceptual evidence accumulation (Kelly & O'Connell, 2013; O'Connell et al., 2012; Philiastides, Heekeren, & Sajda, 2014; Roitman & Shadlen, 2002; Twomey et al., 2015). However, these studies usually employ continuous sensory stimulation for one of two choices, which is thought to be linearly summed until a fixed decision-threshold is reached (Hanks & Summerfield, 2017). Here, we asked participants to detect a discrete signal, which is either present or not (a second flash). It is therefore not entirely clear what pattern of neural activity to expect prior to response. If parietal activity does indeed reflect accumulation of evidence for a particular decision, it is plausible that “see 2” trials show overall higher activity, simply because more evidence has been accumulated. Alternatively, such activity could also be explained by “see 2” representations being linked to stronger sensory evoked traces, which have been shown to be sustained during perceptual decision-making in a stimulus detection task (Mostert et al., 2015). However, these purely perceptual representations decayed relatively early (~350 ms post-stimulus), and later processes were related to the

decision variable (Mostert et al., 2015). We therefore believe it likely that response-locked findings largely reflect a decision variable.

Interestingly, participants also demonstrated substantially higher and close to optimal metacognitive efficiency in the fission condition compared to either unisensory or fusion conditions (Figure 3.2e). This further suggests that the neural mechanisms underlying fission are in part distinct from the fusion condition.

In summary, even though the sound induced flash illusions (fission and fusion) can be explained by the same computational principles (Cuppini et al., 2014; Shams et al., 2005), their underlying neural processes have been implicated to be divergent (Mishra et al., 2008, 2007). Here, we directly probed the perceptual representations of one versus two flashes in the context of variable numbers of beeps (0-2) using multivariate pattern classification. While fusion and fission contexts shared EEG features of one versus two flash perceptual representations, only the fusion context shared features with the unisensory condition, suggesting an intriguing and subtle difference between fusion and fission. Moreover, observers' metacognitive efficiency was substantially higher and close to optimal in the fission condition compared to either unisensory or fusion contexts, further arguing in favour of a partially different neural process for fission compared to fusion.

Methods

Participants

After giving informed consent 20 right-handed healthy adults (11 female, mean age: 22.4; age range 19 - 30) completed the study. All participants had normal or corrected to normal vision and reported unimpaired hearing. Participation was compensated with £7.50 per hour. Ethical approval was granted by the University of Birmingham Science, Technology, Engineering, and Mathematics Review Committee (approval number ERN_11-0429AP22).

Stimuli

The visual stimulus was a truncated Gaussian light-grey circular blob with a diameter of 4° and a standard deviation of 1.34° (maximum luminance 12.91 cd/m^2) presented on a dark-grey background (0.71 cd/m^2) for approximately 2 ms. The auditory stimulus was a 2 ms pure tone (3500 Hz) with a 0.5 ms linear ramp at on- and offset (maximum amplitude at the left earpiece was measured at 80 dB SPL). All stimuli were presented 15° to the left of a central light-grey fixation cross (12.91 cd/m^2). Onset of audio-visual stimulus pairs was synchronous. All stimuli were created in Matlab 2014a (Mathworks, Natick, MA, USA) and presented with Psychtoolbox 3 (<http://psychtoolbox.org>).

Design and procedure

The study consisted of a yes-no task with three experimental conditions (two flashes presented with: 0-2 beeps) and was completed over 2 testing sessions. At the beginning of a session, participants completed 48 practice trials. On each trial participants were presented with either one or two flashes accompanied by 0-2 beeps (Figure 3.1a). Overall 288 trials were acquired per condition and 73 trials were presented in each block including 1 catch trial and 12 trials per flash-beep combination. For each participant and 2-event condition a staircase procedure determined a single SOA close to perceptual threshold at which a participant perceives 2 flashes approximately 50% of the time. Trial order was randomized with the constraint that no two trials of the same condition could follow in direct succession. Participants indicated what they perceived, as well as how confident they were in their decision (not analysed here). Participants had 2.5 s to make a response and a subsequent trial started after another 2.1 s plus jitter (0-250 ms, uniformly sampled; Figure 3.1b).

Behavioural responses and key mappings. Participants responded with all fingers using the ‘A’, ‘S’, ‘D’, ‘F’ keys with the left hand and the ‘J’, ‘K’, ‘L’, ‘;’ keys with the right hand. The

responding hand coded the perceptual response, and the responding finger coded perceptual confidence (not analysed here). An equal number of trials were acquired for each of the two possible key mappings for each task. Key mapping was counterbalanced between sessions and participants. Participants were given at least 48 practice trials to get accustomed to a change in key mapping. They were instructed to correct mistakes if they noticed them.

Behavioural exclusion criteria. All trials with multiple responses on the current or previous trial were discarded from all analyses ($M = 201.35$, *range*: 39, 374). In addition, trials with response times faster than 100 ms after last stimulus onset were excluded ($M = 0.16$, *range*: 0, 2). This led to the following trial distribution: $M = 3294$, *range*: 3068, 3805.

Experimental setup

Testing took place in a darkened room with additional light-shielding the stimulus PC and participant. The presentation screen was a 19'' cathode ray tube (CRT) display with a refresh rate of 120 Hz and a resolution of 1024 x 768 pixels. Beeps were presented via EARtone 3A Insert Earphones (Aearo Company Auditory Systems, 1997). Participants' gaze position was tracked with an EyeLink 1000 Plus eye-tracker to be able to exclude trials where central fixation was not maintained (SR Research Ltd., 2014).

Staircase procedure to titrate perceptual threshold

SOA values were chosen to be as close to the point at which participants responded "see 1" or "see 2" equally often for 2-flash and 2-flash, 1-beep presentations as possible. To this end two staircases were run in parallel for each condition, one with starting values being the psychometric function inflection points obtained from the first two yes-no runs, and one with the inflection points obtained from the last two yes-no runs (exp. 4-6; see chapter 2). SOAs increased by one screen refresh interval after an incorrect response and decreased after a correct response. Each staircase contained 24 trials per condition. Staircases were considered to have

converged after 10 reversals, after which SOA did not change further. SOAs were not allowed to exceed values above or below 5 refresh intervals from the starting value. The average over the last 5 reversals from both staircases determined the final estimate. With regard to the 1-flash, 2-beep experiment this procedure was only followed when the line connecting the perceptual accuracy values at different SOAs was monotonically increasing. Otherwise a guess was taken based on the data from previous yes-no runs. 2-flash, 2-beep SOA was set equal to 1-flash, 2-beep SOA. In some cases, when participants showed markedly unequal response distributions (based on the experimenter's judgment) during subsequent task performance, the task was restarted with an adjustment of 1 screen refresh interval to the relevant conditions.

Behavioural analyses

Signal detection model. For each condition (0-2 beeps), d' (sensitivity) and criterion (C) were estimated under the assumptions of a two equal-variance Gaussian signal detection model (Macmillan & Creelman, 2005). Participants' internal 1-flash, X-beep representations are drawn from the noise distributions and 2-flash, X-beep representations are drawn from the signal distributions (Figure 3.2d). Sensitivity d' is the distance between a given pair of signal and noise distribution means, which can be estimated as follows:

$$(1) \quad d' = z(HR) - z(FAR),$$

where HR denotes hit rate (i.e. proportion of “see 2” responses when 2 flashes were presented) and FAR denotes false alarm rate (i.e. proportion of “see 2” responses when 1 flash was presented). The criterion (bias) is computed as follows:

$$(2) \quad C = -0.5(z(HR) + z(FAR)),$$

where $C > 0$ marks a bias toward “see 1”. A small constant of 0.1 was added to both hits and false alarms to avoid infinities in the calculation of d' and criterion.

Metacognitive efficiency. The level 1 perceptual signal detection model can be extended by adding level 2 criteria depending on participants' confidence judgments. These criteria are determined by type 2 hit rates and false alarm rates on “see 2” response trials and type 2 correct rejections and misses on “see 1” response trials. Type 2 hit rate is the proportion of “confident” responses given that participants reported the signal correctly ($HR2 = P(\text{confident} \mid \text{“see 2” \& 2 flashes})$), whereas type 2 false alarm rate is the proportion of “confident” responses when participants reported the signal incorrectly ($FAR2 = P(\text{confident} \mid \text{“see 2” \& 1 flash})$). The proportions of type 2 correct rejections and misses can correspondingly be computed for trials where participants reported “see 1”. When more than two confidence ratings are given the same calculations apply, but for e.g. $FAR2_{\text{conf}=4} = P(\text{confidence} = 4 \mid \text{“see 2” \& 1 flash})$, effectively adding more criteria to divide up the two equal variance Gaussian signal detection model. Because we are splitting partial Gaussian probability density functions with type 2 criteria, the calculation of type 2 d' cannot be performed with the same equations that we used for the type 1 perceptual decisions (or at least it would lose straightforward interpretability). Instead, we took advantage of the meta- d' model developed and implemented by Maniscalco and Lau (Maniscalco & Lau, 2012). Specifically, we used maximum likelihood estimation to find the value of d' (coined meta d') that would be expected to have given rise to the type 2 confidence criteria if participants were optimal metacognitive observers of their type 1 perceptual evidence. From this metacognitive efficiency can be computed as follows:

$$(3) \quad \text{Metacognitive efficiency} = \frac{\text{meta } d'}{d'}.$$

As such metacognitive efficiency should range between 0 (no access to level 1 perceptual evidence) to 1 (complete access to level 1 perceptual evidence).

Statistical comparisons. Standard parametric statistics (repeated measures ANOVAs and paired t-tests) were used if assumptions were met, as tested by visual inspection of quantile-quantile

plots, residual plots and Shapiro Wilk tests (normality rejected at the 0.001 level). The Greenhouse-Geisser correction was used when Mauchly's test of sphericity was significant at the 0.05 level. Post-hoc tests were corrected with the Bonferroni-Holm method to account for type 1 error inflation due to multiple comparisons (Holm, 1979).

EEG preprocessing

EEG data were recorded with 64 active electrodes (actiCAP, Brain Products Ltd., 2014) arranged in an extended 10-20 montage (ground: AFz, reference: FCz). Data were amplified with two 32 channel BrainAmp DC amplifiers (Brain Products Ltd., 2014) and digitized at 1000 Hz with a high-pass filter of 0.1 Hz. EEG preprocessing and analyses were carried out with the FieldTrip toolbox (Oostenveld et al., 2011) for Matlab and custom written Matlab code. Data were low-pass (99 Hz) and notch filtered (48-52 Hz). Noisy channels and time epochs, including time periods encompassing saccades towards the stimulus, were identified visually and discarded. Independent component analysis (ICA) was applied to the cleaned EEG data and components unrelated to brain activity (e.g. blinks, heart beat) were rejected before back-transforming components to channel space (on average 1.4 components were rejected per recording, range: 1, 5). Signals at positions of excluded channels were estimated with spline interpolation using the standard 10-10 layout and default neighbour positions defined by the FieldTrip toolbox (On average 2.1 channels were interpolated per recording, range: 0, 10). Next, data were segmented into trials for stimulus and response locked analyses (locked to first stimulus: -1.2 to 3.5 s; locked to response: -3 to 0.5 s), and down-sampled to 256 Hz. Trial segments without marked artefacts were demeaned relative to a 0.2 s pre-stimulus baseline, re-referenced to an average over channels and used for further analysis.

Data selection. To avoid confounds due to response related activity (or stimulus related activity when locking time courses to response) all trials with response times < 500 ms relative to second

stimulus onset were discarded (number of trials discarded: $M = 160$; *range*: 3, 969). Furthermore, trial numbers were matched between perceptual outcomes by randomly excluding trials from the larger group, at the same time ensuring that each hand was used equally often per response (trial distributions for stimulus locked data, 0-beep: $M = 231$; *range*: 60, 408; 1-beep: $M = 211.2$, *range*: 68, 380; 2-beep: $M = 166$, *range*: 4, 320; trial distributions for response locked data, 0-beep: $M = 240.4$; *range*: 72, 420; 1-beep: $M = 221.2$, *range*: 68, 384; 2-beep: $M = 175.6$, *range*: 4, 324). The latter is important to prevent motor related activity instead of perceptual or decisional activity to lead to putative condition differences. Only participant data with more than 80 remaining trials per condition (40 per perceptual response) were analysed.

EEG univariate analyses

First, we investigated the neural correlates of perceptual outcome (“see 1” or “see 2”) by computing event related potentials (ERPs) for conditions 2-flashes, 2-flashes & 1-beep, and 1-flash & 2-beeps locked to stimulus 1, stimulus 2 and response. Condition differences were analysed with paired t-tests (two-sided) at each time point over all electrodes of interest. Cluster based permutation statistics (based on summed t-values within clusters, $N_{\text{permutations}} = 2000$) were used to account for multiple comparison problems (Maris & Oostenveld, 2007). Both cluster threshold level and statistical alpha level were set to 0.05.

Functional localiser. ERP analyses were conducted on temporal and spatial periods of interest defined using a functional localizer contrasting physical 1-flash or 2-flash presentations given a “see 1” response (there were not many “see 2” responses on 1-flash trials). Trial numbers were balanced between flash presentations (1 or 2) and response hand combinations (left = 1, left = 2). All channels and time points between 0 and 400 ms were analysed. Any channel being part of a significant cluster, and the time period between the earliest and latest time point over

all significant clusters defined the region of interest over which to statistically contrast perceptual outcomes.

Regressing out response time. To test whether decoded outcome (e.g. respond “see 1” or “see 2”) is driven by different response speeds between groups, we performed linear regression:

$$(4) \quad EEG_{i,j,k} = b0_{j,k} + b1_{j,k}RT_i + e_{i,j,k},$$

where EEG_i denotes amplitude for a given trial i , $b0$ is the intercept, $b1$ the slope coefficient for response time (RT) and e denotes the error term. This was done separately for each time point (j) and channel (k). Residuals were used for further analysis after adding the intercept to preserve the ERPs’ temporal structure:

$$(5) \quad EEG_corrected_{i,j,k} = b0_{j,k} + e_{i,j,k}.$$

EEG multivariate analyses

Decoding perceptual outcome. We used a linear support vector machine (SVM) classifier (with box constraint = 1) to decode perceptual outcome as implemented in the CosmoMVPa toolbox for Matlab (Oosterhof, Connolly, & Haxby, 2016). First, data were low-pass filtered at 48 Hz and down-sampled to 64 Hz. Next, for every participant and condition (2-flash, 2-flash & 1-beep, and 1-flash & 2-beep) trials were randomly grouped into 5 folds, ensuring the same number of trials within folds for each perceptual response and response key mapping by discarding trials at random (note that 1-flash, 1-flash & 1-beep and 2-flash & 2-beep conditions were not used, as the distribution of perceptual reports was too skewed). To reduce the number of features, we performed principal component analysis (PCA) and transformed the EEG time courses into linearly uncorrelated components. We retained the components with the largest eigenvalues that explained at least 90% of the variance in the data. Next, 50% of trials belonging to a perceptual outcome within a fold from each response key mapping were randomly selected and averaged. This process was repeated 100 times and all trial averages were retained as

samples for the classifier. Decoding accuracies were obtained for each time point via leave-one-fold out cross-validation and statistically compared against chance (0.5) with one-sample t-tests (one-sided). Cluster-based permutation statistics were used to control for inflation of type 1 error rate due to multiple comparisons (Maris & Oostenveld, 2007). A complementary analysis was performed for the entire time window of interest. In this case a second PCA was performed to reduce the number of temporal features retaining the components accounting for 50% of the variance in the data.

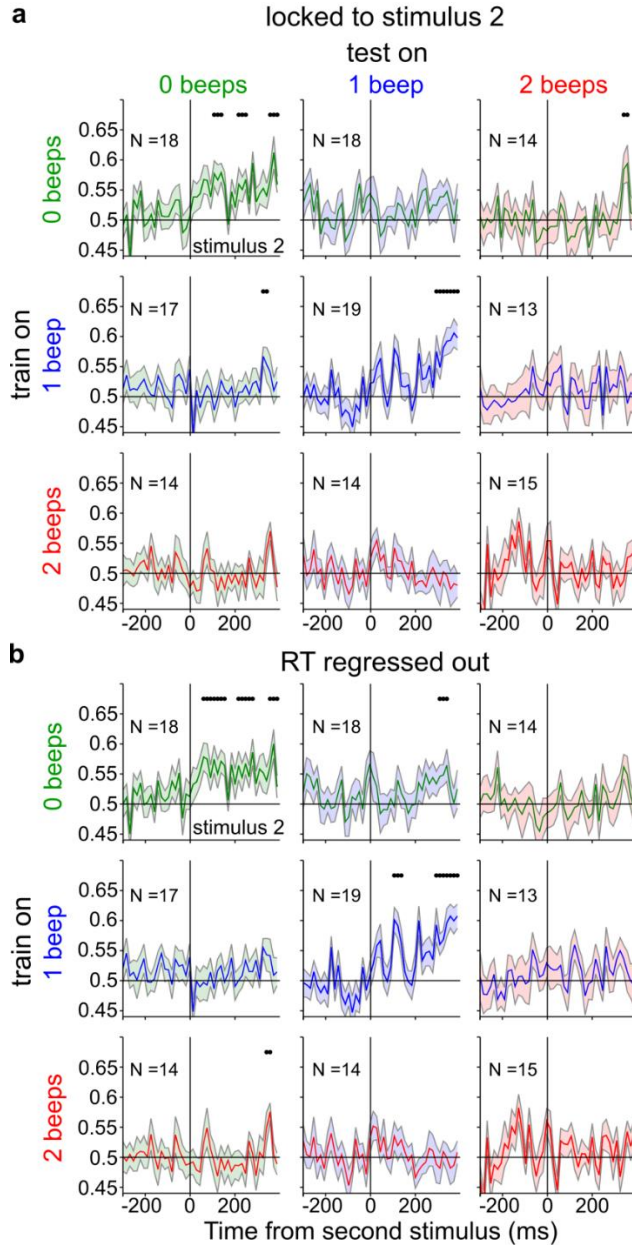
Generalizing across conditions. In addition to decoding perceptual reports within conditions, we trained on one condition (e.g. 2-flash) and generalized to the other conditions (2-flash, 1-beep; 1-flash, 2-beep). To this end the number of trials used to create averages in the test conditions was kept identical to the number used in the training condition to allow for direct comparisons of decoding accuracies within training condition generalizations.

Regressing out response time. Similar to univariate analyses, RT was regressed out from EEG amplitude to assess perceptual differences irrespective of response speed. In this case, the regression was performed only on the training set over both outcomes (e.g. “see 1” and “see 2”), and beta coefficients were used to apply the same correction to the corresponding test set. Note that a regression over the full dataset would create dependencies between training and test data (Snoek et al., 2019).

Bayes factors. To quantify evidence for (or against) the null hypothesis, we computed two-sided Bayes factors (BFs) for dependent samples t-tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009). A BF is a ratio of evidence between the alternative hypothesis (H_a) and the null hypothesis (H_0). For example, a value of 3:1 means that H_a is three times more likely than H_0 . As a general rule BFs < 3.2 or > 0.32 are considered inconclusive (Jeffreys, 1998). We used a standard JZS prior consisting of a Jeffreys prior on variance for H_a and H_0 , a Cauchy prior on

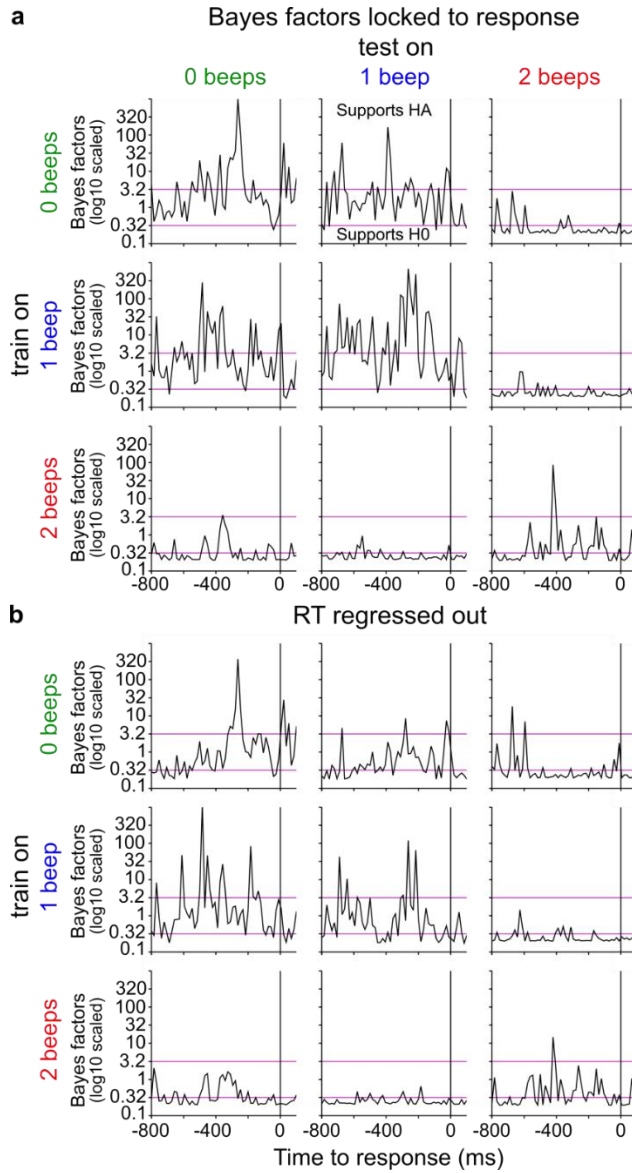
effect size for H_a and a point-zero prior on effect size for H_0 (Rouder et al., 2009).

Supplementary materials



Supplementary Figure 3.1. Decoding time courses locked to second stimulus onset.

Ordinates show proportion correctly decoded labels (“see 1” or “see 2”). Chance level is 0.5. **a-b**, Within and across condition decoding accuracy time courses locked to second stimulus onset. **b**, Response times were regressed out before training the classifier. Samples with decoding accuracies significantly above chance are marked with black dots. Error bars denote ± 1 SEM. RT, response time.



Supplementary Figure 3.2. Bayes factors for decoding accuracies locked to response.

a-b, Two-sided Bayes factors (BFs) were computed for the within and across condition decoding accuracy time courses locked to response to quantify evidence for the point-Null hypothesis $H_0 = 0.5$. BFs > 3.2 provide substantial evidence for H_A , while BFs < 0.32 provide substantial evidence for H_0 (marked by purple lines). **b**, Same as in **a**, but response times were regressed out before training the classifier. Note that cross-condition generalizations between 0 and 1-beep with the 2-beep condition mostly show support for H_0 , suggesting that internal representations of one versus two-flash perceptual choices are different between conditions. RT, response time.

Chapter 4 - What is the role of posterior alpha frequency in visual temporal resolution?

The following chapter is going to be submitted for scientific publication in near identical form and is the result of collaboration between Uta Noppeney and Steffen Bürgers. U.N. conceived the study. S.B. and U.N. designed the experiments. S.B. collected the data. S.B. analysed the data. S.B. and U.N. interpreted the data. S.B. wrote the manuscript. U.N. provided direct supervision at all stages of the process.

Abstract

Posterior alpha oscillations (~6-14 Hz) have been suggested as an electrophysiological substrate of temporal resolution in vision (VanRullen, 2018; White, 1963). If true, this would establish alpha oscillations as an excellent biomarker for studying and understanding the discreteness and rhythmicity of visual perception. Yet, despite several pieces of evidence pointing to a relationship between alpha frequency and visual temporal integration, the nature of this link remains contentious due to i. variability in paradigms and perceptual contexts, ii. conflicting results, and iii. confounding of perceptual sensitivity with perceptual and decisional biases. To clarify whether alpha frequency might indeed be a fundamental index of visual temporal resolution, we combined electroencephalography and psychophysics to systematically replicate, combine and extend previous work. Importantly, we dissociated perceptual sensitivity from perceptual and response bias using signal detection theory and two-interval forced choice designs. Results provide robust evidence that *neither* inter-trial *nor* inter-subject variability in alpha frequency influence temporal resolution in two-flash discrimination tasks. Null-findings are supported by Bayes factors. These results challenge current notions of a simple, generic relationship between posterior alpha oscillation frequency and visual temporal resolution in two-flash integration or segregation.

Introduction

Alpha oscillations (~6-14 Hz) are elicited by fluctuations in postsynaptic activity of large neural populations and can prominently be observed in the raw electroencephalogram (EEG), especially over posterior brain regions (Berger, 1929). Given such a strong rhythmic signal over parieto-occipital cortex, investigators of the temporal resolution of visual perception early on conjectured that alpha cycles mark discrete perceptual windows in time (VanRullen, 2018; White, 1963). According to this alpha temporal resolution hypothesis (α TRH), if two visual stimuli fall within the same alpha cycle they will be integrated and perceived as one temporal event. Conversely, when two stimuli fall in different alpha cycles they will be segregated and perceived as two distinct events (Hirsh & Sherrick Jr., 1961b; Kristofferson, 1967; Samaha & Postle, 2015; VanRullen, 2018; Varela et al., 1981; White, 1963) (Figure. 4.1a). Recently, research employing unisensory visual and audio-visual stimulation provided novel evidence in support of the α TRH (Cecere et al., 2015; Keil & Senkowski, 2017; Ronconi, Busch, & Melcher, 2018; Samaha & Postle, 2015; Wutz, Melcher, & Samaha, 2018), even suggesting that alpha waves drive visual perceptual cycles causally (Cecere et al., 2015). Nevertheless, aggregating results to form solid support for the α TRH remains challenging.

A major caveat is that substantial variability exists in experimental paradigms and consequently perceptual outcomes, which may or may not be subject to the same underlying temporal framing mechanism. At present, there are multiple candidate sources of alpha activity with at least one thalamo-cortical (Hughes et al., 2004; Lőrincz, Kékesi, Juhász, Crunelli, & Hughes, 2009; Vijayan & Kopell, 2012) and one cortico-cortical alpha component (Bollimunta, Mo, Schroeder, & Ding, 2011; Lopes da Silva et al., 1973; Naruse et al., 2010), which might plausibly interact (Sigala, Haufe, Roy, Dinse, & Ritter, 2014). Moreover, even though evidence in favour of electrophysiological oscillations playing a role in perceptual cycles is mounting,

the field is steering away from a pure focus on alpha oscillations and embracing the view that different perceptual phenomena map onto different frequencies and underlying brain activity (VanRullen, 2016, 2018). Appropriately, studies investigating the α TRH differ in the frequency ranges and scalp locations where significant effects were observed. It is therefore important to show a link between alpha frequency and visual temporal resolution in multiple contexts, while holding as many experimental and analysis parameters constant as possible.

Notwithstanding the unknown physiological generators of posterior alpha oscillations, alpha waves are ascribed a role in modulating (or driving) visual information processing, with both alpha power and phase marking the level of cortical excitability (Hanslmayr, Gross, Klimesch, & Shapiro, 2011; Iemi et al., 2017; Jensen, Bonnefond, & VanRullen, 2012; Jensen, Gips, Bergmann, & Bonnefond, 2014; Klimesch, Sauseng, & Hanslmayr, 2007). The proposed link between alpha phase and cortical excitability is predicted by the α TRH (Figure. 4.1b). However, pre-stimulus phase differences in studies where multiple events are to be dissociated in time are sparse (Gulbinaite, İlhan, & VanRullen, 2017; VanRullen, 2018), presumably because of a phase reset initiated by the first stimulus (Mercier et al., 2013; Sauseng et al., 2007).

Another challenge for the α TRH is that findings are partially conflicting, with one study showing faster alpha oscillations for larger temporal integration windows (Ronconi & Melcher, 2017). Furthermore, there were several reports of null results (Baumgarten, Schnitzler, & Lange, 2017; Gulbinaite et al., 2017; Keil & Senkowski, 2017). In addition, some findings suggest alpha frequency is a trait marker of perceptual resolution (Cecere et al., 2015; Keil & Senkowski, 2017; Ronconi & Melcher, 2017; Samaha & Postle, 2015), while others suggest that it fluctuates as a state marker within individuals over time (Ronconi et al., 2018; Ronconi & Melcher, 2017; Samaha & Postle, 2015; Wutz et al., 2018).

Finally, previous work did not unequivocally dissociate perceptual sensitivity, perceptual bias and response bias. This is problematic, as a difference in response bias might mask as an effect of sensitivity and vice versa. Perceptual resolution, the rate at which participants resolve visual events in time, is associated with sensitivity, not bias.

Given these controversies, we tried to clarify the role of posterior alpha frequency in visual temporal resolution by focusing on simple two-flash integration or segregation for both unisensory visual, as well as audio-visual contexts of the sound induced fission (Shams et al., 2002) and fusion illusion (Andersen et al., 2004). Importantly, we dissociated perceptual sensitivity and bias using signal detection theory and two-interval forced choice designs. If alpha waves in visual cortex really mark the temporal resolution of visual perception, we would expect to see higher sensitivity for faster alpha waves in any two-flash versus one-flash perceptual context both within and between participants.

Results

To elucidate the role of alpha frequency in visual temporal resolution we conducted a multi-day psychophysics EEG study that included 3 x 3 experiments (Figure. 4.1). Twenty human participants were presented with one or two flashes i. in a unisensory context, ii. with one sound, and iii. with two sounds (columns in Figure. 4.1). Each of those three perceptual stimulation contexts was presented in three designs: i. two-interval forced choice (2IFC) with variable stimulus onset asynchrony (SOA), ii. yes-no with multiple SOAs and iii. yes-no with single SOA at perceptual threshold (rows in Figure. 4.1). The 2IFC design is crucial to control for perceptual and decisional biases. Only yes-no experiments included the acquisition of EEG data during task performance, as pre-stimulus EEG activity in the 2IFC design cannot unambiguously be assigned to either the first or second interval (for an overview of a participant's testing schedule see Supplementary Figure 4.1).

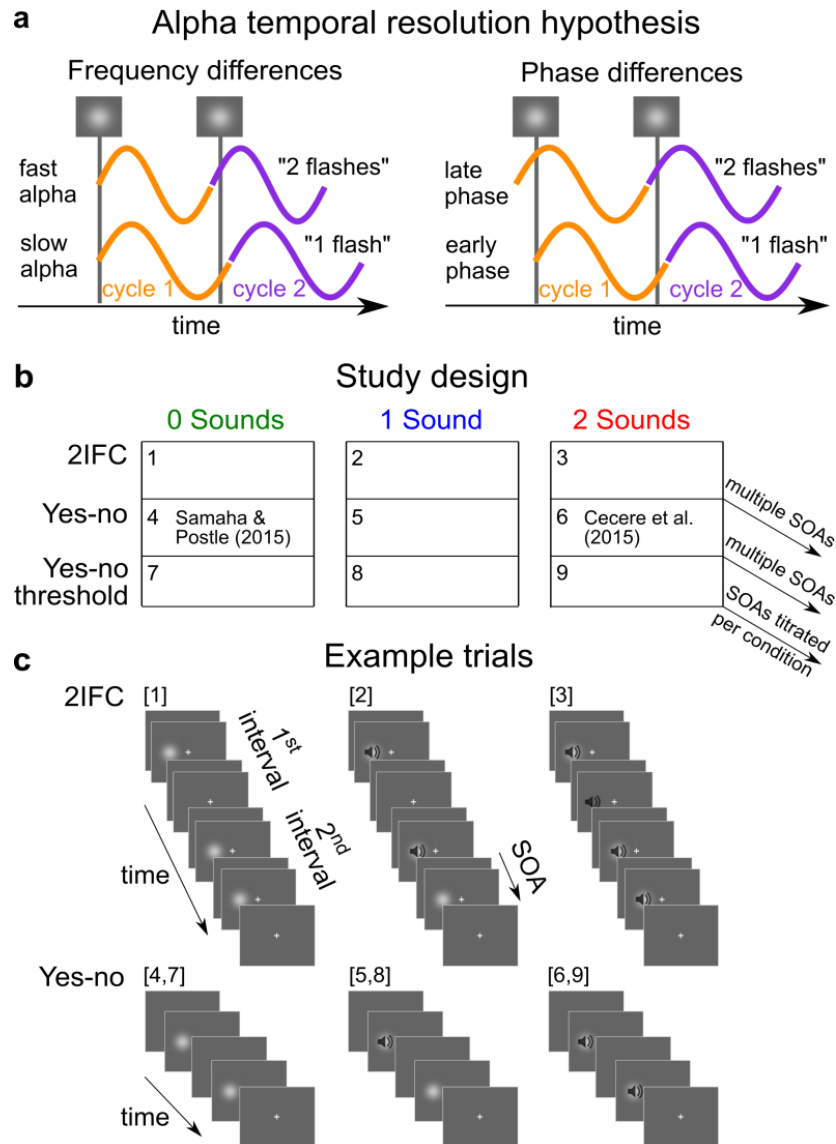


Figure. 4.1. Hypothesis, study design and trial examples. *a*, Illustration of the α TRH. Two identical visual stimuli are closely presented in time. If they fall within the same alpha cycle, they are temporally integrated and perceived as a single flash. Conversely when the second stimulus falls in a new alpha cycle the two stimuli are perceived as separate temporal events. In this framework both frequency and phase should influence two-flash discriminability. *b*, Illustration of the 9 experiments included in this study with paradigm coded by rows and number of sounds coded by columns. EEG was acquired only during yes-no tasks (exp. 4-9). Two yes-no designs were deployed, one with multiple SOAs to allow for the estimation of psychometric functions and one with single SOA titrated to yield 50% 1-flash and 50%

2-flash responses per participant and experiment (only for the following stimulus combinations: 2 flashes; 2 flashes, 1 sound; 1 flash, 2 sounds). c, Example trials for 2IFC (exp. 1-3) and yes-no tasks (exp. 4-9). In the 2IFC task the first or second interval contained 2 flashes and the other interval contained 1 flash (examples only show 2 flashes in second interval). Participants reported which interval contained 2 flashes. In the yes-no task either one or two flashes were presented and participants reported how many flashes they perceived (only 2-flash examples shown).

State alpha frequency does not influence visual temporal resolution

First, we asked whether spontaneous fluctuations in pre-stimulus alpha frequency influence the temporal resolution of subsequent visual perception, as previously reported (Ronconi et al., 2018; Ronconi & Melcher, 2017; Samaha & Postle, 2015; Wutz et al., 2018). To this end, we ordered trials by pre-stimulus frequency and computed perceptual sensitivity (d') for the first tercile (low frequency) and third tercile (high frequency). If the α TRH is correct, we would expect larger sensitivity (higher temporal resolution) for trials with high alpha frequency (third tercile > first tercile). Two-sided paired t-tests were computed for the 600-100 ms window prior to first stimulus onset in steps of 40 ms. To correct for multiple comparisons, cluster-based permutation tests were performed (Maris & Oostenveld, 2007). Bayes factors (BFs) quantify evidence for or against H_0 , with $BF < 0.32$ or > 3.2 substantially supporting H_0 or H_a , respectively (Jeffreys, 1998).

Sensor level analysis. Sensor level results were computed for the average over channels O2, PO4 and PO8, as the visual stimulus is expected to elicit the strongest cortical response in right visual cortex. Results show no significant evidence for faster pre-stimulus alpha oscillations preceding responses that indicate a higher perceptual resolution (Figure. 4.2a). Moreover, BFs consistently support H_0 , (exp. 4, 8), a small trend in the opposite direction of what the α TRH predicts (third tercile < first tercile; exp. 5, 6), or a small trend in line with the α TRH (exp. 7,

9). Importantly, no trend is internally replicated in the correspondingly second yes-no experiment (e.g. 4 and 7), further arguing in favour of H_0 .

As the α TRH is not unambiguously supported by the literature, we tested the possibility that a subset of participants showed the expected frequency effects reliably over experiments, while the remaining participants either show no effect or an effect in the opposite direction. To investigate this, we collapsed sensitivity over time and computed the difference between the first and third tercile ($\Delta d'$). Following this, we correlated participants' $\Delta d'$ scores between the yes-no and yes-no threshold experiments. Correlation coefficients are all negative ($r_0(20) = -0.22$, $p = 0.35$; $r_1(20) = -0.11$, $p = 0.66$; $r_2(20) = -0.36$, $p = 0.12$), indicating lower than chance consistency of the relationship between alpha frequency and temporal resolution over experiments (Figure. 4.2b).

Source analysis. Volume conduction leads to a spread of alpha source activity over a large region of electrodes at the scalp surface. Therefore, we tried to source localize occipital and parietal alpha components, which may subserve different cortical processes (Gulbinaite et al., 2017; Haegens, Cousijn, Wallis, Harrison, & Nobre, 2014). Source estimates were computed using dipole fitting of independent components filtered in the alpha band (Delorme & Makeig, 2004; Delorme, Palmer, Onton, Oostenveld, & Makeig, 2012). Both occipital and parietal source level results are in line with findings at the sensor level, showing no significant effect of alpha frequency on perceptual sensitivity (Supplementary Figure 4.2).

Alpha frequency is not related to bias. If perceptual sensitivity is not linked to the speed of alpha oscillations, can previously reported findings be explained by bias? Our results indicate that this is not the case. Similar to sensitivity, criterion does not significantly differ between high and low alpha frequency bins for any of the yes-no experiments (exp. 4-9; Supplementary Figure 4.3).

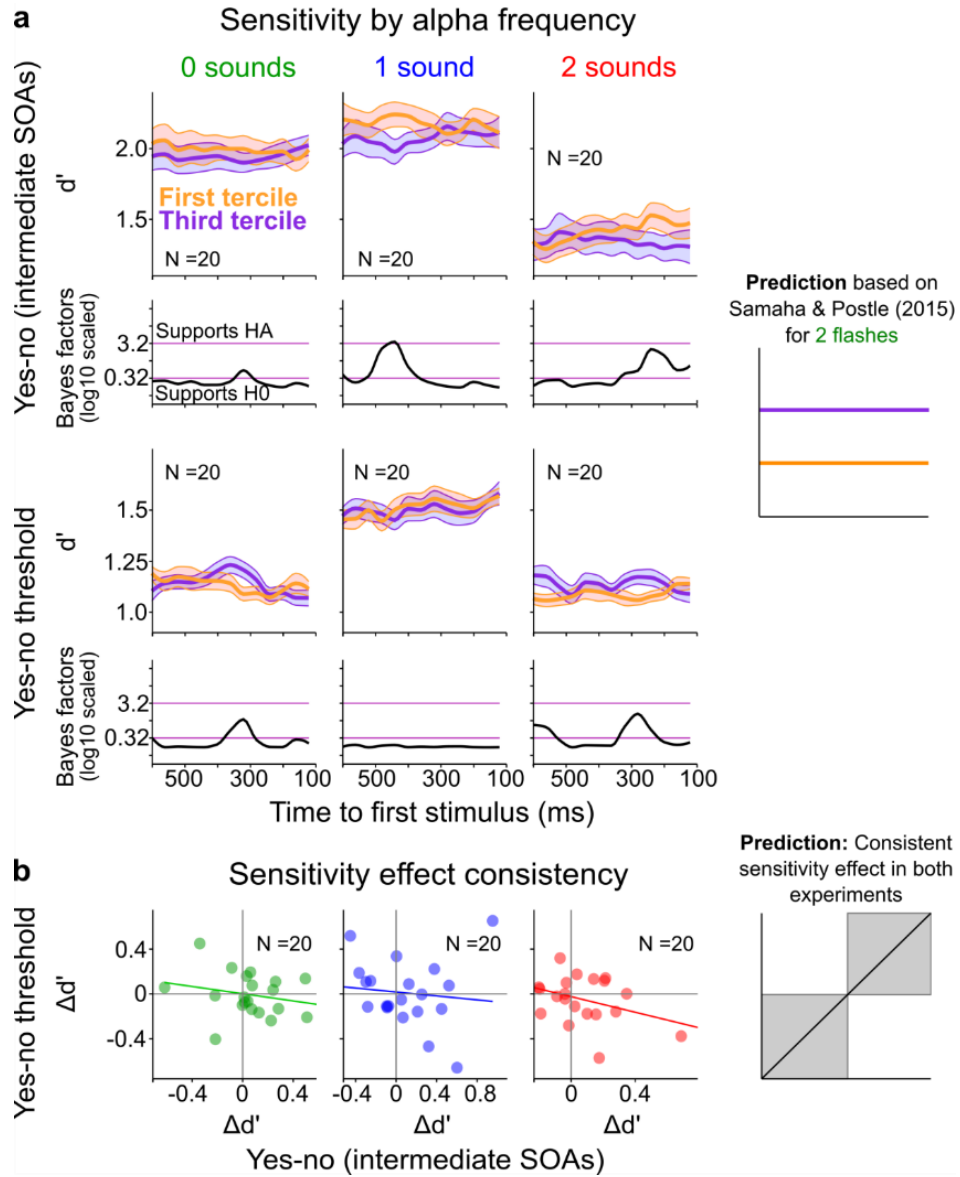


Figure 4.2. Perceptual sensitivity by pre-stimulus alpha frequency. *a*, Perceptual sensitivity does not differ between low (first tercile) and high (third tercile) pre-stimulus alpha frequency trials for all yes-no experiments (exp. 4-9). The α TRH predicts higher sensitivity for high alpha frequency. Significance of paired *t*-tests (two-sided) is cluster corrected (all tests *n. s.*). Error bars denote ± 1 within subjects SEM. Pink lines mark Jeffrey's thresholds of substantial evidence for H_0 (0.32) or H_a (3.2) (Jeffreys, 1998). *b*, For each experiment $\Delta d'$ ($d'_{bin1} - d'_{bin3}$) indicates support for the α TRH ($\Delta d' < 0$), or for slower

alpha oscillations to lead to higher sensitivity ($\Delta d' > 0$). Participants showed no consistent relationship between alpha frequency and perceptual sensitivity over experiments (negative correlation).

State alpha power fluctuations do not mask alpha frequency effects. Spontaneous power fluctuations could potentially mask frequency related effects (Nelli, Itthipuripat, Srinivasan, & Serences, 2017). We therefore repeated the sensor analyses for alpha power instead of frequency. There was one significant cluster showing higher perceptual sensitivity for high compared to low pre-stimulus alpha power in the 2 sound yes-no threshold experiment (exp. 9). State frequency related results for this experiment should therefore be interpreted with caution. All other experiments show no significant clusters for alpha power and BFs support H_0 or small trends in either direction (Supplementary Figure 4.4, left).

State alpha power does not generally index cortical excitability

Alpha power has previously been shown to index baseline excitability (Hanslmayr et al., 2011; Iemi et al., 2017; Jensen et al., 2012, 2014; Klimesch et al., 2007), and low alpha power was associated with more 2-flash responses in the tactile induced fusion and fission illusion contexts (J. Lange et al., 2013). From a signal detection point of view this corresponds to a shift in criterion (Iemi et al., 2017). Here, we found a significant pre-stimulus power effect on criterion for the sound induced fusion illusion, but only in the yes-no threshold experiment (exp. 8), and not for unisensory flash perception or the sound induced fission illusion (exp. 4-7, 9; Supplementary Figure 4.4, right).

Alpha phase might play a role in visual temporal resolution

In addition to frequency, the α TRH predicts a modulation of temporal resolution by alpha phase (Figure. 4.1b). If there is no complete phase reset after the first stimulus this should already be visible in the pre-stimulus window. To test this prediction, we ordered trials by pre-stimulus alpha phase and aggregated them into three equally sized bins, where bin 1 and 3 were

maximally opposite each other on the unit circle (e.g. bin 1 contains trials close to 0 and bin 3 contains trials close to π). Next, we computed d' for both bins. As we were agnostic about which phase would influence perceptual resolution for each participant, we repeated this procedure for 17 different phase splits, rotating in steps of 10° around the unit circle. For each participant we chose the phase split with maximal absolute difference in d' between phase bins 1 and 3. To test whether this difference was statistically significant, we randomly shuffled the phase relationships between trials 1000 times, and repeated the process of finding the maximum d' difference for each iteration. From this permutation distribution we selected the median and computed paired t-tests (one-sided) for pre-stimulus time points between 600 and 100 ms prior to first stimulus onset in steps of 40 ms. Cluster correction was used to account for multiple comparison problems (Maris & Oostenveld, 2007).

There was one cluster showing significant phase opposition when 1 sound was presented (exp. 5; Supplementary Figure 4.5, left). However, this effect was not replicated in the yes-no threshold task (exp. 8), where BFs consistently and substantially support H_0 (BFs < 0.32). However, in all other experiments (exp. 4-7, exp. 9), BFs give mixed results, largely falling in the inconclusive range ($0.32 < \text{BF} < 3.2$). Regarding the relationship between pre-stimulus alpha phase and criterion, a similar pattern emerged. Two experiments showed significant clusters of phase opposition with respect to criterion (exp. 4, 6), but these results were not replicated in the yes-no threshold experiments and BFs frequently fall in the inconclusive range (Supplementary Figure 4.5, right). Given these results it is difficult to conclude that phase does not influence perceptual sensitivity or bias. Instead, our study design might simply not be suited to test pre-stimulus alpha phase effects, which might be strongly attenuated by a phase reset initiated by the first flash/beep event (Mercier et al., 2013; Sauseng et al., 2007).

Trait alpha frequency does not index visual temporal resolution

In addition to spontaneous fluctuations in alpha frequency, trait peak frequency differences have been hypothesized to mark differences in visual temporal resolution between individuals (Cecere et al., 2015; Keil & Senkowski, 2017; Ronconi & Melcher, 2017; Samaha & Postle, 2015). To test this hypothesis, we correlated individual alpha peak frequency with the lengths of participants' temporal integration windows. Individual alpha peak frequency was estimated over all recording sessions using the peak fitting algorithm developed and implemented by Corcoran and colleagues (Corcoran, Alday, Schlesewsky, & Bornkessel-Schlesewsky, 2018). Separate peak estimates were obtained for eyes-closed recordings, and task related alpha activity prior to first stimulus onset. Perceptual thresholds were defined as the inflection points of psychometric functions (Figure. 4.3, exp. 1-6) or obtained from a staircase procedure (exp. 7-9). For the yes-no task with multiple SOAs (exp. 4-6) we pooled over 1-flash and 2-flash trials for each of the 0-2 sound conditions in order to mitigate the influence of perceptual and response bias. Statistical comparisons between perceptual threshold and trait alpha frequency were carried out using Pearson's correlation coefficient and corresponding BF (Wetzels & Wagenmakers, 2012).

We found no significant relationship between individual alpha peak frequency and perceptual threshold for neither eyes-closed, nor eyes-open recordings over parieto-occipital channels (Figure. 4.3b, Table 4.1). BFs consistently showed substantial support for H_0 (except for exp. 1, where evidence falls in the inconclusive range. Note, however, that the correlation coefficient is positive, and therefore opposite of what the α TRH predicts). Similarly, occipital and parietal source estimates of individual alpha peak frequency were not related to individual perceptual resolution (Supplementary Table 1).

Trait alpha power does not mask alpha frequency effects. As power can potentially mask frequency related effects (Nelli et al., 2017), we repeated the between subject analysis for alpha power. There was no significant correlation between eyes-closed or eyes-open alpha power and perceptual window length (Supplementary Table 2). It is therefore unlikely that power effects mask a potential link between alpha frequency and visual temporal resolution between participants.

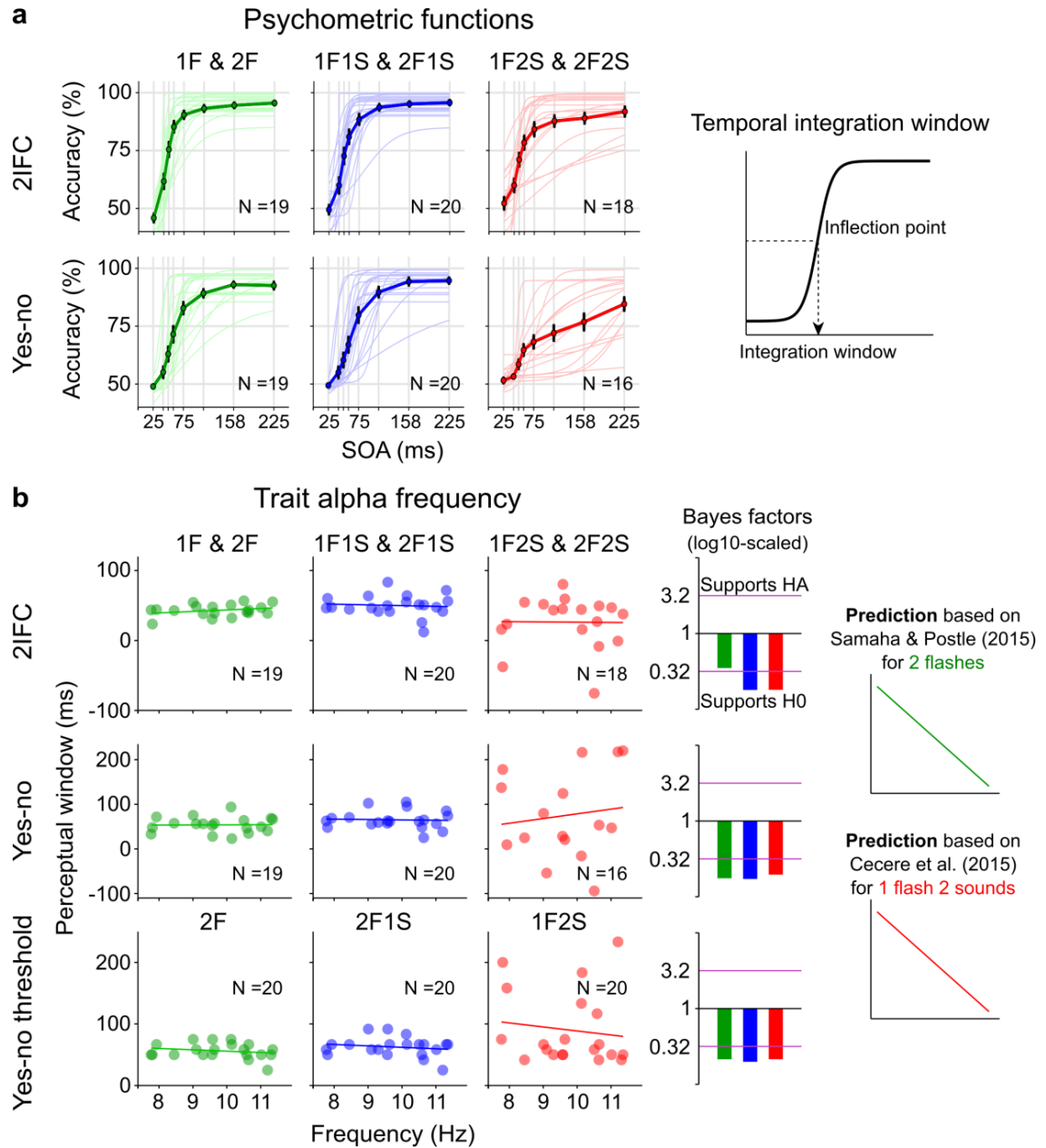


Figure. 4.3. Psychometric functions and trait alpha frequency results. *a*, Grand average response accuracies (connected with bold lines) and single subject psychometric functions (thin lines) for 2IFC (exp. 1-3, row 1) and yes-no (exp. 4-6, row 2) tasks. Error bars denote ± 1 SEM. Psychometric function inflection points quantify temporal integration windows for experiments 1-6. Missing data in rows 1 and 2 are due to poor psychometric curve fits. *b*, Trait alpha frequency correlations with perceptual windows. Row 3 shows alpha frequency against the SOA estimated for the yes-no threshold paradigm at which participants responded “see 2” approximately 50% of the time. BF thresholds of substantial

evidence for H_0 (0.32) or H_a (3.2) are marked in pink. BFs consistently support H_0 . The α TRH, based on previous studies, predicts a negative correlation between alpha frequency and size of the temporal integration window.

Table 4.1. Trait alpha peak frequency versus perceptual window length.

Threshold definition	Eyes-closed sensor level			Pre-stimulus sensor level		
	N	r	BF	N	r	BF
2IFC						
1F & 2F	19	-0.36	0.55	19	-0.28	0.35
1F1S & 2F1S	20	0.13	0.20*	20	0.08	0.18*
1F2S & 2F2S	18	-0.21	0.26*	18	-0.03	0.18*
Yes-no pooled						
1F & 2F	19	-0.03	0.18*	19	-0.02	0.18*
1F1S & 2F1S	20	0.08	0.18*	20	0.03	0.17*
1F2S & 2F2S	16	0.07	0.20*	16	-0.07	0.20*
Staircase SOA						
2F	20	0.07	0.18*	20	0.16	0.21*
2F1S	20	0.00	0.17*	20	0.13	0.20*
1F2S	20	0.32	0.43	20	0.16	0.21*
2F2S	20	0.34	0.43	20	0.16	0.21*

N, number of participants; *r*, Pearson's correlation coefficient; *BF*, Bayes factor. Asterisks mark substantial evidence for the Null-hypothesis.

Control analyses

To establish that the reported null findings cannot trivially be explained by human or technical errors during data acquisition or analysis we performed a control analysis on event related potentials (ERPs). Specifically, we conducted a cluster-based permutation test using a repeated measures ANOVA comparing all SOAs for each of the two event conditions. We expected a significant difference between ERPs at time points after second stimulus onset depending on SOA. Indeed, this is what we found for each 2-event presentation (Supplementary Figure 4.6).

Discussion

Alpha oscillations have been postulated to reflect a marker of visual temporal resolution (Cecere et al., 2015; Samaha & Postle, 2015; VanRullen, 2018; Varela et al., 1981; White,

1963). In line with this view, recent evidence suggests that discrete perceptual cycles are linked to the frequency of ongoing alpha oscillations (Cecere et al., 2015; Keil & Senkowski, 2017; Minami & Amano, 2017; Ronconi et al., 2018; Ronconi & Melcher, 2017; Samaha & Postle, 2015; Wutz et al., 2018) (Figure. 4.1a). However, the generality of this relationship over different perceptual contexts, whether alpha frequency constitutes a constant individual trait or fluctuates over time, and whether it relates to perceptual sensitivity, perceptual or decisional bias remains controversial. Here, we demonstrated that temporal integration windows for 2-flash segregation in both unisensory and multisensory contexts are *not* associated with alpha frequency neither between nor within individuals. This was true both for the α TRH (resolution \propto alpha wavelength), as well as its inverse (resolution \propto alpha frequency), and there was no indication of different groups of participants either consistently showing the former or latter relationship over experiments.

These null findings are not easily explained by alpha power differences. Type 2 error was guarded against by i. conducting 9 two flash segregation experiments, ii. analysing both occipital scalp electrodes, occipital and parietal sources, and iii. computing Bayes factors to quantify evidence for or against H_0 . Importantly, we dissociated perceptual sensitivity (temporal resolution) from perceptual and decisional bias, which may have confounded earlier findings (Cecere et al., 2015; Ronconi & Melcher, 2017; Samaha & Postle, 2015). Here, we failed to show a link between bias (c) and alpha frequency, similarly arguing against a general relationship between these factors. It is important to note, however, that response bias was minimized by deploying factorial designs and randomizing trial sequences.

The reasons why we failed to replicate earlier results are ultimately unknown, but it is worth considering differences between our study and the work of Samaha and Postle (Samaha

& Postle, 2015), and Cecere, Rees and Romei (Cecere et al., 2015), whose tasks largely correspond to our experiments 4 and 6, respectively.

Samaha and Postle showed higher alpha frequencies both within and between participants for improved two flash segregation (Samaha & Postle, 2015). On one flash trials the authors presented a single long stimulus of 90-130 ms, and two flash trials were of similar overall duration, but contained a gap (10-50 ms). This is a significant contrast to our 1.3 ms low contrast flashes and might affect how participants integrate flashes perceptually, as well as how they place their internal decision criterion based on response strategy. A concern with a short, low-luminance flash is that two flashes are likely temporally summed (Gorea, 2015), appearing as a single bright flash for short SOAs. This issue is less pronounced (or even absent) for stronger visual stimuli (with larger luminance, size, contrast and duration) as used by Samaha and Postle. In terms of participants' strategy Samaha and Postle's design might be construed as a gap-detection task rather than a temporal discrimination task of one versus two flashes. However, it is unclear why this should lead to different response criteria and it should not affect temporal resolution. Regarding trait alpha frequency, Samaha and Postle pooled over one flash and two flash trials before estimating psychometric functions (similar to our yes-no exp. 4). This is not ideal as both sensitivity and bias influence threshold estimates (Macmillan & Creelman, 2005).

All in all, it seems likely that the results of Samaha and Postle are specific to their chosen stimulus parameters and do not generalize to different two-flash integration settings. This is also supported by the audio-visual entrainment study conducted by Romei and Melcher (Ronconi & Melcher, 2017), in which within subject alpha frequency was proportional to temporal integration window length (opposite to Samaha and Postle (Samaha & Postle, 2015)). Interestingly, a link between alpha frequency and temporal resolution within participants in line

with the α TRH was shown repeatedly in the missing element / odd element task (Ronconi et al., 2018; Wutz et al., 2018). However, these effects are driven by top-down task goals and are located more centrally than results relating to simple two-flash discrimination accuracy (Samaha & Postle, 2015). In addition, alpha frequency differences between correct and incorrect two-frame binding were not reported (Wutz et al., 2018).

Cecere, Rees and Romei (Cecere et al., 2015) showed a between subject correlation between alpha wavelength and perceptual integration window in the sound induced fission illusion. However, the authors used a purely subjective design, only presenting 1-flash, 2-sound trials. Such a design is particularly prone to large variations in response bias between participants, which might explain the discrepant results.

It was previously shown that alpha power increases and alpha frequency decreases over the course of a psychophysics task, effects that were partially inter-dependent (Benwell et al., 2019). Is it possible that these slow power or frequency drifts mask effects of interest? In principle, drifts in frequency should still reflect shifts in temporal resolution, unless they are elicited by different alpha sources or confounded by power changes (Benwell et al., 2019; Nelli et al., 2017), for example due to fluctuations in fatigue or sustained attention (Luo & Maunsell, 2015; Veksler & Gunzelmann, 2018). For within subject analyses of frequency we found identical results (significant versus non-significant) regardless of whether temporal effects of recording day, block and trial were regressed out. It seems therefore unlikely that slow drifts in alpha frequency (unrelated to temporal resolution) or power mask putative frequency effects on temporal resolution (although non-linear confounding effects may still exist).

Can alpha power fluctuations from day to day confound individual alpha peak frequency estimates? Assuming that an individual's alpha peak frequency constitutes a stable trait, a potential confounding role by alpha power should be mitigated by obtaining multiple peak

frequency estimates. On the other hand, if the true alpha peak frequency fluctuates from day to day, averaging over multiple recordings is suboptimal. Unfortunately, due to the relatively small number of trials on a given recording day we could not successfully model psychometric functions for individual recording sessions to compare them to single-day alpha frequency estimates.

Is it possible that individual alpha peak frequency estimates were simply too noisy to obtain reliable estimates? In most cases, alpha peak frequency estimates were relatively stable over posterior channels and recording days and the peak fitting routine could easily identify a single dominant peak. However, for some participants there was a clear bimodal pattern with either peak dominating in some recordings or channels thereby effectively leading to peak frequency estimates in-between these two modes (Supplementary Figure 4.7a). The occurrence of two alpha peaks in the periodogram is well-known and a challenge for any alpha peak fitting routine (Chiang et al., 2008; Chiang, Rennie, Robinson, van Albada, & Kerr, 2011). While there is no evidence to suggest that either peak is more important for temporal resolution, if only one peak is linked to temporal resolution this approach is suboptimal. However, when analysing trait alpha peak frequency links to temporal resolution using only peak estimates from unimodal distributions, results remained largely unchanged and Bayes factors still supported the null hypothesis in all cases (albeit not substantially in 3 out of 9; Supplementary Figure 4.7b).

We are convinced that our findings imply a real lack of a relationship between alpha frequency and visual temporal resolution in the context of two-flash segregation, especially with respect to within subject effects. Our sample size is equal to previous work (Samaha & Postle, 2015), and we performed substantially more experiments with larger trial pools, including internal replications. Regarding between subject effects our findings need to be interpreted more cautiously, because individual differences impact all analyses equally.

However, given that null results have been reported previously (Baumgarten et al., 2017), and considering that BFs consistently provide substantial support for H_0 , it seems likely that there is no general relationship between alpha frequency and visual temporal resolution between individuals.

Our findings of pre-stimulus alpha phase are less conclusive, and do not convincingly rule out an effect on either sensitivity or bias. Especially considering that a putative phase reset elicited by the first stimulus attenuates the link between pre- and post-stimulus phase (Mercier et al., 2013; Sauseng et al., 2007). Moreover, alpha phase results might be affected by slow drifts in power over the course of the experiment (Benwell et al., 2019). Given our alpha frequency results, and previous literature linking alpha phase to cortical excitability (Jensen et al., 2014), we hypothesize that pre-stimulus alpha phase has an impact only on bias, but not sensitivity in the context of two-flash perceptual segregation. Although this relationship was recently questioned by a pre-registration study that attempted to replicate two influential studies showing pre-stimulus alpha phase effects on subsequent target detection (Ruzzoli, Torralba, Fernández, & Soto-Faraco, 2019).

In summary, the alpha temporal resolution hypothesis – the notion that individual alpha frequency marks discrete temporal cycles in visual perception – has been researched for more than half a century without a clear conclusion (VanRullen, 2018; White, 1963). We aimed to clarify whether alpha frequency indexes the rate of visual temporal sampling in two-event perceptual contexts. Over 9 two-flash discrimination experiments we consistently found that alpha frequency is *not* related to visual temporal resolution neither within nor between participants. These results argue against a general mapping between posterior alpha frequency and visual temporal resolution in the context of two-flash integration or segregation.

Methods

Participants

After giving informed consent 20 right handed healthy adults (11 female, mean age: 22.4; age range 19 - 30) completed the study. An additional six participants were excluded after the first testing session, because the eye-tracker could not be reliably calibrated (three participants) or because participants responded too slowly, multiple times or not at all on > 10% of trials in the first two-interval forced choice run (Supplementary Figure 4.1). All participants had normal or corrected to normal vision and reported unimpaired hearing. Participation was compensated with £7.50 per hour. Ethical approval was granted by the University of Birmingham Science, Technology, Engineering, and Mathematics Review Committee (approval number ERN_11-0429AP22).

Stimuli

The visual stimulus was a truncated Gaussian light-grey circular blob with a diameter of 4° and a standard deviation of 1.34° (maximum luminance 12.91 cd/m^2) presented on a dark-grey background (0.71 cd/m^2) for approximately 2 ms. The auditory stimulus was a 2 ms pure tone (3500 Hz) with a 0.5 ms linear ramp at on- and offset (maximum amplitude at the left earpiece was measured at 80 dB SPL). All stimuli were presented 15° to the left of a central light-grey fixation cross (12.91 cd/m^2). Onset of audio-visual stimulus pairs was synchronous. All stimuli were created in Matlab 2014a (Mathworks, Natick, MA, USA) and presented with Psychtoolbox 3 (<http://psychtoolbox.org>).

Design and procedure

The study consisted of 3 (task: Two-interval forced choice, yes-no, yes-no at perceptual threshold) x 3 (two flashes presented with: 0-2 sounds) experiments (Figure. 4.1). A normal testing schedule consisted of 5 days, each including two task runs (~70 min of task performance per run, see Supplementary Figure 4.1 for details). Before the first task run of a session,

participants completed 48 practice trials. EEG was recorded only during yes-no and yes-no threshold tasks (exp. 4-9). In addition, 2 minutes of relaxed, eyes-closed EEG activity were recorded before and after each yes-no and yes-no threshold session.

Two-interval forced choice task. In the two-interval forced choice (2IFC) task (exp. 1-3) a target (2-flash, X-sound) and distractor (1-flash, X-sound) interval were presented with 800 ms delay in random order. Participants responded which interval contained two flashes. Stimulus presentations followed a factorial $2 \times 3 \times 8$ design, denoting the interval order (1-2 or 2-1 flashes), number of sounds (0-2), and stimulus onset asynchronies (SOAs; 0.025, 0.042, 0.05, 0.058, 0.075, 0.108, 0.158, 0.225 ms). The response window was 1.5 s. It took another 1.5 s plus jitter (uniformly sampled between 0 and 250 ms) until the next trial started. To distinguish inter-trial intervals from 2IFC-interval delays the fixation cross was rotated by 45° after each trial. A run consisted of 12 blocks of 96 fully randomized trials of each type. Two 2IFC runs were acquired per participant for a total of 48 trials per condition. Importantly, the 2IFC task design leads to theoretically unbiased responses after pooling over trials with different interval orders. No EEG was acquired, as it is difficult to relate spontaneous brain activity prior to stimulus presentation to perception of either the first or second 2IFC interval.

Yes-no task (multiple SOAs). In the yes-no task (exp. 4-6) a single interval was presented per trial and participants indicated whether they perceived one or two flashes (Figure. 4.1). Note that in this case response bias does play a role. To ensure that pre-stimulus EEG analyses are not confounded by response related activity, an additional 0.6 s delay was included between trials. Each block consisted of 97 trials and included one catch trial where no stimulus was presented. Overall 96 trials were acquired per condition.

Yes-no threshold task (SOA at perceptual threshold). In the yes-no threshold task (exp. 7-9), for each participant and 2-event condition a staircase procedure determined a single SOA close

to perceptual threshold at which a participant perceives 2 flashes approximately 50% of the time. In this task 73 trials were presented in each block including 1 catch trial and 12 trials per flash-sound combination. Trial order was randomized with the constraint that no two trials of the same condition could follow in direct succession. In total 288 trials were acquired per condition. Participants indicated what they perceived, as well as how confident they were in their perceptual estimates (not analysed here). Consequently, response windows were extended to 2.5 s. Otherwise, timings were equal to the yes-no task.

Behavioural responses and key mappings. In the 2IFC and yes-no task (exp. 1-6) participants responded with their index fingers using the ‘F’ and ‘J’ keys on a standard computer keyboard. In the yes-no-threshold task (exp. 7-9) all fingers were deployed using the ‘A’, ‘S’, ‘D’, ‘F’ keys with the left hand and the ‘J’, ‘K’, ‘L’, ‘;’ keys with the right hand. The responding hand coded the perceptual response, and the responding finger coded perceptual confidence. An equal number of trials were acquired for each of the two possible key mappings for each task. Key mapping was counterbalanced between sessions and participants. Participants were given at least 48 practice trials to get accustomed to a change in key mapping. They were instructed to correct mistakes if they noticed them.

Behavioural exclusion criteria. All trials with multiple responses on the current or previous trial were discarded from all analyses (2IFC: $M = 109.9$, range: 20, 260; Yes-no: $M = 196.6$, range: 23, 590; Yes-no threshold: $M = 201.35$, range: 39, 374). In addition, trials with response times faster than 100 ms after last stimulus onset were excluded (2IFC: $M = 14.89$, range: 1, 36; Yes-no: $M = 0.54$, range: 0, 7; Yes-no threshold: $M = 0.16$, range: 0, 2). This left the following trial distributions: 2IFC (exp. 1-3; $M = 2170.3$, range: 1940, 2278); Yes-no (exp. 4-6; $M = 4410.6$, range: 3860, 4756); Yes-no threshold (exp. 7-9; $M = 3294$, range: 3068, 3805).

Experimental setup

Testing took place in a darkened room with additional light-shielding the stimulus PC and participant. The presentation screen was a 19'' cathode ray tube (CRT) display with a refresh rate of 120 Hz and a resolution of 1024 x 768 pixels. Sounds were presented via EARtone 3A Insert Earphones (Aearo Company Auditory Systems, 1997). Participants' gaze position was tracked with an EyeLink 1000 Plus eye-tracker to be able to exclude trials where central fixation was not maintained (SR Research Ltd., 2014).

EEG recording and preprocessing

EEG data were recorded with 64 active electrodes (actiCAP, Brain Products Ltd., 2014) arranged in an extended 10-20 montage (ground: AFz, reference: FCz). Data were amplified with two 32 channel BrainAmp DC amplifiers (Brain Products Ltd., 2014) and digitized at 1000 Hz with a high-pass filter of 0.1 Hz. EEG preprocessing and analyses were carried out with the FieldTrip toolbox (Oostenveld et al., 2011) for Matlab and custom written Matlab code. Data were low-pass (99 Hz) and notch filtered (48-52 Hz). Noisy channels and time epochs, including time periods encompassing saccades towards the stimulus, were identified visually and discarded. Independent component analysis (ICA) was applied to the cleaned EEG data and components unrelated to brain activity (e.g. blinks, heart beat) were rejected before back-transforming components to channel space (on average 1.4 components were rejected per recording, range: 1, 5). Signals at positions of excluded channels were estimated with spline interpolation using the standard 10-10 layout and default neighbour positions defined by the FieldTrip toolbox (On average 2.1 channels were interpolated per recording, range: 0, 10). Next, data were segmented into trials (-1.2 s to 0.5 s relative to first stimulus onset), and down-sampled to 250 Hz. Trial segments without marked artefacts were detrended linearly, re-referenced to an average over channels and used for further analysis.

Computation of pre-stimulus spectral features

Instantaneous frequency was calculated using the method and implementation developed by Cohen (Cohen, 2014). First, the pre-stimulus period (0.6 – 0.1 s prior to first flash onset) was filtered with a 6-14 Hz plateau shaped window (15% transition zones). Next, a Hilbert transform was applied and the first temporal derivative of the resulting phase angle time series was taken, which gives an estimate of frequency between adjacent time points (to obtain Hz the sampling rate and 2π are used for scaling). Ten median filters were applied independently with linearly spaced orders between 5 and 200 sample points (approx. 10-400 ms), and the median of the resulting median filtered time courses was computed to minimize the influence of noise driven jumps on phase estimates.

We computed power and phase estimates by applying a discrete Fourier transformation on 0.5 s Hanning tapered sliding time windows shifted in 40 ms steps over the 0.6 - 0.1 s pre-stimulus time window. To avoid smearing of stimulus evoked activity into the pre-stimulus period, post-stimulus values were set to zero. Each 0.5 s epoch was zero-padded to 8 s (leading to a frequency resolution of 0.125 Hz).

Controlling for temporal confounds. To ensure that changes in alpha frequency or power over time did not mask the effects of interest, we regressed out temporal nuisance factors. To this end, frequency and power estimates were averaged over the pre-stimulus window (0.6 – 0.1 s) and power was log-transformed. Next, two linear regressions were performed with frequency and log-power as dependent variables. In both cases independent variables were participant (categorical), session (categorical), block number (numerical) and trial number (numerical). All possible interactions were included.

Disambiguating between perceptual sensitivity and bias

To disambiguate between perceptual sensitivity (i.e. temporal resolution) and response or perceptual bias, trials were ordered separately for frequency, power and phase and divided into

3 equally sized bins for each time point, condition and SOA. Next, d' (sensitivity) and criterion c (bias) were estimated for the first and third bin (e.g. low and high frequency) assuming a two equal-variance Gaussian signal detection model (Macmillan & Creelman, 2005). To study phase opposition the intermediate phase bin (excluded from analysis) contained trials closest to a line dividing the unit circle, leaving bin 1 and 3 with maximally opposing phase angles. According to the signal detection framework in our yes-no tasks, participants internal 1-flash, X-sound representations are drawn from the noise distributions and 2-flash, X-sound representations are drawn from the signal distributions. Sensitivity d' is the distance between a given pair of signal and noise distribution means, which can be estimated as follows:

$$(1) \quad d' = z(HR) - z(FAR)$$

where HR denotes hit rate (proportion of “see 2” responses when 2 flashes were presented) and FAR denotes false alarm rate (proportion of “see 2” responses when 1 flash was presented).

The estimate of criterion c is computed as follows:

$$(2) \quad c = -0.5(z(HR) + z(FAR))$$

and quantifies the bias toward responding “see 2” ($c > 0$) or “see 1” ($c < 0$). At $c = 0$ the observer is unbiased. A small constant of 0.1 was added to both hits and false alarms to avoid infinities in the calculation of d' and c .

Relating spontaneous oscillatory fluctuations to perceptual sensitivity and bias

At the scalp level pre-stimulus time courses of d' and c were averaged over three channels of interest (O2, PO4, PO8), chosen based on the position of the visual stimulus in the left visual field (see supplementary materials for source analysis). In the yes-no task d' and c were further averaged over intermediate SOAs (0.05, 0.058, 0.075, 0.108 ms).

We contrasted sensitivity and bias in the first and third bin with paired t-tests at each time point. To find opposing phase bins that maximally differ in sensitivity or bias we first

rotated the threshold for dividing trials into bins in steps of 10° and selected the split with the largest absolute d' or c difference. Next, we randomly shuffled the phase relationships between trials and repeated this process 1000 times. The largest d' or c difference was then compared against the median of these 1000 iterations with dependent samples t-tests (one-sided).

To correct for multiple comparisons, a cluster-based permutation test was performed (Maris & Oostenveld, 2007). Clusters were given as the sum of adjacent t-values that passed the 0.05 threshold. The null distribution was constructed by shuffling the condition labels (bin 1 or 3) randomly for each participant 5000 times, repeating the clustering of t-values and taking the maximum cluster value of each iteration. Significant clusters were selected at the 0.025 level (or 0.05 for phase).

Bayes factors. To quantify evidence for (or against) the null hypothesis, we computed Bayes factors (BFs) for dependent samples t-tests (Rouder et al., 2009). A BF is a ratio of evidence between the alternative hypothesis (H_a) and the null hypothesis (H_0). For example, a value of 3:1 means that H_a is three times more likely than H_0 . As a general rule BFs < 3.2 or > 0.32 are considered inconclusive (Jeffreys, 1998). We used a standard JZS prior consisting of a Jeffreys prior on variance for H_a and H_0 , a Cauchy prior on effect size for H_a and a point-zero prior on effect size for H_0 (Rouder et al., 2009).

Trait alpha peak frequency and power estimation

To identify each participant's trait alpha peak frequency and power we used the algorithm and implementation developed by Corcoran and colleagues (Corcoran et al., 2018). This method has been validated with real and simulated data and was shown to provide reliable estimates (Corcoran et al., 2018). If not otherwise specified, parameters used in the peak-fitting routine were all set to default values (Corcoran et al., 2018). Eyes-closed data were segmented into 2 s pieces with 1 s overlap, Hanning tapered, Fourier transformed and averaged to obtain power

spectral density (PSD) estimates (Welch, 1967). Task related PSD was computed over the 0.7 s pre-stimulus time windows. In both cases data were zero-padded to 8 s resulting in a frequency resolution of 0.125 Hz. We obtained peak and power estimates for all occipital and parietal channels (O1, O2, Oz, PO9, PO7, PO3, POz, PO4, PO8, PO10, P7, P5, P3, P1, Pz, P2, P4, P6, P8; for eyes-closed the following channels were additionally included: TP7, TP8, TP9, TP10, CP1, CP2, CP3, CP4, CP5, CP6, CPz) separately for each session. We used all posterior channels (as opposed to the three channels used for within subject analyses) to improve estimation of trait peak frequency, which sometimes fails for single channels (see source analyses for a separate investigation of occipital and parietal alpha sources). For an overview of individual alpha peak frequency estimates see Supplementary Figure 4.7.

Individual trait alpha peak frequency was computed as the average over session peak estimates weighted by the number of sessions' successful channel peak estimates. Session peak estimates were averages over channels weighted by channels' peak power. Peak-power was defined as the sum of power values between inflection points of the peak's flanks. Individual trait power was obtained by averaging over channels and sessions, similar to frequency. For source analyses the same procedure was followed, except that only a single independent component was chosen per session for occipital and parietal sources, respectively.

Fitting psychometric functions to quantify individual windows of perception

To obtain an estimate of perceptual window length for each participant and flash-sound pairing (i.e. 1FXS & 2FXS) in the 2IFC (exp. 1-3) and yes-no task (exp. 4-6) we described response accuracy as a psychometric function of SOA. Formally, this can be written as follows:

$$(3) \quad \psi(x; \alpha, \beta, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda) F_L(x; \alpha, \beta)$$

with

$$F_L(x; \alpha, \beta) = \frac{1}{1 + \exp(-\beta(x - \alpha))},$$

where x denotes SOA, F_L is the logistic function, and parameters α, β, γ and λ describe threshold (inflection-point), slope, guess-rate and lapse-rate, respectively. The threshold parameter α is the point at which participants can distinguish 1 from 2 flashes 50% of the time, discounting guesses and lapses. It therefore serves as a measure of perceptual resolution (Samaha & Postle, 2015). However, this measure is affected by participants' bias. Hence, the 2IFC task, which theoretically leads to unbiased responses (Macmillan & Creelman, 2005) provides an estimate of α (almost) purely driven by perceptual sensitivity, while the yes-no task α estimates are probably tainted by bias.

Psychometric function parameters were estimated using maximum likelihood estimation (MLE), as implemented in the Palamedes toolbox (Prins & Kingdom, 2018) for Matlab.

Contrasting trait alpha frequency and power with perceptual window length

To assess whether individual trait alpha frequency or power estimates are correlated with participants' temporal windows of perception, we computed Pearson's correlation coefficients with threshold parameters α for each condition and task (2IFC or yes-no). Assumptions of parametric statistics were assessed via visual inspection of residuals (quantile-quantile-plots and residual plots). Normality of residuals was assessed using visual inspection and Shapiro-Wilk tests. As normality of residuals was in doubt in some cases, we furthermore computed Spearman's rank correlation coefficients.

Bayes factors. Bayes factors for Pearson's correlation coefficients were computed according to the recommendations by Wetzels and Wagenmakers (Wetzels & Wagenmakers, 2012), using a JZS prior.

Alpha source separation

To disambiguate between occipital and parietal alpha sources, we localized independent

component (IC) dipoles (Gulbinaite et al., 2017), using the DIPFIT 2.3 plug-in of the EEGLAB toolbox v14.1.1 (Delorme & Makeig, 2004) for each recording session. First, artefact free data were band-pass filtered between 5 and 15 Hz and the 1 s windows prior to first flash onset were selected. Next, the first 20 principal components of the data were subjected to independent component analysis and source localized. To this end it was assumed that each IC had a single-equivalent current dipole (Delorme et al., 2012). The dipoles best explaining scalp topographies were localized within a 3-layer boundary element volume conduction model based on canonical values (in MNI space (Evans, Collins, Milner, & Milner, 1992)) from the EEGLAB toolbox. Electrode scalp positions were fine-tuned visually by linearly morphing standard 10-10 electrode positions to align with digitized electrode positions (morphed to MNI space) measured with a Polhemus FASTRAK system (Polhemus Inc.). Dipole model fits that explained less than 85% of an IC's variance were rejected. In such cases a symmetric two-dipole model was fit mirrored at the x-axis (anterior commissure to nasion), and tested against the same threshold. Occipital and parietal dipoles were selected based on proximity to regions of interest (ROIs) defined in MNI coordinates based on previous work (occipital left: -20, -90, 0; occipital right: 20, -90, 0; parietal left: -20, -70, 50; parietal right: 20, -70, 50) (Gulbinaite et al., 2017). The dipoles closest to each ROI were selected with the constraint of not being closer to the other ROI and the source location being within one centimetre of a participant's cortical grey matter. Grey matter volumes were obtained with SPM12 ("Statistical Parametric Mapping: The Analysis of Functional Brain Images—1st Edition," n.d.) after segmenting T1 scans with tissue probability maps, normalizing to MNI space and thresholding grey matter probabilities at 0.05. For three participants structural scans were not available. We therefore used the canonical template T1 from the EEGLAB toolbox. Finally, only dipole estimates within 4 cm of a ROI were considered valid. In the yes-no task (exp. 4-6) 44/58 occipital and 58/58 parietal

dipole fits were accepted. In the yes-no threshold task (exp. 7-9) 32/41 occipital and 40/41 parietal dipole fits were accepted.

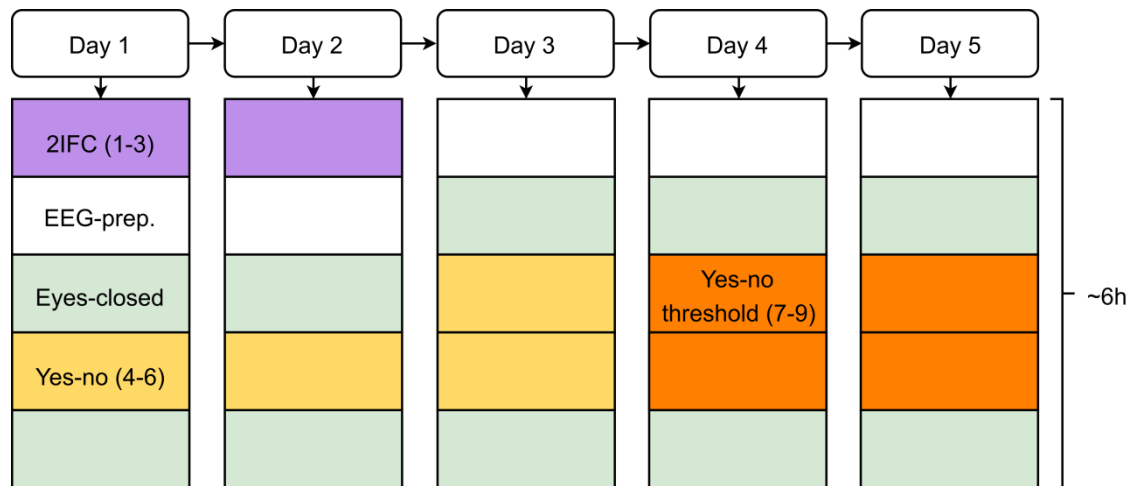
Control analysis of event related potentials

For each two-event context in the yes-no experiments 4-6, we computed ERPs and tested for significant differences between SOAs. Specifically, repeated measures ANOVAs were computed at each time point between 0 and 400 ms. A cluster based permutation test was performed to correct for multiple comparisons (Maris & Oostenveld, 2007). Clusters were given as the sum of adjacent F-values that passed the 0.05 threshold. Significant clusters were selected at the 0.05 level when comparing against a null distribution obtained over 5000 permutations.

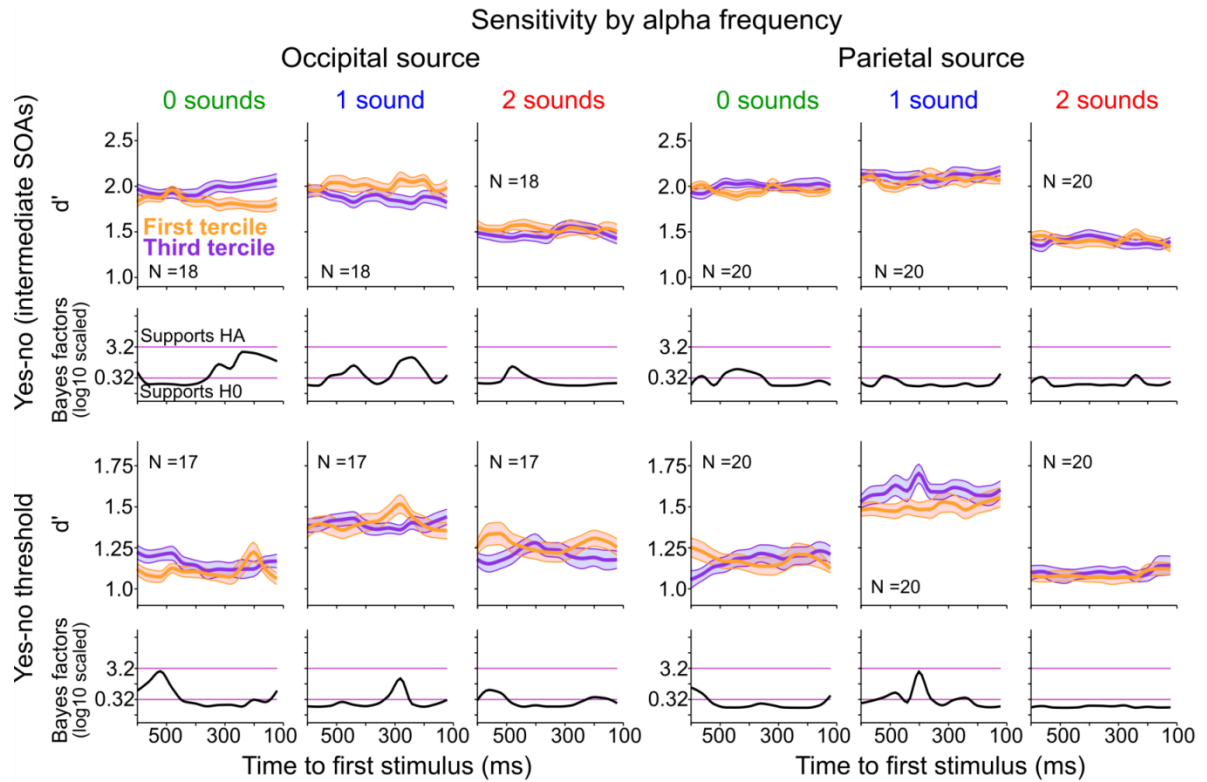
Supplementary materials

Interpreting source level analyses in EEG data

While our source level analysis provides us with a way of separating occipital and parietal alpha band components a few notes of caution in their interpretation are warranted. When projecting scalp EEG (or MEG) data into source space one needs to solve the so-called inverse problem. This problem arises, because in principle there is an infinite number of possible current dipoles that give rise to the exact same pattern measured at the scalp surface. Therefore, any source modelling technique needs to make certain assumptions to restrict the number of possible solutions. In the dipole fitting of independent components approach the restriction is that a single dipole is assumed to underlie a single independent component. This assumption is reasonable, because scalp topographies of independent components frequently support a unique underlying current dipole (Delorme et al., 2012), but may not always be correct. Moreover, it is possible that the regions of interest (occipital and parietal source coordinates) might not reflect the optimal dissociation points of occipital and parietal sources, thereby leading to misclassifications of true occipital source activity as parietal source activity and vice versa. Nevertheless, given that these coordinates have been used previously and our criterion of not accepting components as either parietal or occipital when they were closer to the respectively other region of interest, we have reasonable confidence that our source estimates are at least moderately accurate. In addition, our criterion for accepting only dipoles that explain at least 85% of an independent component's variance ensured that a single dipole fit was indeed a good approximation to the component's generators, which per definition, were either located mostly occipitally or parietally.

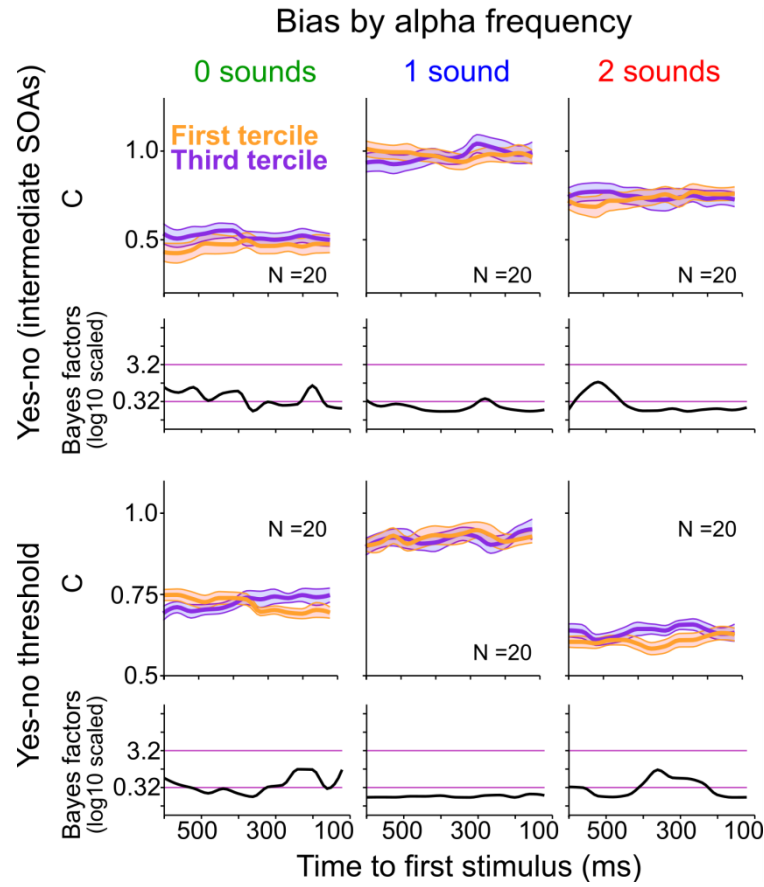


Supplementary Figure 4.1. Typical study time table. A planned testing schedule consisted of 5 days with approximately 6 hours per day. Square blocks indicate experimental runs, EEG preparation and eyes-closed resting state recordings. An experimental run included roughly 70 minutes of pure task time and breaks were flexibly offered in-between task blocks (~6 min task time per block). Eyes-closed runs lasted between 2 and 3 minutes.

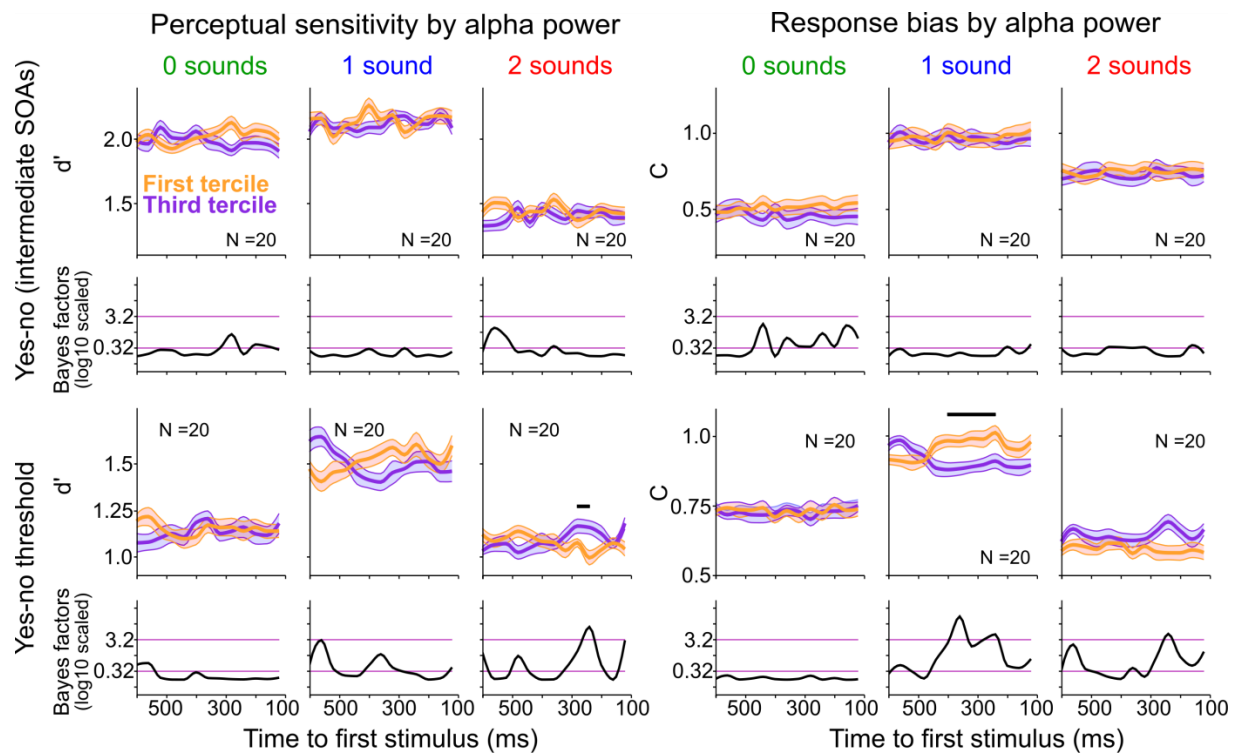


Supplementary Figure 4.2. Perceptual sensitivity by occipital and parietal source alpha frequency.

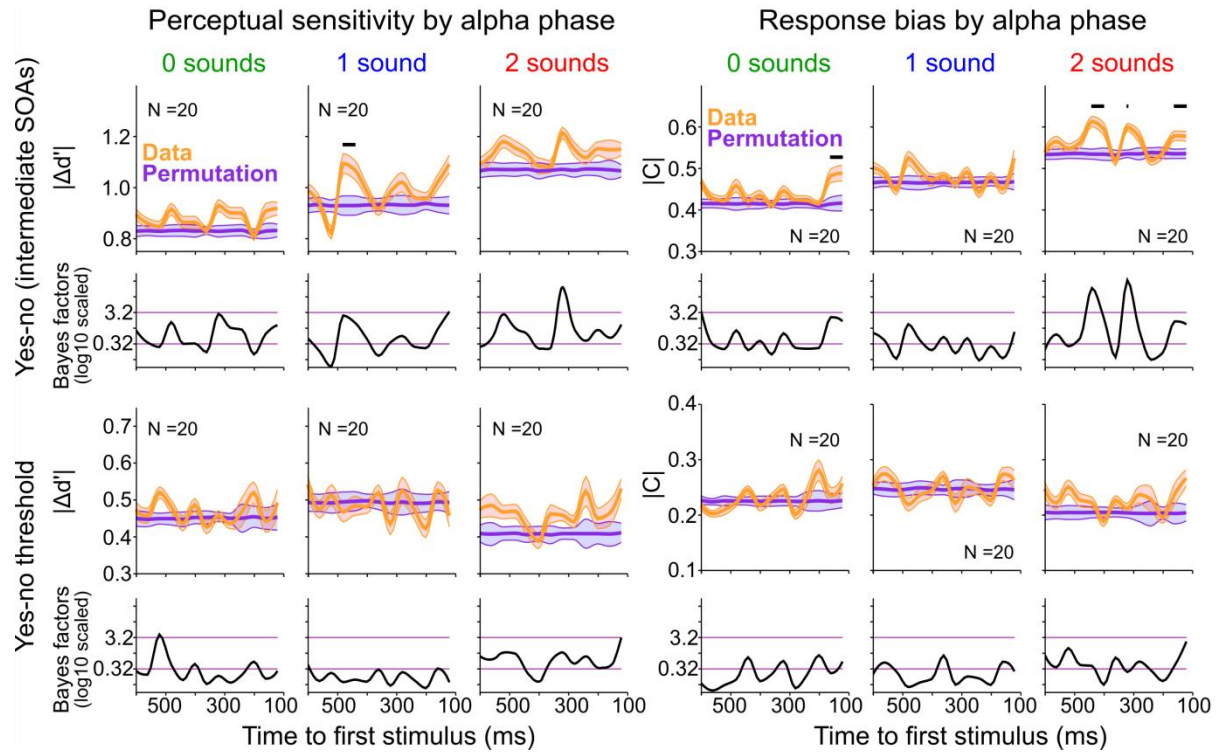
Pre-stimulus time courses of perceptual sensitivity for low (first tercile) and high (third tercile) pre-stimulus alpha frequency at occipital and parietal sources. Significant time points are marked with black lines (all tests *n. s.*). Error bars denote ± 1 within subjects SEM. Pink lines mark BF thresholds of substantial evidence for H_0 (0.32) or H_a (3.2).



Supplementary Figure 4.3. Bias by pre-stimulus alpha frequency. Pre-stimulus time courses of bias for low (first tercile) and high (third tercile) pre-stimulus alpha frequency at electrodes O2, PO4 and PO8. Bias does not statistically differ between low and high alpha frequency for all yes-no experiments (exp. 4-9). Error bars denote ± 1 within subjects SEM. Pink lines mark BF thresholds of substantial evidence for H_0 (0.32) or H_a (3.2).



Supplementary Figure 4.4. Perceptual sensitivity and bias by pre-stimulus alpha power. Pre-stimulus perceptual sensitivity (left) and bias (right) for low (first tercile) and high (third tercile) pre-stimulus alpha power. Error bars denote ± 1 within subjects SEM. Black lines denote significant clusters. Pink lines mark BF thresholds of substantial evidence for H_0 (0.32) or H_a (3.2).



Supplementary Figure 4.5. Perceptual sensitivity and bias by pre-stimulus alpha phase. Pre-stimulus perceptual sensitivity and bias for opposing phase bins with maximally different absolute sensitivity and bias values (data) are compared against median values obtained from a permutation distribution (permutation). Error bars denote ± 1 within subjects SEM. Black lines denote significant clusters. Pink lines mark BF thresholds of substantial evidence for H_0 (0.32) or H_a (3.2).

Supplementary Table 4.1. Source level between subject analyses of alpha peak frequency and perceptual discrimination threshold.

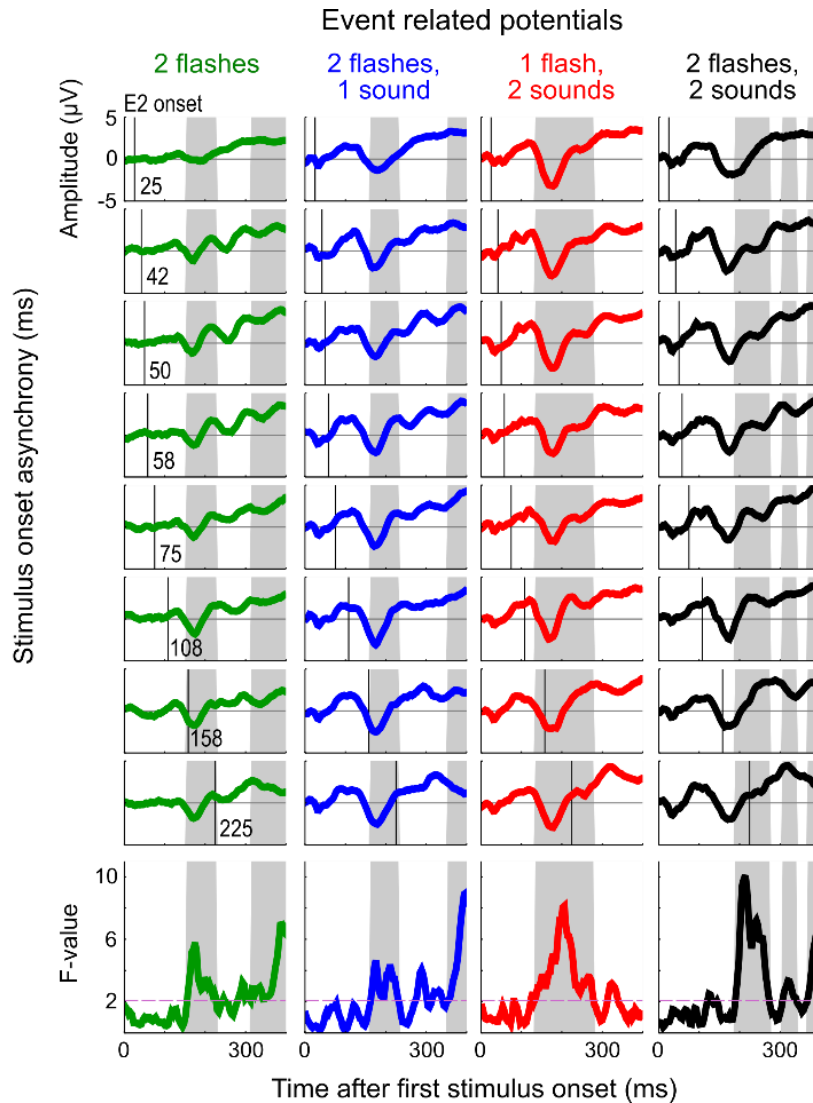
Threshold definition	Pre-stimulus occipital source			Pre-stimulus parietal source		
	N	r	BF	N	r	BF
2IFC						
1F & 2F	16	-0.50	1.22	19	-0.39	0.67
1F1S & 2F1S	17	0.17	0.23	20	0.14	0.20*
1F2S & 2F2S	16	-0.15	0.22	18	0.07	0.19*
Yes-no pooled						
1F & 2F	16	-0.03	0.19	19	-0.11	0.19*
1F1S & 2F1S	17	-0.05	0.19	20	0.03	0.17*
1F2S & 2F2S	13	0.17	0.24	16	0.15	0.22*
Staircase SOA						
2F	17	-0.08	0.19	20	0.20	0.25*
2F1S	17	-0.02	0.18	20	0.20	0.24*
1F2S	17	0.45	0.93	20	0.12	0.19*
2F2S	17	0.45	0.93	20	0.12	0.19*

N, number of participants; *r*, Pearson's correlation coefficient; *BF*, Bayes factor

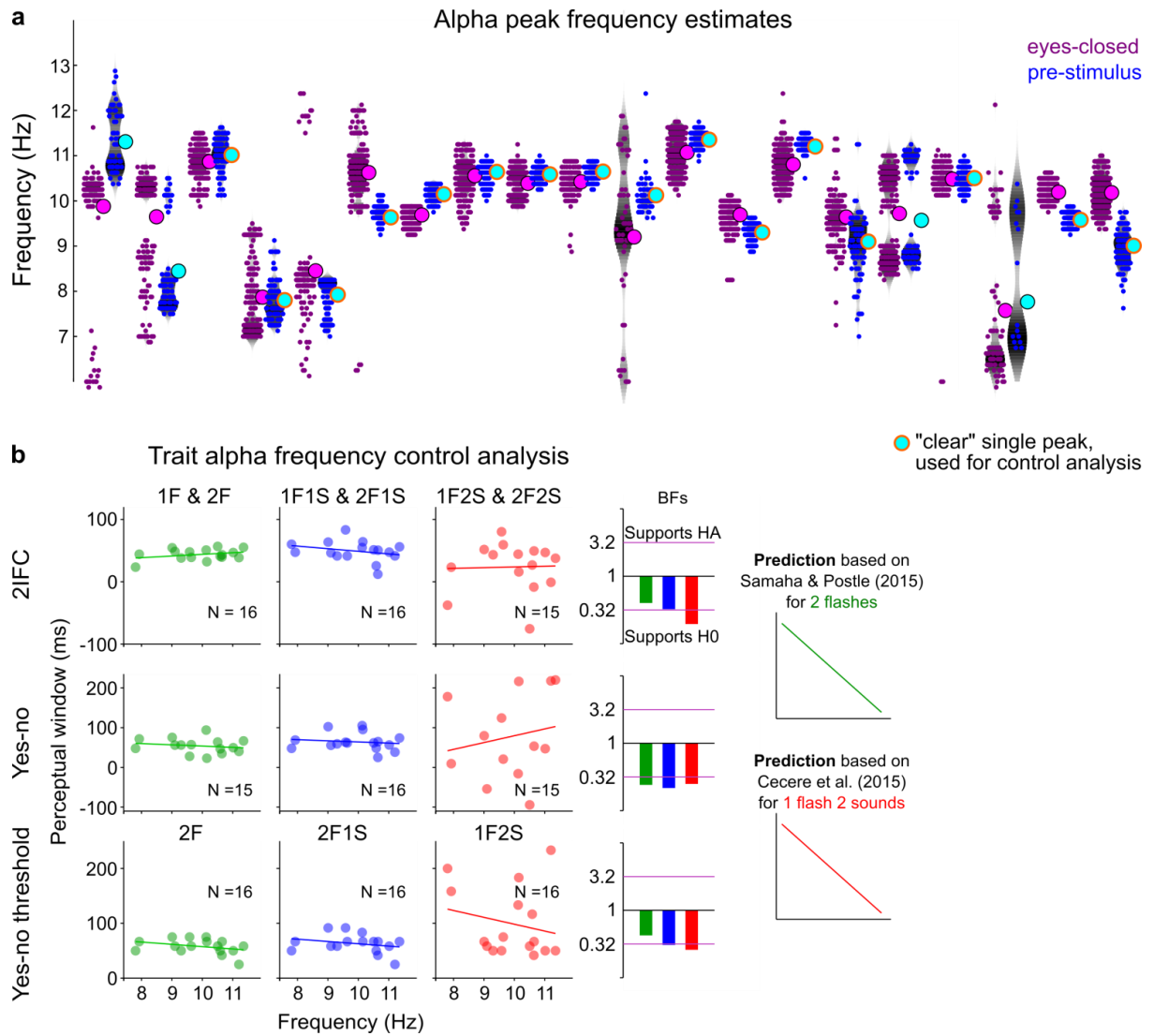
Supplementary Table 4.2. Between subject analyses of alpha power and perceptual discrimination threshold.

Threshold definition	Eyes-closed sensor level			Pre-stimulus sensor level		
	N	r	BF	N	r	BF
2IFC						
1F & 2F	19	-0.223	0.27	19	-0.31	0.40
1F1S & 2F1S	20	-0.01	0.17	20	0.05	0.17*
1F2S & 2F2S	18	0.10	0.19	18	0.00	0.18*
Yes-no pooled						
1F & 2F	19	-0.49	1.72	19	-0.24	0.28*
1F1S & 2F1S	20	-0.17	0.22	20	-0.07	0.18*
1F2S & 2F2S	16	0.40	0.19	16	0.41	0.65
Staircase SOA						
2F	20	0.08	0.18	20	0.13	0.20*
2F1S	20	-0.06	0.18	20	-0.00	0.17*
1F2S	20	-0.06	0.18	20	0.06	0.18*
2F2S	20	-0.06	0.18	20	0.06	0.18*

N, number of participants; *r*, Pearson's correlation coefficient; *BF*, Bayes factor



Supplementary Figure 4.6. Event related potentials differ between SOAs. Event related potentials (average over channels O2, PO4 and PO8) for all 2-event conditions in the yes-no paradigm (4-6), unpacked for each stimulus onset asynchrony (SOA). Black vertical lines denote the onset of the second flash/sound event. Grey shaded areas denote significant differences between any of the 8 SOAs. F-statistics are depicted in the last row. The purple dashed line denotes the critical F-value used for clustering.



Supplementary Figure 4.7. Trait alpha frequency control analysis. *a*, Individual alpha peak frequency estimates for eyes-closed (purple) and pre-stimulus EEG activity (blue). Light coloured disks denote final estimates, whereas dark coloured dots denote single channel estimates pooled over all recording sessions. Note that for some participants the alpha peak estimate distribution is bimodal (e.g. third to last). *b*, Prestimulus trait alpha frequency correlations with temporal binding window lengths obtained from each experiment for the subset of participants showing a clear unimodal alpha peak estimate distribution over channels and recordings (as seen in *a*; light blue disks with brown outline). Two-sided Bayes factor (BF) thresholds of substantial evidence for H_0 (0.32) or H_A (3.2) are marked in purple.

Chapter 5 - General discussion

The research presented in this thesis aimed to improve our understanding of visual and audio-visual perceptual decision making in time. In chapter 2 we scrutinized the temporal determinants of one versus two flash temporal segregation in a unisensory and multisensory context (0-2 beeps). In chapter 3 we tracked and compared the neural representations of illusory and real flashes over the course of perceptual choice formation. In chapter 4 we examined the role of posterior alpha frequency in visual temporal resolution. In this last chapter I briefly summarize the findings from the empirical work, evaluate their contribution to the field, and outline future directions for further inquiry.

Summary of empirical findings

Chapter 2: Temporal determinants of the sound induced fission and fusion illusion

In the sound induced fission illusion, a second illusory flash is reported when one flash and two beeps are presented (Shams et al., 2002). In the sound induced fusion illusion a single flash is reported despite two flashes and one beep being presented, even when two unisensory flashes can already be resolved in time (Andersen et al., 2004; Shams et al., 2002). These phenomena are associated with both perceptual sensitivity changes, and bias toward reporting the presented number of beeps, as indicated by studies where the temporal asynchrony between stimuli was fixed (Chen et al., 2017; Kaposvári et al., 2014; McCormick & Mamassian, 2008; Vanes et al., 2016). However, little is known about how these factors unfold as a function of asynchrony between the two flashes or beeps, which this chapter sought to address.

We discovered that the fission illusion has a broader temporal binding window than either unisensory two-flash fusion or the fusion illusion. Moreover, the fusion illusion showed higher sensitivity compared to unisensory contexts, which was explained by improved

perceptual reliability of 1-flash, 1-beep congruent presentations (as opposed to 1-flash, 0-beep). Finally, both fission and fusion responses showed biases toward the presented number of beeps. Note, however, that caution is required when interpreting bias, as its precise origin cannot be discerned by the signal detection model (Witt et al., 2015). These findings extend previous work deploying fixed SOAs and highlight the importance of dissociating perceptual sensitivity from bias, which earlier research on the TBWs of fission or fusion has largely neglected (Apthorp et al., 2013; Bidelman, 2016; Cecere et al., 2015; Foss-Feig et al., 2010; Hamilton et al., 2013; Narinesingh et al., 2017; Shams et al., 2000, 2002).

Chapter 3: Decoding real and illusory flash representations during perceptual decision making

Both fission and fusion illusion can be explained by computational models of multisensory decision making (Cuppini et al., 2014; Shams et al., 2005; Wozny et al., 2008). Despite this, evidence suggests that the underlying neural mechanisms are not entirely equivalent. Both illusions are associated with changes in BOLD activity in early visual cortex, as well as polysensory regions such as STS (Watkins et al., 2007, 2006). Yet, the time courses of illusion versus non-illusion difference waves, as indicated by ERP research, diverge within the first 400 ms after first stimulus onset (Mishra et al., 2008, 2007). Such early effects are likely to be strongly influenced by sensory processes, and it is unknown how the perceptual decision evolves in relation to the commitment of a response.

In chapter 3 we used multivariate pattern decoding to track the perceptual outcomes of one versus two flashes relative to stimulus and response. We found that perceptual decisions could be decoded long before a choice was enacted for each flash-beep condition, with very similar neural patterns for unisensory two flash and fusion trials, but very different patterns for fission. Hence, apart from early stimulus evoked activity (Mishra et al., 2008, 2007), the

representation of the perceptual outcome itself is different between fission and both fusion illusion and unisensory two-flash contexts. However, this finding was strongly influenced by differences in confidence between perceptual outcomes, establishing metacognition as a crucial variable to consider when studying perceptual decision making. After controlling for confidence there was significant cross-condition generalization demonstrating similar neural representations between 2-flash and 2-flash & 1-beep, as well as 2-flash & 1-beep and 1-flash & 2-beep, but not between 2-flash and 1-flash & 2-beep. The fusion illusion condition therefore appears to share similar neural patterns with both unisensory flash fusion and the fission illusion condition, yet these patterns are not shared between unisensory and fission conditions. Speculatively, this could be explained by 2-flash presentations and simple presence of auditory stimuli, respectively.

Finally, we found that metacognitive awareness of level 1 perceptual evidence is substantially higher (on average close to optimal) in the fission condition compared to either unisensory or fusion contexts. This finding provides an interesting perspective on previous work showing that the illusory second flash “feels” real and is resistant to feedback training unless participants are monetarily rewarded (Rosenthal et al., 2009). Even though perceptual representations might be largely similar (at least between fusion and fission), metacognitive access appears to be superior for fission, which could be leveraged when given a strong incentive to learn.

Chapter 4: What is the role of posterior alpha frequency in visual temporal resolution?

An influential theory in cognitive neuroscience posits that the frequency of alpha oscillations determines the temporal resolution of visual perception (VanRullen, 2018). Yet, scientific evidence to date has been inconclusive. Research has associated a high alpha frequency with *better* (Samaha & Postle, 2015) and *worse* (Ronconi & Melcher, 2017) two flash temporal

segregation in visual perception, and fewer double flash illusions in audio-visual perception (Cecere et al., 2015; Keil & Senkowski, 2017). Moreover, current conclusiveness is limited, because of different experimental designs and analysis approaches. In addition, research did not (fully) dissociate perceptual sensitivity from bias (Cecere et al., 2015; Keil & Senkowski, 2017; Ronconi & Melcher, 2017; Samaha & Postle, 2015). In light of these controversies, in chapter 4 we scrutinized the influence of alpha frequency on perceptual resolution in a series of two-flash segregation experiments. Across nine experiments and various analysis approaches we found robust evidence that alpha frequency is *not* related to perceptual resolution. These results challenge the influential notion that the frequency of alpha oscillations is a fundamental marker of perceptual resolution in two-flash integration or segregation.

Contributions, limitations and future directions

The sound induced flash illusion has been researched extensively over the past 20 years and significant advances have been made with respect to its computational principles, neuroanatomical and functional neural substrates (Shams, 2012); just as significant progress has been made in the larger field of multisensory integration (Spence, 2018) and perceptual decision-making (Hanks & Summerfield, 2017). Computational accounts of MSI (including the SIFI) argue that simple perceptual decisions are often close to optimal and follow the rules of Bayesian inference (Rohe et al., 2018; Shams, 2012). That is, together with prior expectations, the reliability of each individual signal is considered when computing a multisensory perceptual estimate, thereby creating a more reliable representation. In the SIFI this results in an incorrect numerical estimate of presented visual flashes, because visual temporal resolution is low compared to audition. Yet, surprisingly little work has thus far attempted to track the perceptual decision-making process in the context of the SIFI from its start (prior to stimulus presentation)

to its conclusion (response initiation), which is where this thesis makes a significant contribution.

Before placing the empirical findings in the larger theoretical landscape, however, it is prudent to consider potential caveats and limitations. Throughout this work we have employed signal detection theory to segregate perceptual sensitivity from biases in multiple full factorial psychophysics tasks with intermixed audio-visual trial presentations. While effects of perceptual sensitivity are straightforward to interpret as the temporal reliability or precision of neural activity in sensory systems, interpretation of bias is less straightforward. This is true, even though we tried to limit the scope of potential biases by counter-balancing response keys and intermixing trial types, which should minimize motor biases, as well as differences in prior expectations or strategy based on task context.

For example, in chapter 2 we demonstrated that both fusion and fission illusion are associated with a bias toward reporting the perceived number of beeps (in line with previous work). However, the underlying neural mechanism for this could be both cognitive (i.e. simply report the number of beeps, irrespective of visual evidence) or perceptual (i.e. both signal and noise are shifted by concurrent beeps, while criterion stays constant). Although the latter (perceptual bias) is likely only a good explanation in detection tasks as it is related to overall neural excitability, which should not change performance in discrimination tasks (Iemi et al., 2017). While our task can be construed as a detection task of the second flash or a gap in-between two flashes this is likely not an entirely accurate description. There is no reason that overall excitability should lead to improved temporal discrimination per se, but rather a stronger perceptual outcome, i.e. if stimuli are bound a brighter flash is elicited in a high excitability state. As such, bias likely represents response bias toward the perceived number of beeps or a generic decisional bias favouring either one or two flashes. Given that participants were biased

toward reporting the number perceived beeps in the auditory conditions, this strongly suggests decisional multimodal biases in fission and fusion.

While intermixing different trial types (unimodal and multimodal flash-beep combinations) is a good safeguard against differences in strategy or prior expectations between different conditions (0-2 beeps), it also hampers comparability with previous work where perceptual context was less variable (Cecere et al., 2015; Samaha & Postle, 2015). For example, our findings regarding alpha frequency strongly argue against a generic relationship between alpha frequency and visual temporal resolution (i.e. sensitivity of neural processes), but are limited in their conclusiveness with respect to task strategy, which should mostly manifest in the bias parameter and might strongly depend on task context, especially when designs are unbalanced (Cecere et al., 2015). In the same context, it is important to note that while perceptual resolution is expected to play a similar role throughout different audio-visual one versus two flash segregation settings, multimodal decisional bias would be expected to differ between audio-visual contexts. Hence, if alpha frequency is related to the bias parameter, it likely does not reflect multimodal bias, but rather a generic bias of reporting more temporal events. This hypothesis fits well with alpha frequency effects found in tasks where top-down expectations are likely to determine the perceptual outcome (Ronconi et al., 2018; Shen, Han, Chen, & Chen, 2019; Wutz et al., 2018). Yet, it cannot be properly tested with our study design optimized for investigating bottom-up perceptual influences.

Recently, it was discovered that an abstract decision variable can be measured with EEG over centro-parietal areas in the form of the centro-parietal positivity (CPP) when sensory stimulation is continuously updated (Kelly & O'Connell, 2013; O'Connell et al., 2012). In addition, the classical late component of the P300 was shown to capture a decision variable when evidence is transient, task relevant and rare (Twomey et al., 2015). These results are

supplemented by research of a simple perceptual decision-making task, in which decoding analyses demonstrated that sensory input representations are sustained when task relevant and followed in time by a representation of the perceptual decision, even when stimuli are transient and frequent (Mostert et al., 2015).

Inspired by these discoveries, in chapter 3, we harnessed the power of multivariate pattern analysis to track the neuronal activity leading up to perceptual outcomes, and showed that fission (i.e. illusory) flash representations differ significantly from real flash encoding. Interestingly, “see 2” ERPs showed overall higher amplitude compared to “see 1” ERPs, especially over occipital, parietal and central regions. Speculatively, this could reflect a decision variable, which in the case of a detection task would be expected to reach a higher value when a stimulus is indeed detected. We hope that future studies will supplement these results with MEG, fMRI and neurophysiological recordings to illuminate the underlying spatial sources and neuronal population characteristics. An intriguing question that we ourselves want to investigate is whether the support vector machines trained on physical flash presentations (e.g. distinguish one from two physical flashes, given that participants reported “see 1”) are able to generalize to perceptual outcomes (e.g. distinguish “see 1” from “see 2”, given that two flashes were presented). This analysis would test the similarity of the representation of a “see 1” versus “see 2” response to that of physically evoked activity by one or two flashes. Another interesting avenue for future research would be to present fixed SOAs to all participants, allowing an investigation of early perceptual processes with high temporal resolution using MVPA. Such an endeavour would be challenging, however, given the variability in illusion susceptibility between participants with regard to SOA.

In chapter 4 we addressed the long-standing questions of whether (pre-stimulus) alpha oscillations index cortical excitability (power) and temporal frames (frequency). Surprisingly

we did not replicate previous work on either power (Iemi et al., 2017) or frequency (Cecere et al., 2015; Samaha & Postle, 2015), but rather found substantial evidence in favour of the null hypotheses. This is not a trivial contribution as both alpha power and frequency have been implicated in perceptual detection tasks many times (Iemi et al., 2017; Joachim Lange, Keil, Schnitzler, van Dijk, & Weisz, 2014; VanRullen, 2018), but it is unclear exactly how often they have not been. Due to publication bias null results are (probably vastly) underreported, an issue that fortunately is acknowledged by the community (VanRullen, 2018). Moreover, many studies frequently focus on a single alpha oscillation characteristic (power, phase, frequency), even though they are interdependent (Jensen et al., 2014; Nelli et al., 2017). In addition, absences of effects are not always quantified statistically (Gulbinaite et al., 2017). Again, however, we see an encouraging change in this pattern, which might be partially due to the increased use and accessibility of software and theory for computing Bayes factors (Baumgarten et al., 2017; Iemi et al., 2017; Keil & Senkowski, 2017), as well as more general efforts to promote replications and open science (Munafò et al., 2017). Intriguingly, a recent study showed that alpha power and phase prior to stimulus onset are correlated with the common source prior, i.e. participants' prior belief about whether beeps and flashes come from the same outside source (Rohe et al., 2018), providing yet another perspective on how alpha oscillations influence (multisensory) visual perception. Another promising recent development in the field is the acknowledgement that brain oscillations in specific frequency bands, such as the alpha rhythm, are i. embedded in a broad-band background signal that is functionally relevant (Haller et al., 2018), and ii. do not necessarily exhibit sinusoidal properties, but more complex recurrent wave patterns (Cole & Voytek, 2017, 2018). We look forward to future work disentangling the contributions of alpha oscillations and its putative sub-components on passive (audio-) visual stimulation tasks. To this end we encourage researchers to take a holistic view

and consider all characteristics of the alpha rhythm, as well as to embrace new methodological developments (Cole & Voytek, 2018; Haller et al., 2018).

In addition to tracking the neural correlates of the perceptual decision-making process in time, we characterized participants' perceptual sensitivity as a function of stimulus onset asynchrony (chapter 2). We thereby corroborate and expand previous findings that show stronger perceptual sensitivity changes for fission compared to unisensory and fusion contexts, especially when flashes are displayed in the periphery (Chen et al., 2017; Kaposvári et al., 2014; McCormick & Mamassian, 2008; Mishra et al., 2008; Vanes et al., 2016). An interesting and logical next step would be to characterize the temporal determinants of fission and fusion when flashes and beeps are presented foveally, as visual field location of the flashes affects perceptual sensitivity differently between the illusions when SOA is fixed (Chen et al., 2017; Kaposvári et al., 2014). Similarly, it would be intriguing to repeat the MVPA and ERP analyses conducted in chapter 3 to study centrally presented flashes, given that there is top-down connectivity between both auditory and polysensory cortices and peripheral, but not central V1 (Clavagnier et al., 2004; Falchier et al., 2002).

A strong point of the psychophysical analyses in this thesis is that we carefully measured perceptual performance in several ways using psychometric function and two equal-variance Gaussian signal detection models, aiming to segregate perceptual sensitivity from bias. This allowed a more thorough representation of the data than any single approach, and highlights the usefulness of conducting multiple, complementary analyses. If we had only used psychometric functions to describe the temporal binding window, for example, interpretation of threshold and slope would have been problematic, as both parameters can be influenced by biases (García-Pérez & Alcalá-Quintana, 2011; Klein, 2001). Similarly, it was valuable to conduct analyses on response times to take into account the speed-accuracy trade-off. Future work should aim to

incorporate response times into the Bayesian causal inference model, for example by integrating it with drift diffusion accounts. Such an approach would allow for all behavioural data to feed into computational model fitting and provide more realistic insights garnered from estimated parameters. It could be coupled with MVPA and neuroimaging techniques to track the time courses and brain regions that represent different computational processes, as has been done previously (Aller & Noppeney, 2019; Rohe et al., 2018; Rohe & Noppeney, 2015a).

Conclusion

In this thesis we examined the multisensory perceptual decision-making process of resolving one versus two flashes in time in the context of variable numbers of acoustic stimuli (0-2 beeps). We investigated the temporal evolution of this decision making-process from its start (prior to stimulus presentation) to its conclusion (behavioural response) by probing EEG activity. In particular, pre-stimulus alpha band oscillations were considered as a putative substrate of visual temporal resolution (alpha frequency) and cortical excitability (alpha power); and the time courses of perceptual outcomes were tracked and compared between conditions by analysing multivariate EEG patterns at each time point. We demonstrated that the fission illusion has a wider temporal binding window compared to the fusion illusion or unisensory two-flash fusion. In the pre-stimulus period neither alpha power nor frequency contributed to the subsequent processing of flash-beep pairings. Similarly, decoding of perceptual outcomes was only possible after stimulus presentation. Remarkably, the neural patterns of “see 1” versus “see 2” reports were substantially similar (i.e. generalized) between unisensory two flash presentations and the fusion condition (2-flash, 1-beep), and between the fusion and fission condition (1-flash, 2-beep), but not between unisensory two flash presentations and the fission condition. This suggests, that even though perceptual outcomes are represented in very similar ways between the fusion and fission contexts, there are also differences that make the neural

representations comparable to unisensory representations in the fusion, but not the fission condition. Such subtle differences might be leveraged for metacognitive judgments, which are close to optimal in fission, but not fusion. Furthermore, it might be exploited by participants that get monetarily rewarded for learning how to distinguish illusory double flashes from real ones (Rosenthal et al., 2009).

References

- Abadi, R. V., & Murphy, J. S. (2014). Phenomenology of the sound-induced flash illusion. *Experimental Brain Research*, 232(7), 2207–2220. <https://doi.org/10.1007/s00221-014-3912-2>
- Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, 14(3), 257–262. <https://doi.org/10.1016/j.cub.2004.01.029>
- Alais, D., Newell, F., & Mamassian, P. (2010). *Multisensory Processing in Review: From Physiology to Behaviour* (Vol. 23). <https://doi.org/10.1163/187847510X488603>
- Aller, M., & Noppeney, U. (2019). To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian Causal Inference. *BioRxiv*, 504118. <https://doi.org/10.1101/504118>
- Alvarado, J. C., Stanford, T. R., Vaughan, J. W., & Stein, B. E. (2007). Cortex Mediates Multisensory But Not Unisensory Integration in Superior Colliculus. *Journal of Neuroscience*, 27(47), 12775–12786. <https://doi.org/10.1523/JNEUROSCI.3524-07.2007>
- Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, 21(3), 301–308. <https://doi.org/10.1016/j.cogbrainres.2004.06.004>
- Apthorp, D., Alais, D., & Boenke, L. T. (2013). Flash illusions induced by visual, auditory, and audiovisual stimuli. *Journal of Vision*, 13(5), 3–3. <https://doi.org/10.1167/13.5.3>
- Arnold, D. H., & Yarrow, K. (2011). Temporal recalibration of vision. *Proceedings of the Royal Society B: Biological Sciences*, 278(1705), 535–538. <https://doi.org/10.1098/rspb.2010.1396>

- Baumgarten, T. J., Schnitzler, A., & Lange, J. (2017). Beyond the Peak – Tactile Temporal Discrimination Does Not Correlate with Individual Peak Frequencies in Somatosensory Cortex. *Frontiers in Psychology*, 8.
<https://doi.org/10.3389/fpsyg.2017.00421>
- Bennur, S., & Gold, J. I. (2011). Distinct Representations of a Perceptual Decision and the Associated Oculomotor Plan in the Monkey Lateral Intraparietal Area. *Journal of Neuroscience*, 31(3), 913–921. <https://doi.org/10.1523/JNEUROSCI.4417-10.2011>
- Benwell, C. S. Y., London, R. E., Tagliabue, C. F., Veniero, D., Gross, J., Keitel, C., & Thut, G. (2019). Frequency and power of human alpha oscillations drift systematically with time-on-task. *NeuroImage*, 192, 101–114.
<https://doi.org/10.1016/j.neuroimage.2019.02.067>
- Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87(1), 527–570. <https://doi.org/10.1007/BF01797193>
- Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, 29(6), 578–584.
<https://doi.org/10.3758/BF03207374>
- Bidelman, G. M. (2016). Musicians have enhanced audiovisual multisensory binding: Experience-dependent effects in the double-flash illusion. *Experimental Brain Research*, 234(10), 3037–3047. <https://doi.org/10.1007/s00221-016-4705-6>
- Bidelman, G. M., & Heath, S. T. (2018). Enhanced temporal binding of audiovisual information in the bilingual brain. *Bilingualism: Language and Cognition*, 1–11.
<https://doi.org/10.1017/S1366728918000408>

- Bollimunta, A., Mo, J., Schroeder, C. E., & Ding, M. (2011). Neuronal Mechanisms and Attentional Modulation of Corticothalamic Alpha Oscillations. *Journal of Neuroscience*, 31(13), 4935–4943. <https://doi.org/10.1523/JNEUROSCI.5580-10.2011>
- Bolognini, N., Convento, S., Casati, C., Mancini, F., Brighina, F., & Vallar, G. (2016). Multisensory integration in hemianopia and unilateral spatial neglect: Evidence from the sound induced flash illusion. *Neuropsychologia*, 87(Supplement C), 134–143. <https://doi.org/10.1016/j.neuropsychologia.2016.05.015>
- Bolognini, N., Rossetti, A., Casati, C., Mancini, F., & Vallar, G. (2011). Neuromodulation of multisensory perception: A tDCS study of the sound-induced flash illusion. *Neuropsychologia*, 49(2), 231–237. <https://doi.org/10.1016/j.neuropsychologia.2010.11.015>
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652. <https://doi.org/10.1037/0033-295X.108.3.624>
- Bruns, P., Maiworm, M., & Röder, B. (2014). Reward expectation influences audiovisual spatial integration. *Attention, Perception, & Psychophysics*, 76(6), 1815–1827. <https://doi.org/10.3758/s13414-014-0699-y>
- Carriere, B. N., Royal, D. W., & Wallace, M. T. (2008). Spatial Heterogeneity of Cortical Receptive Fields and Its Impact on Multisensory Interactions. *Journal of Neurophysiology*, 99(5), 2357–2368. <https://doi.org/10.1152/jn.01386.2007>
- Cecere, R., Rees, G., & Romei, V. (2015). Individual Differences in Alpha Frequency Drive Crossmodal Illusory Perception. *Current Biology*, 25(2), 231–235. <https://doi.org/10.1016/j.cub.2014.11.034>

- Chen, Y.-C., Maurer, D., Lewis, T. L., Spence, C., & Shore, D. I. (2017). Central–peripheral differences in audiovisual and visuotactile event perception. *Attention, Perception, & Psychophysics*, 79(8), 2552–2563. <https://doi.org/10.3758/s13414-017-1396-4>
- Chiang, A. K. I., Rennie, C. J., Robinson, P. A., Roberts, J. A., Rigozzi, M. K., Whitehouse, R. W., ... Gordon, E. (2008). Automated characterization of multiple alpha peaks in multi-site electroencephalograms. *Journal of Neuroscience Methods*, 168(2), 396–411. <https://doi.org/10.1016/j.jneumeth.2007.11.001>
- Chiang, A. K. I., Rennie, C. J., Robinson, P. A., van Albada, S. J., & Kerr, C. C. (2011). Age trends and sex differences of alpha rhythms including split alpha peaks. *Clinical Neurophysiology*, 122(8), 1505–1517. <https://doi.org/10.1016/j.clinph.2011.01.040>
- Churchland, Anne. K., Kiani, R., Chaudhuri, R., Wang, X.-J., Pouget, A., & Shadlen, M. N. (2011). Variance as a Signature of Neural Computations during Decision Making. *Neuron*, 69(4), 818–831. <https://doi.org/10.1016/j.neuron.2010.12.037>
- Clavagnier, S., Falchier, A., & Kennedy, H. (2004). Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive, Affective, & Behavioral Neuroscience*, 4(2), 117–126. <https://doi.org/10.3758/CABN.4.2.117>
- Cohen, M. X. (2014). Fluctuations in Oscillation Frequency Control Spike Timing and Coordinate Neural Networks. *Journal of Neuroscience*, 34(27), 8988–8998. <https://doi.org/10.1523/JNEUROSCI.0261-14.2014>
- Cole, S. R., & Voytek, B. (2017). Brain Oscillations and the Importance of Waveform Shape. *Trends in Cognitive Sciences*, 21(2), 137–149. <https://doi.org/10.1016/j.tics.2016.12.008>

- Cole, S. R., & Voytek, B. (2018). Cycle-by-cycle analysis of neural oscillations. *BioRxiv*, 302000. <https://doi.org/10.1101/302000>
- Corcoran, A. W., Alday, P. M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2018). Toward a reliable, automated method of individual alpha frequency (IAF) quantification. *Psychophysiology*, 55(7), e13064. <https://doi.org/10.1111/psyp.13064>
- Cuppini, C., Magosso, E., Bolognini, N., Vallar, G., & Ursino, M. (2014). A neurocomputational analysis of the sound-induced flash illusion. *NeuroImage*, 92, 248–266. <https://doi.org/10.1016/j.neuroimage.2014.02.001>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG Sources Are Dipolar. *PLOS ONE*, 7(2), e30135. <https://doi.org/10.1371/journal.pone.0030135>
- DeLoss, D. J., & Andersen, G. J. (2015). Aging, Spatial Disparity, and the Sound-Induced Flash Illusion. *PLOS ONE*, 10(11), e0143773. <https://doi.org/10.1371/journal.pone.0143773>
- Deroy, O., Spence, C., & Noppeney, U. (2016). Metacognition in Multisensory Perception. *Trends in Cognitive Sciences*, 20(10), 736–747. <https://doi.org/10.1016/j.tics.2016.08.006>
- Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, 66(8), 1388–1404. <https://doi.org/10.3758/BF03195006>

- Ding, L. (2015). Distinct dynamics of ramping activity in the frontal cortex and caudate nucleus in monkeys. *Journal of Neurophysiology*, 114(3), 1850–1861.
<https://doi.org/10.1152/jn.00395.2015>
- Driver, J., & Noesselt, T. (2008). Multisensory Interplay Reveals Crossmodal Influences on ‘Sensory-Specific’ Brain Regions, Neural Responses, and Judgments. *Neuron*, 57(1), 11–23. <https://doi.org/10.1016/j.neuron.2007.12.013>
- Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., & Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *ELife*, 3, e03005.
<https://doi.org/10.7554/eLife.03005>
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4), 162–169. <https://doi.org/10.1016/j.tics.2004.02.002>
- Evans, A. C., Collins, D. L., Milner, B., & Milner, B. (1992). *An MRI-based stereotactic atlas from 250 young normal subjects*. Retrieved from
<https://www.scienceopen.com/document?vid=6a571d6f-66b5-4a26-8962-418725b4bcbc>
- Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical Evidence of Multimodal Integration in Primate Striate Cortex. *Journal of Neuroscience*, 22(13), 5749–5759. <https://doi.org/10.1523/JNEUROSCI.22-13-05749.2002>
- 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. Experimentelle Hirnforschung. Experimentation Cerebrale*, 203(2), 381–389. <https://doi.org/10.1007/s00221-010-2240-4>

- Freyer, F., Roberts, J. A., Becker, R., Robinson, P. A., Ritter, P., & Breakspear, M. (2011). Biophysical Mechanisms of Multistability in Resting-State Cortical Rhythms. *Journal of Neuroscience*, 31(17), 6353–6361. <https://doi.org/10.1523/JNEUROSCI.6693-10.2011>
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, 7(7), 773–778. <https://doi.org/10.1038/nn1268>
- Gandhi, N. J., & Katnani, H. A. (2011). Motor Functions of the Superior Colliculus. *Annual Review of Neuroscience*, 34(1), 205–231. <https://doi.org/10.1146/annurev-neuro-061010-113728>
- García-Pérez, M. A., & Alcalá-Quintana, R. (2011). Improving the Estimation of Psychometric Functions in 2AFC Discrimination Tasks. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00096>
- García-Pérez, M. A., & Alcalá-Quintana, R. (2011). Interval bias in 2AFC detection tasks: Sorting out the artifacts. *Attention, Perception, & Psychophysics*, 73(7), 2332–2352. <https://doi.org/10.3758/s13414-011-0167-x>
- Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *NeuroImage*, 124(Part A), 876–886. <https://doi.org/10.1016/j.neuroimage.2015.09.045>
- Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory Integration of Dynamic Faces and Voices in Rhesus Monkey Auditory Cortex. *Journal of Neuroscience*, 25(20), 5004–5012. <https://doi.org/10.1523/JNEUROSCI.0799-05.2005>
- Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10(6), 278–285. <https://doi.org/10.1016/j.tics.2006.04.008>

- Gho, M., & Varela, F. J. (1988). A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *Journal de Physiologie*, 83(2), 95–101.
Retrieved from Scopus.
- Ghose, D., & Wallace, M. T. (2014). Heterogeneity in the spatial receptive field architecture of multisensory neurons of the superior colliculus and its effects on multisensory integration. *Neuroscience*, 256, 147–162.
<https://doi.org/10.1016/j.neuroscience.2013.10.044>
- Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews. Neuroscience*, 14(5). <https://doi.org/10.1038/nrn3476>
- Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The Neural Basis of Perceptual Learning. *Neuron*, 31(5), 681–697. [https://doi.org/10.1016/S0896-6273\(01\)00424-X](https://doi.org/10.1016/S0896-6273(01)00424-X)
- Gold, J. I., & Ding, L. (2013). How mechanisms of perceptual decision-making affect the psychometric function. *Progress in Neurobiology*, 103, 98–114.
<https://doi.org/10.1016/j.pneurobio.2012.05.008>
- Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, 36(2), 299–308.
- Gorea, A. (2015). A Refresher of the Original Bloch's Law Paper (Bloch, July 1885). *I-Perception*, 6(4), 2041669515593043. <https://doi.org/10.1177/2041669515593043>
- Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2016). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, 29(4), 677–697.
https://doi.org/10.1162/jocn_a_01068

- Gulbinaite, R., İlhan, B., & VanRullen, R. (2017). The Triple-Flash Illusion Reveals a Driving Role of Alpha-Band Reverberations in Visual Perception. *Journal of Neuroscience*, 37(30), 7219–7230. <https://doi.org/10.1523/JNEUROSCI.3929-16.2017>
- Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J., & Nobre, A. C. (2014). Inter- and intra-individual variability in alpha peak frequency. *NeuroImage*, 92(Supplement C), 46–55. <https://doi.org/10.1016/j.neuroimage.2014.01.049>
- Haller, M., Donoghue, T., Peterson, E., Varma, P., Sebastian, P., Gao, R., ... Voytek, B. (2018). Parameterizing neural power spectra. *BioRxiv*, 299859. <https://doi.org/10.1101/299859>
- Hamilton, R. H., Wiener, M., Drebing, D. E., & Coslett, H. B. (2013). Gone in a flash: Manipulation of audiovisual temporal integration using transcranial magnetic stimulation. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00571>
- Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, 520(7546), 220–223. <https://doi.org/10.1038/nature14066>
- Hanks, T. D., & Summerfield, C. (2017). Perceptual Decision Making in Rodents, Monkeys, and Humans. *Neuron*, 93(1), 15–31. <https://doi.org/10.1016/j.neuron.2016.12.003>
- Hanslmayr, S., Gross, J., Klimesch, W., & Shapiro, K. L. (2011). The role of alpha oscillations in temporal attention. *Brain Research Reviews*, 67(1), 331–343. <https://doi.org/10.1016/j.brainresrev.2011.04.002>
- Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, 8. <https://doi.org/10.3389/fnins.2014.00150>

- Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, 298(5598), 1627–1630.
- Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia*, 49(3), 461–467. <https://doi.org/10.1016/j.neuropsychologia.2010.11.041>
- Hirsh, I. J., & Sherrick Jr., C. E. (1961a). Perceived order in different sense modalities. *Journal of Experimental Psychology*, 62(5), 423–432. <https://doi.org/10.1037/h0045283>
- Hirsh, I. J., & Sherrick Jr., C. E. (1961b). Perceived order in different sense modalities. *Journal of Experimental Psychology*, 62(5), 423–432. <https://doi.org/10.1037/h0045283>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics. Theory and Applications*, 6(2), 65–70.
- Hughes, S. W., Lörincz, M., Cope, D. W., Blethyn, K. L., Kékesi, K. A., Parri, H. R., ... Crunelli, V. (2004). Synchronized Oscillations at α and θ Frequencies in the Lateral Geniculate Nucleus. *Neuron*, 42(2), 253–268. [https://doi.org/10.1016/S0896-6273\(04\)00191-6](https://doi.org/10.1016/S0896-6273(04)00191-6)
- Iemi, L., Chaumon, M., Crouzet, S. M., & Busch, N. A. (2017). Spontaneous Neural Oscillations Bias Perception by Modulating Baseline Excitability. *Journal of Neuroscience*, 37(4), 807–819. <https://doi.org/10.1523/JNEUROSCI.1432-16.2016>
- Innes-Brown, H., & Crewther, D. (2009). The Impact of Spatial Incongruence on an Auditory-Visual Illusion. *PLOS ONE*, 4(7), e6450. <https://doi.org/10.1371/journal.pone.0006450>

- Jeffreys, S. H. (1998). *The Theory of Probability* (Third Edition). Oxford, New York: Oxford University Press.
- Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences*, 16(4), 200–206. <https://doi.org/10.1016/j.tics.2012.03.002>
- Jensen, O., Gips, B., Bergmann, T. O., & Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends in Neurosciences*, 37(7), 357–369. <https://doi.org/10.1016/j.tins.2014.04.001>
- Jiang, W., Jiang, H., & Stein, B. E. (2006). Neonatal Cortical Ablation Disrupts Multisensory Development in Superior Colliculus. *Journal of Neurophysiology*, 95(3), 1380–1396. <https://doi.org/10.1152/jn.00880.2005>
- Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W., & Stein, B. E. (2001). Two Cortical Areas Mediate Multisensory Integration in Superior Colliculus Neurons. *Journal of Neurophysiology*, 85(2), 506–522. <https://doi.org/10.1152/jn.2001.85.2.506>
- Kaposvári, P., Bognár, A., Csibri, P., Utassy, G., & Sárosi, G. (2014). Fusion and fission in the visual pathways. *Physiological Research/Academia Scientiarum Bohemoslovaca*, 63(5), 625–635.
- Katz, L. N., Yates, J. L., Pillow, J. W., & Huk, A. C. (2016). Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature*, 535(7611), 285–288. <https://doi.org/10.1038/nature18617>
- Kayser, C., & Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Structure and Function*, 212(2), 121–132. <https://doi.org/10.1007/s00429-007-0154-0>

- Keil, J., Müller, N., Hartmann, T., & Weisz, N. (2014). Prestimulus Beta Power and Phase Synchrony Influence the Sound-Induced Flash Illusion. *Cerebral Cortex*, 24(5), 1278–1288. <https://doi.org/10.1093/cercor/bhs409>
- Keil, J., & Senkowski, D. (2017). Individual Alpha Frequency Relates to the Sound-Induced Flash Illusion. *Multisensory Research*, 30. <https://doi.org/10.1163/22134808-00002572>
- Kelly, S. P., & O’Connell, R. G. (2013). Internal and External Influences on the Rate of Sensory Evidence Accumulation in the Human Brain. *Journal of Neuroscience*, 33(50), 19434–19441. <https://doi.org/10.1523/JNEUROSCI.3355-13.2013>
- Kingdom, F. A. A., & Prins, N. (2016). Chapter 4—Psychometric Functions*. In F. A. A. Kingdom & N. Prins (Eds.), *Psychophysics (Second Edition)* (pp. 55–117). <https://doi.org/10.1016/B978-0-12-407156-8.00004-9>
- Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, 63(8), 1421–1455. <https://doi.org/10.3758/BF03194552>
- Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Research Reviews*, 53(1), 63–88. <https://doi.org/10.1016/j.brainresrev.2006.06.003>
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE*, 2(9), e943. <https://doi.org/10.1371/journal.pone.0000943>
- Kristofferson, A. B. (1967). Successiveness Discrimination as a Two-State, Quantal Process. *Science*, 158(3806), 1337–1339. <https://doi.org/10.1126/science.158.3806.1337>

- Krueger, J., Royal, D. W., Fister, M. C., & Wallace, M. T. (2009). Spatial receptive field organization of multisensory neurons and its impact on multisensory interactions. *Hearing Research*, 258(1), 47–54. <https://doi.org/10.1016/j.heares.2009.08.003>
- Kumpik, D. P., Roberts, H. E., King, A. J., & Bizley, J. K. (2014). Visual sensitivity is a stronger determinant of illusory processes than auditory cue parameters in the sound-induced flash illusion. *Journal of Vision*, 14(7), 12–12. <https://doi.org/10.1167/14.7.12>
- Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, 384(6604), 74. <https://doi.org/10.1038/384074a0>
- Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. *Neuron*, 53(2), 279–292. <https://doi.org/10.1016/j.neuron.2006.12.011>
- Lange, J., Oostenveld, R., & Fries, P. (2013). Reduced Occipital Alpha Power Indexes Enhanced Excitability Rather than Improved Visual Perception. *Journal of Neuroscience*, 33(7), 3212–3220. <https://doi.org/10.1523/JNEUROSCI.3755-12.2013>
- Lange, Joachim, Keil, J., Schnitzler, A., van Dijk, H., & Weisz, N. (2014). The role of alpha oscillations for illusory perception. *Behavioural Brain Research*, 271, 294–301. <https://doi.org/10.1016/j.bbr.2014.06.015>
- Law, C.-T., & Gold, J. I. (2010). Shared Mechanisms of Perceptual Learning and Decision Making. *Topics in Cognitive Science*, 2(2), 226–238. <https://doi.org/10.1111/j.1756-8765.2009.01044.x>
- Liu, C. C., & Watanabe, T. (2012). Accounting for speed-accuracy tradeoff in perceptual learning. *Vision Research*, 61, 107–114. <https://doi.org/10.1016/j.visres.2011.09.007>
- Lopes da Silva, F. H., van Lierop, T. H. M. T., Schrijer, C. F., & Storm van Leeuwen, W. (1973). Organization of thalamic and cortical alpha rhythms: Spectra and coherences.

- Electroencephalography and Clinical Neurophysiology*, 35(6), 627–639.
[https://doi.org/10.1016/0013-4694\(73\)90216-2](https://doi.org/10.1016/0013-4694(73)90216-2)
- Lőrincz, M. L., Kékesi, K. A., Juhász, G., Crunelli, V., & Hughes, S. W. (2009). Temporal Framing of Thalamic Relay-Mode Firing by Phasic Inhibition during the Alpha Rhythm. *Neuron*, 63(5), 683–696. <https://doi.org/10.1016/j.neuron.2009.08.012>
- Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, 54(1), 146–157.
<https://doi.org/10.1111/psyp.12639>
- Luo, T. Z., & Maunsell, J. H. R. (2015). Neuronal Modulations in Visual Cortex Are Associated with Only One of Multiple Components of Attention. *Neuron*, 86(5), 1182–1188. <https://doi.org/10.1016/j.neuron.2015.05.007>
- Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The Curious Incident of Attention in Multisensory Integration: Bottom-up vs. Top-down. *Multisensory Research*, 29(6–7), 557–583.
<https://doi.org/10.1163/22134808-00002528>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed). Mahwah, N.J: Lawrence Erlbaum Associates.
- Maiworm, M., Bellantoni, M., Spence, C., & Röder, B. (2012). When emotional valence modulates audiovisual integration. *Attention, Perception, & Psychophysics*, 74(6), 1302–1311. <https://doi.org/10.3758/s13414-012-0310-3>
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, 21(1), 422–430. <https://doi.org/10.1016/j.concog.2011.09.021>

- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.
<https://doi.org/10.1016/j.jneumeth.2007.03.024>
- McCormick, D., & Mamassian, P. (2008). What does the illusory-flash look like? *Vision Research*, 48(1), 63–69. <https://doi.org/10.1016/j.visres.2007.10.010>
- McGovern, D. P., Roudaia, E., Stapleton, J., McGinnity, T. M., & Newell, F. N. (2014). The sound-induced flash illusion reveals dissociable age-related effects in multisensory integration. *Frontiers in Aging Neuroscience*, 6.
<https://doi.org/10.3389/fnagi.2014.00250>
- Mercier, M. R., Foxe, J. J., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., & Molholm, S. (2013). Auditory-driven phase reset in visual cortex: Human electrocorticography reveals mechanisms of early multisensory integration. *NeuroImage*, 79, 19–29.
<https://doi.org/10.1016/j.neuroimage.2013.04.060>
- 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), 3215–3229. <https://doi.org/10.1523/JNEUROSCI.07-10-03215.1987>
- Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, 221(4608), 389–391.
<https://doi.org/10.1126/science.6867718>
- Meredith, M. A., & Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, 365(2), 350–354.
[https://doi.org/10.1016/0006-8993\(86\)91648-3](https://doi.org/10.1016/0006-8993(86)91648-3)

- Meylan, R. V., & Murray, M. M. (2007). Auditory–visual multisensory interactions attenuate subsequent visual responses in humans. *NeuroImage*, 35(1), 244–254.
<https://doi.org/10.1016/j.neuroimage.2006.11.033>
- Minami, S., & Amano, K. (2017). Illusory Jitter Perceived at the Frequency of Alpha Oscillations. *Current Biology*, 27(15), 2344–2351.e4.
<https://doi.org/10.1016/j.cub.2017.06.033>
- Mishra, J., Martinez, A., & Hillyard, S. A. (2008). Cortical processes underlying sound-induced flash fusion. *Brain Research*, 1242, 102–115.
<https://doi.org/10.1016/j.brainres.2008.05.023>
- Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early Cross-Modal Interactions in Auditory and Visual Cortex Underlie a Sound-Induced Visual Illusion. *Journal of Neuroscience*, 27(15), 4120–4131.
<https://doi.org/10.1523/JNEUROSCI.4912-06.2007>
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research*, 14(1), 115–128.
[https://doi.org/10.1016/S0926-6410\(02\)00066-6](https://doi.org/10.1016/S0926-6410(02)00066-6)
- Morrell, F. (1972). Visual System’s View of Acoustic Space. *Nature*, 238(5358), 44.
<https://doi.org/10.1038/238044a0>
- Mostert, P., Kok, P., & de Lange, F. P. (2015). Dissociating sensory from decision processes in human perceptual decision making. *Scientific Reports*, 5, 18253.
<https://doi.org/10.1038/srep18253>
- Müller, B. C. N., Tsalas, N. R. H., van Schie, H. T., Meinhardt, J., Proust, J., Sodian, B., & Paulus, M. (2016). Neural correlates of judgments of learning – An ERP study on

- metacognition. *Brain Research*, 1652, 170–177.
<https://doi.org/10.1016/j.brainres.2016.10.005>
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., ... Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 0021. <https://doi.org/10.1038/s41562-016-0021>
- Narinesingh, C., Goltz, H. C., & Wong, A. M. F. (2017). Temporal Binding Window of the Sound-Induced Flash Illusion in Amblyopia. *Investigative Ophthalmology & Visual Science*, 58(3), 1442–1448. <https://doi.org/10.1167/iovs.16-21258>
- Naruse, Y., Matani, A., Miyawaki, Y., & Okada, M. (2010). Influence of coherence between multiple cortical columns on alpha rhythm: A computational modeling study. *Human Brain Mapping*, 31(5), 703–715. <https://doi.org/10.1002/hbm.20899>
- Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7(4), 308–313. <https://doi.org/10.1093/comjnl/7.4.308>
- Nelli, S., Itthipuripat, S., Srinivasan, R., & Serences, J. T. (2017). Fluctuations in instantaneous frequency predict alpha amplitude during visual perception. *Nature Communications*, 8(1), 2071. <https://doi.org/10.1038/s41467-017-02176-x>
- Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H.-J., & Driver, J. (2010). Sound-Induced Enhancement of Low-Intensity Vision: Multisensory Influences on Human Sensory-Specific Cortices and Thalamic Bodies Relate to Perceptual Enhancement of Visual Detection Sensitivity. *Journal of Neuroscience*, 30(41), 13609–13623.
<https://doi.org/10.1523/JNEUROSCI.4524-09.2010>
- O’Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, 15(12), 1729–1735. <https://doi.org/10.1038/nn.3248>

- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data [Research article]. <https://doi.org/10.1155/2011/156869>
- Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, 10. <https://doi.org/10.3389/fninf.2016.00027>
- Park, I. M., Meister, M. L. R., Huk, A. C., & Pillow, J. W. (2014). Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nature Neuroscience*, 17(10), 1395–1403. <https://doi.org/10.1038/nn.3800>
- Philiastides, M. G., Heekeren, H. R., & Sajda, P. (2014). Human Scalp Potentials Reflect a Mixture of Decision-Related Signals during Perceptual Choices. *Journal of Neuroscience*, 34(50), 16877–16889. <https://doi.org/10.1523/JNEUROSCI.3012-14.2014>
- Pöppel, E. (1973). Comment on “Visual System’s View of Acoustic Space.” *Nature*, 243(5404), 231. <https://doi.org/10.1038/243231a0>
- Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual Training Narrows the Temporal Window of Multisensory Binding. *Journal of Neuroscience*, 29(39), 12265–12274. <https://doi.org/10.1523/JNEUROSCI.3501-09.2009>
- Prins, N., & Kingdom, F. A. A. (2018). Applying the Model-Comparison Approach to Test Specific Research Hypotheses in Psychophysical Research Using the Palamedes Toolbox. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01250>
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108. <https://doi.org/10.1037/0033-295X.85.2.59>

- Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, 20(4), 873–922.
<https://doi.org/10.1162/neco.2008.12-06-420>
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, 20(4), 260–281.
<https://doi.org/10.1016/j.tics.2016.01.007>
- Reeves, A. (1996). Chapter 1 Temporal resolution in visual perception. In W. Prinz & B. Bridgeman (Eds.), *Handbook of Perception and Action* (pp. 11–24).
[https://doi.org/10.1016/S1874-5822\(96\)80004-1](https://doi.org/10.1016/S1874-5822(96)80004-1)
- Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology*, 50(1), 19–26.
[https://doi.org/10.1016/S0167-8760\(03\)00121-1](https://doi.org/10.1016/S0167-8760(03)00121-1)
- Rohe, T., Ehrlis, A.-C., & Noppeney, U. (2018). The neural dynamics of hierarchical Bayesian inference in multisensory perception. *BioRxiv*, 504845. <https://doi.org/10.1101/504845>
- Rohe, T., & Noppeney, U. (2015a). Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLOS Biology*, 13(2), e1002073.
<https://doi.org/10.1371/journal.pbio.1002073>
- Rohe, T., & Noppeney, U. (2015b). Sensory reliability shapes perceptual inference via two mechanisms. *Journal of Vision*, 15(5), 22–22. <https://doi.org/10.1167/15.5.22>
- Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of Neuroscience*, 22(21), 9475–9489.

- Ronconi, L., Busch, N. A., & Melcher, D. (2018). Alpha-band sensory entrainment alters the duration of temporal windows in visual perception. *Scientific Reports*, 8(1), 11810. <https://doi.org/10.1038/s41598-018-29671-5>
- Ronconi, L., & Melcher, D. (2017). The Role of Oscillatory Phase in Determining the Temporal Organization of Perception: Evidence from Sensory Entrainment. *Journal of Neuroscience*, 37(44), 10636–10644. <https://doi.org/10.1523/JNEUROSCI.1704-17.2017>
- Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-Induced Flash Illusion is Resistant to Feedback Training. *Brain Topography*, 21(3–4), 185–192. <https://doi.org/10.1007/s10548-009-0090-9>
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *tests for accepting and rejecting the null hypothesis*. *Psychonomic Bulletin & Review*, 16(2), 225–237. <https://doi.org/10.3758/PBR.16.2.225>
- Royal, D. W., Carriere, B. N., & Wallace, M. T. (2009). Spatiotemporal architecture of cortical receptive fields and its impact on multisensory interactions. *Experimental Brain Research*, 198(2), 127–136. <https://doi.org/10.1007/s00221-009-1772-y>
- Ruzzoli, M., Torralba, M., Fernández, L. M., & Soto-Faraco, S. (2019). The relevance of alpha phase in human perception. *Cortex*. <https://doi.org/10.1016/j.cortex.2019.05.012>
- Samaha, J., & Postle, B. R. (2015). The Speed of Alpha-Band Oscillations Predicts the Temporal Resolution of Visual Perception. *Current Biology*, 25(22), 2985–2990. <https://doi.org/10.1016/j.cub.2015.10.007>
- Sauseng, P., Klimesch, W., Gruber, W. R., Hanslmayr, S., Freunberger, R., & Doppelmayr, M. (2007). Are event-related potential components generated by phase resetting of

- brain oscillations? A critical discussion. *Neuroscience*, 146(4), 1435–1444.
<https://doi.org/10.1016/j.neuroscience.2007.03.014>
- Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences*, 109(42), E2904–E2913. <https://doi.org/10.1073/pnas.1210467109>
- Shams, L. (2012). Early Integration and Bayesian Causal Inference in Multisensory Perception. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. Retrieved from
<http://www.ncbi.nlm.nih.gov/books/NBK92847/>
- Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, 14(9), 425–432. <https://doi.org/10.1016/j.tics.2010.07.001>
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, 408(6814), 788. <https://doi.org/10.1038/35048669>
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14(1), 147–152. [https://doi.org/10.1016/S0926-6410\(02\)00069-1](https://doi.org/10.1016/S0926-6410(02)00069-1)
- Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *NeuroReport*, 16(17), 1923–1927.
- Shen, L., Han, B., Chen, L., & Chen, Q. (2019). Perceptual inference employs intrinsic alpha frequency to resolve perceptual ambiguity. *PLOS Biology*, 17(3), e3000025.
<https://doi.org/10.1371/journal.pbio.3000025>
- Sigala, R., Haufe, S., Roy, D., Dinse, H. R., & Ritter, P. (2014). The role of alpha-rhythm states in perceptual learning: Insights from experiments and computational models. *Frontiers in Computational Neuroscience*, 8.
<https://doi.org/10.3389/fncom.2014.00036>

- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27(3), 161–168. <https://doi.org/10.1016/j.tins.2004.01.006>
- Snoek, L., Miletić, S., & Scholte, H. S. (2019). How to control for confounds in decoding analyses of neuroimaging data. *NeuroImage*, 184, 741–760. <https://doi.org/10.1016/j.neuroimage.2018.09.074>
- Spence, C. (2018). Multisensory Perception. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–56). <https://doi.org/10.1002/9781119170174.epcn214>
- Stankevich, L. A., Sonkin, K. M., Shemyakina, N. V., Nagornova, Zh. V., Khomenko, J. G., Perets, D. S., & Koval, A. V. (2016). EEG pattern decoding of rhythmic individual finger imaginary movements of one hand. *Human Physiology*, 42(1), 32–42. <https://doi.org/10.1134/S0362119716010175>
- Statistical Parametric Mapping: The Analysis of Functional Brain Images—1st Edition. (n.d.). Retrieved April 25, 2018, from <https://www.elsevier.com/books/statistical-parametric-mapping-the-analysis-of-functional-brain-images/penny/978-0-12-372560-8>
- Stein, B. E., Burr, D., Constantinidis, C., Laurienti, P. J., Meredith, M. A., Perrault, T. J., ... Lewkowicz, D. J. (2010). Semantic confusion regarding the development of multisensory integration: A practical solution. *The European Journal of Neuroscience*, 31(10), 1713–1720. <https://doi.org/10.1111/j.1460-9568.2010.07206.x>
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA, US: The MIT Press.
- Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews. Neuroscience*, 15(8), 520–535.

- Stein, B. E., Wallace, M. W., Stanford, T. R., & Jiang, W. (2002). Book Review: Cortex Governs Multisensory Integration in the Midbrain. *The Neuroscientist*, 8(4), 306–314. <https://doi.org/10.1177/107385840200800406>
- 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. <https://doi.org/10.1007/s00221-012-3072-1>
- Tsalas, N. R. H., Müller, B. C. N., Meinhardt, J., Proust, J., Paulus, M., & Sodian, B. (2018). An ERP study on metacognitive monitoring processes in children. *Brain Research*, 1695, 84–90. <https://doi.org/10.1016/j.brainres.2018.05.041>
- Twomey, D. M., Murphy, P. R., Kelly, S. P., & O’Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, 42(1), 1636–1643. <https://doi.org/10.1111/ejn.12936>
- Urbantschitsch, V. (1880). Über den Einfluss einer Sinneserregung auf die übrigen Sinnesempfindungen. *Archiv f d Gesch Physiol*, 42, 155.
- van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory Integration: Flexible Use of General Operations. *Neuron*, 81(6), 1240–1253. <https://doi.org/10.1016/j.neuron.2014.02.044>
- Vanes, L. D., White, T. P., Wigton, R. L., Joyce, D., Collier, T., & Shergill, S. S. (2016). Reduced susceptibility to the sound-induced flash fusion illusion in schizophrenia. *Psychiatry Research*, 245(Supplement C), 58–65. <https://doi.org/10.1016/j.psychres.2016.08.016>
- VanRullen, R. (2016). Perceptual Cycles. *Trends in Cognitive Sciences*, 20(10), 723–735. <https://doi.org/10.1016/j.tics.2016.07.006>

- VanRullen, R. (2018). Perceptual Rhythms. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–44).
<https://doi.org/10.1002/9781119170174.epcn212>
- Varela, F. J., Toro, A., Roy John, E., & Schwartz, E. L. (1981). Perceptual framing and cortical alpha rhythm. *Neuropsychologia*, 19(5), 675–686.
[https://doi.org/10.1016/0028-3932\(81\)90005-1](https://doi.org/10.1016/0028-3932(81)90005-1)
- Veksler, B. Z., & Gunzelmann, G. (2018). Functional Equivalence of Sleep Loss and Time on Task Effects in Sustained Attention. *Cognitive Science*, 42(2), 600–632.
<https://doi.org/10.1111/cogs.12489>
- Vijayan, S., & Kopell, N. J. (2012). Thalamic model of awake alpha oscillations and implications for stimulus processing. *Proceedings of the National Academy of Sciences*, 109(45), 18553–18558. <https://doi.org/10.1073/pnas.1215385109>
- Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, 158(2). <https://doi.org/10.1007/s00221-004-1899-9>
- Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, 64(Supplement C), 105–123.
<https://doi.org/10.1016/j.neuropsychologia.2014.08.005>
- Walsh, E. G. (1952). Visual reaction time and the α -rhythm, an investigation of a scanning hypothesis. *The Journal of Physiology*, 118(4), 500–508.
- Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *NeuroImage*, 37(2), 572–578.
<https://doi.org/10.1016/j.neuroimage.2007.05.027>

- Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, 31(3), 1247–1256. <https://doi.org/10.1016/j.neuroimage.2006.01.016>
- Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2), 70–73. <https://doi.org/10.1109/TAU.1967.1161901>
- Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 19(6), 1057–1064. <https://doi.org/10.3758/s13423-012-0295-x>
- White, C. T. (1963). Temporal numerosity and the psychological unit of duration. *Psychological Monographs: General and Applied*, 77(12), 1–37. <https://doi.org/10.1037/h0093860>
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8), 1293–1313.
- Wichmann, F. A., & Jäkel, F. (2018). Methods in Psychophysics. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–42). <https://doi.org/10.1002/9781119170174.epcn507>
- Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal Detection Measures Cannot Distinguish Perceptual Biases from Response Biases. *Perception*, 44(3), 289–300. <https://doi.org/10.1068/p7908>
- Wozny, D. R., Beierholm, U. R., & Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *Journal of Vision*, 8(3), 24–24. <https://doi.org/10.1167/8.3.24>

- Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability Matching as a Computational Strategy Used in Perception. *PLoS Computational Biology*, 6(8), e1000871. <https://doi.org/10.1371/journal.pcbi.1000871>
- Wutz, A., Melcher, D., & Samaha, J. (2018). Frequency modulation of neural oscillations according to visual task demands. *Proceedings of the National Academy of Sciences*, 115(6), 1346–1351. <https://doi.org/10.1073/pnas.1713318115>
- Yang, C.-T. (2017). 10—Attention and Perceptual Decision Making. In D. R. Little, N. Altieri, M. Fifić, & C.-T. Yang (Eds.), *Systems Factorial Technology* (pp. 199–217). <https://doi.org/10.1016/B978-0-12-804315-8.00013-6>